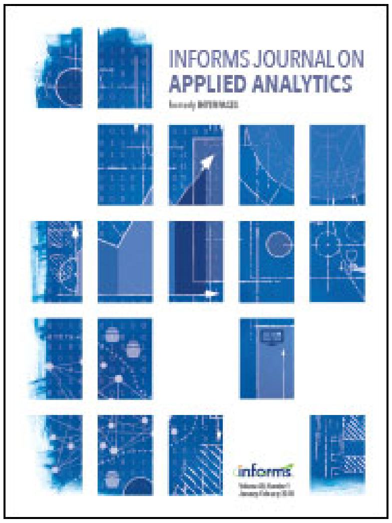


Interfaces

Business Analytics for Streamlined Assort Packing and Distribution of Fashion Goods at Kolon Sport

Shin Woong Sung, Young Jae Jang, Jung Hoon Kim, Juyeong Lee

To cite this article:
Shin Woong Sung, Young Jae Jang, Jung Hoon Kim, Juyeong Lee (2017) Business Analytics for Streamlined Assort Packing and Distribution of Fashion Goods at Kolon Sport. Interfaces 47(6):555-573. https://doi.org/10.1287/inte.2017.0904

informs
http://pubsonline.informs.org/journal/inte/

INTERFACES
Vol. 47, No. 6, November–December 2017, pp. 555–573
ISSN 0092-2102 (print), ISSN 1526-551X (online)

Business Analytics for Streamlined Assort Packing and Distribution of Fashion Goods at Kolon Sport

Shin Woong Sung,[a] Young Jae Jang,[a] Jung Hoon Kim,[b] Juyeong Lee[c]

[a] Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea; [b] Kolon Industries Inc, Seoul 06168, South Korea; [c] SK Telecom, Seoul 04539, South Korea

Contact: sw.sung@kaist.ac.kr, http://orcid.org/0000-0001-9757-2109 (SWS); yjang@kaist.ac.kr, http://orcid.org/0000-0002-2342-1444 (YJJ); justsayit@kolon.com, http://orcid.org/0000-0003-1986-6536 (JHK); skt.juyeonglee@sk.com, http://orcid.org/0000-0002-0270-7777 (JL)

Received: March 28, 2016
Revised: September 16, 2016; January 28, 2017
Accepted: February 15, 2017
Published Online in Articles in Advance: September 18, 2017

https://doi.org/10.1287/inte.2017.0904

Abstract. Kolon Sport (K/S), a leading outdoor fashion brand in South Korea, must deal with a large variety of items during each selling season. In doing so, the company has encountered a challenging operational problem—the assort-packing and distribution problem. The problem involves making decisions on the optimal method to use in packing a set of different items in a box and allocating the boxes to stores to meet the stores' demands. In this paper, we introduce an analytics project initiated to improve the assort-packing and distribution process, and we describe the formulation and solution approach we developed to solve this industrial problem in a timely manner. We validated the proposed approach by computational and onsite pilot testing, which demonstrated that the decisions made using this approach are superior to those made with the manual method K/S used previously. Inventory is distributed to all stores in a more balanced way by considering the demands of the stores. K/S, which implemented the proposed approach into its internal system in July 2015, estimates that the new system improves sales by approximately eight percent.

History: This paper was refereed.

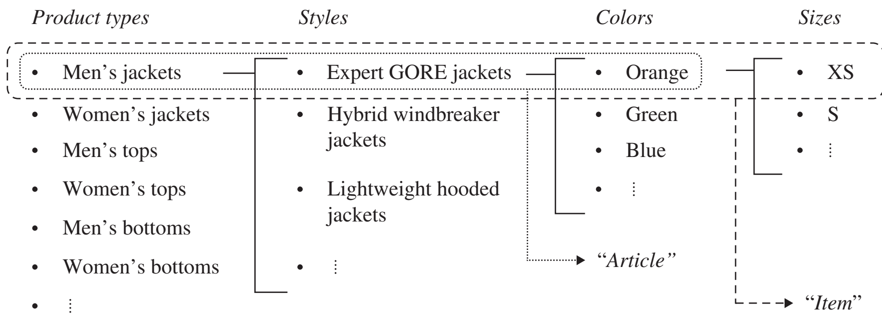
Funding: The authors thank KOLON Corporation, Korea, for providing funding for this research through the KOLON-KAIST Lifestyle Innovation Center Project [LSI14-LSJYJ0001].

Keywords: industries • retail • apparel • inventory and production • applications • programming • integer • supply chain management

Kolon Sport (K/S) is the signature fashion brand of Kolon Industries, Inc. (KRX: KOLONIND [120110]), which was founded in 1957 in South Korea and operates more than 10 fashion brands under its business unit. Total sales at Kolon Industries were approximately $5 billion in 2014 (Kolon Industries 2015). K/S was established in 1973 in Korea. As of 2014, Korea had the world's second-largest market share in the global outdoor fashion industry after the United States (Samsung Fashion Research Institute 2014, Financial Supervisory Service of Korea 2015). Among dozens of domestic outdoor fashion brands in Korea, K/S had a market share that exceeded nine percent share in 2014 and is one of the top three outdoor fashion brands in Korea (Samsung Fashion Research Institute 2014, Financial Supervisory Service of Korea 2015).

K/S operates more than 250 stores nationwide, and designs, manufactures, and distributes more than 4,000 stock-keeping units (SKUs) during each selling season. Its product line is classified hierarchically (Figure 1). It calls an *article* a unit of a product line with a specific product type, style, and color. Each article is further broken down into individual *items* that are distinguished by size. Each product is therefore distinguished by item level with its unique product type, style, color, and size. In this paper, the item level

Figure 1. Product Types, Such as Men's Jackets and Women's Bottoms, Are at the Highest Level of Aggregation



Notes. Each product type is broken down into styles; for example, men's jackets include the hybrid windbreaker jacket as a style. Styles are further broken down into colors; for example, men's hybrid windbreaker jackets can be orange, green, or blue.

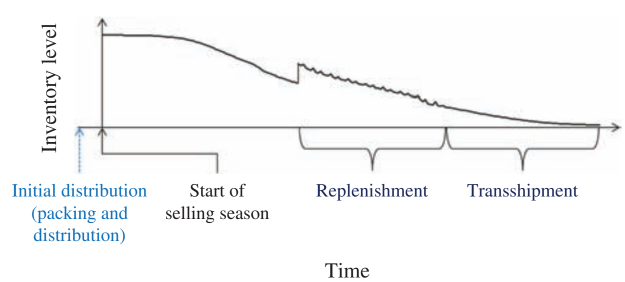
represents a SKU level. Because of industry standards and K/S's business operations, a box for distribution contains multiple items of the same article. The project we describe in this paper addresses the problem of configuring item combinations in boxes to maximize sales, while also meeting the company's business needs.

Because K/S operates multiple stores and its products include a large variety of items, making decisions on how to distribute the items to its stores has been one of its primary challenges. K/S headquarters is responsible for distributing the products to the stores during each selling season. K/S uses three main serial distribution methods during a selling season: initial distribution, replenishment, and transshipment (Figure 2). Conventionally, about 80 percent of its total volume is distributed by initial distribution, and the remaining 20 percent is stored in its distribution center for replenishment during the selling season. Except for some transshipments, headquarters makes every distribution decision. When the selling season is over, headquarters collects the leftover inventory from every store. Because one of the company's brand policies is to position K/S as a premium brand, it rarely uses markdowns to reduce inventory. Therefore, we did not consider the effect of markdowns in our research. In this paper, we discuss the analysis, development, and application of a new approach for the initial distribution at K/S. Our approach comprehensively handles the decisions that involve the packing and distribution of products to stores; these decisions are made just prior the selling season.

Previously, for the initial distribution, all manufactured products at K/S were shipped to the distribution center using bulk boxes that contain products of single items; this is called solid packing. At the distribution center, the items were unpacked, picked, sorted, and repacked into individual boxes for delivery to stores, as we show in the upper panel of Figure 3. Given that K/S distributes more than 1.5 million products of 4,000 different items during each selling season, this repacking process required significant time and effort. Moreover, as a result of long delays in the repacking process, some stores did not receive products until after a selling season had begun.

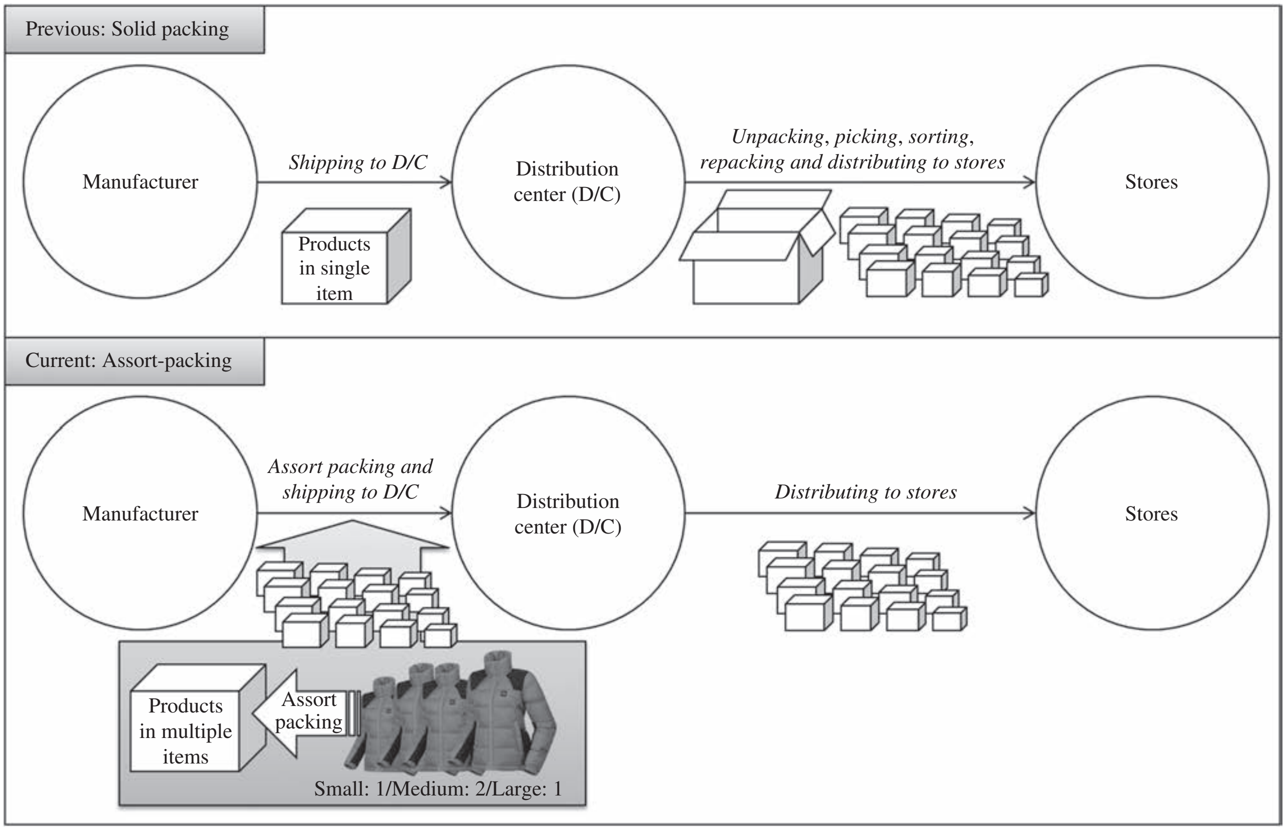
To efficiently distribute various products to multiple stores, K/S recently introduced assort packing into its supply chain. In contrast to solid packing, the term assort packing in this paper refers to putting a set of different items into a box to facilitate the efficient

Figure 2. (Color online) Initial Distribution Refers to Distribution of the Products from the Distribution Center to the Stores Immediately Before the Beginning of the Selling Season



Notes. Because the inventory level drops during the selling season, replenishment from the distribution center to the stores occurs in accordance with the inventory-replenishment policy and the sales manager's decision. After the inventory in the distribution center has been depleted, transshipment between stores occurs by communication between the store managers.

Figure 3. Previous and Current Supply Chain Processes



Notes. The upper panel depicts K/S's previous process, which used solid packing. The lower panel shows its current process, which uses assort packing. An assort-packed box contains multiple items of the same article.

distribution of products. The assort-packed boxes are shipped to the distribution center and then distributed to stores without any repacking process, as we show in the lower panel of Figure 3. Assort packing, therefore, provides significant time and cost savings. It also allows more efficient supply chain operations, such as direct store delivery from the manufacturer or cross-docking operations (Kunz 2009). Assort packing has been used widely in the retail fashion and general merchandise industries, especially where many stores require a variety of items with a relatively small volume (Chettri and Sharma 2008; Kunz 2009; Wang 2010; Vakhutinsky 2011; Vakhutinsky et al. 2011; Hoskins et al. 2014; McMains et al. 2014a, b; Pratt 2014; Fischetti et al. 2015). Usually, an assort-packed box contains items distinguished by only one attribute, such as size (Chettri and Sharma 2008, Pratt 2014). At K/S, items belonging to the same article are packed together. For example, if an article is available in three sizes, such as small, medium, and large, then the assort-packed boxes for the article are made with varying combinations of quantities for the three sizes, such as 3, 6, 4 or 4, 5, 4 for small, medium, and large, respectively.

Although assort packing has provided significant operational benefits, it has generated a challenge. Because manufacturers, consisting of mostly third-party original equipment manufacturing (OEM) suppliers, pack the boxes, K/S cannot request that these suppliers make all the differing *box configuration*s (i.e., possible ways to fill a box with different items) for every article for every store. Therefore, it is common practice to restrict the number of different box configurations used per article in the packing process. That is, of the numerous possible box configurations, only a fixed number of configurations are used to distribute an article. However, the number of each item shipped to each store (i.e., the distribution amount) might not be precisely the required number. If an item is shipped to a

Table 1. This Example Shows an Assort-Packing Decision for Article A, Which Is Available in Five Sizes, When Four Different Box Configurations Can be Used in Its Distribution

	Items (sizes)					
Article A	XS	S	M	L	XL	No. of corresponding boxes
Box configuration 1	2	5	9	7	3	25
Box configuration 2	1	2	4	3	2	80
Box configuration 3	0	2	3	2	1	95
Box configuration 4	0	1	2	1	0	55

Notes. Box configuration 1 contains two extra small (XS), five small (S), nine medium (M), seven large (L), and three extra-large (XL) items. For this configuration, 25 boxes are made.

store in an amount that is more (less) than that store's demand, overshipment (undershipment) can result. Considering that the distribution amount depends on both packing and allocating the boxes to stores, the *assort-packing decision* and the *distribution decision* must be made simultaneously to minimize over- and undershipments. The assort-packing decision involves determining a set of box configurations that will be used in the distribution process (Table 1). The distribution decision involves determining which box configuration and how many of each box configuration to allocate to each store (Table 2). Additionally, the *demand estimation* for each item in each store must be considered in advance to provide input to the assort-packing and distribution decisions. These serial decisions are collectively called *streamlined assort packing and distribution* (*SAPD*), which involves demand estimation, assort-packing, and distribution decisions. In summary, for each article, K/S seeks to find the answers to the following three SAPD questions:

1. Demand estimation: For each item, how many products should be supplied to each store?

Table 2. This Example Shows the Distribution Decision for Article A Using the Box Configurations Listed in Table 1 When There Are 250 Stores

Article A	Box configurations and their numbers
Store 1	Box configuration 1: 1
Store 2	Box configuration 2: 1, Box configuration 3: 1
...	...
Store 250	Box configuration 4: 1

Note. Store 1 receives only one box for Box configuration 1, Store 2 receives one box for Box configuration 2 and one box for Box configuration 3, and Store 250 receives one box for Box configuration 4.

2. Assort-packing decision: For each item, how many products should be in each box configuration, and how many of each box configuration should be made?

3. Distribution decision: Which box configurations and how many box configurations should be allocated to each store?

This paper discusses the development, implementation, and effects of the new decision-making process for SAPD. This work is based on an industry-university collaborative project between Kolon Industries and Korea Advanced Institute of Science and Technology (KAIST), initiated in February 2014. The objective of the project was to improve the efficiency of the assort-packing and distribution operation at K/S using business analytics. The primary team units were the big data analytics team at Kolon Industries (Kolon analytics team) and the system design management laboratory at KAIST (KAIST team). During the project, various SAPD managerial and operational issues at in K/S were investigated through data analysis and interviews. We developed a solution approach for SAPD that has two components: (1) demand estimation, and (2) assort-packing and distribution optimization. We focused mainly on the modeling and optimization of the assort-packing and distribution process, and modeled the demand estimation using a standard multiple linear regression. We validated our proposed approach by computational and onsite pilot testing, which demonstrated that the approach provides K/S with better decisions than its previous manual method, because it allows the company to distribute inventory to all stores in a more balanced way by considering the stores' demands. K/S has fully implemented the proposed approach as a decision support system for

SAPD. It estimates that the proposed SAPD method increased sales by eight percent during the fall and winter seasons of 2015.

Packing and distribution are the critical components in the supply chain within the fashion and apparel industries. Although assort packing is a common practice in these industries and voluminous literature is available on the distribution of inventory, research dealing simultaneously with assort packing and distribution is limited. To the best of our knowledge, this is the first paper to analytically formulate the problem and implement it into an information technology (IT) solution for operations.

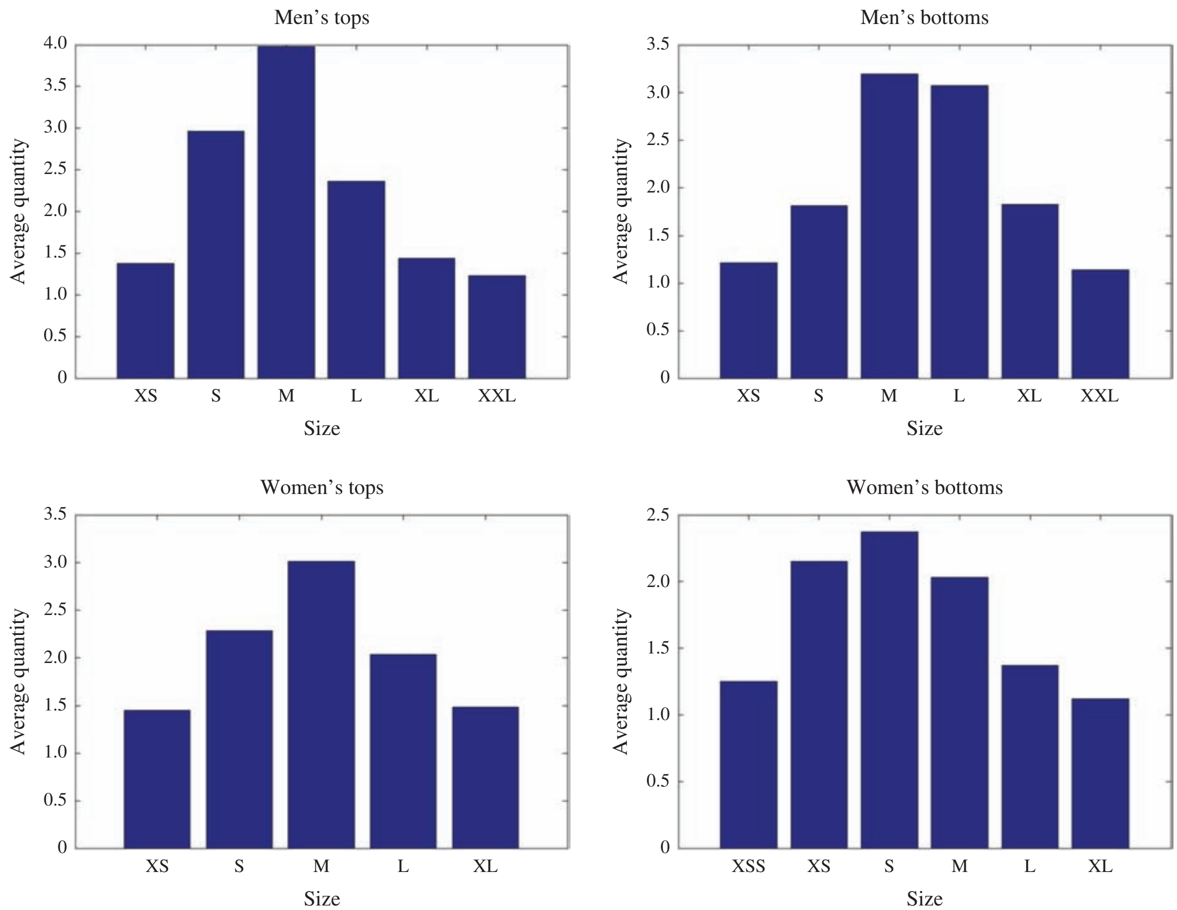
We organized the remainder of this paper as follows. We present the relevant work on SAPD and the manual method that K/S used to solve the SAPD problem in the *Background and Related Work* section. We describe the analytical development of the new decision-making process for SAPD in *The Solution for SAPD* section. The validation of the modeling, analysis, and solution approach is provided in the *Experimental Results and Analysis of the Effect on Sales* section in which we discuss a computational experiment and an onsite pilot experiment that we conducted. The implementation process and its effects on organization are presented in the *Implementation and Organizational Effects* section. Finally, we conclude the paper in the *Summary and Concluding Remarks* section.

Background and Related Work

As we previously mention, SAPD mainly involves (1) the assort-packing decision of allocating many products for each item to each box configuration, and (2) the distribution decision on how many and which box configurations to distribute to each store. These two decisions have been studied widely, but separately, for decades. Most of the packing-decision literature has focused on efficient methods of packing a set of items, while minimizing the number of containers (bins) or the wasted space in the containers (Dyckhoff 1990, Lodi et al. 2002, Wäscher et al. 2007, Bortfeldt and Wäscher 2013). This problem is known familiarly as the *bin packing problem* or *container loading problem*. For the distribution decision, a variety of inventory-management theories have been researched in the context of replenishment policy or integrated models with location decision, vehicle routing, and production planning (Silver 1981, Caro and Gallien 2010, Caro et al. 2010, Federgruen and Zipkin 1984, Chandra and Fisher 1994, Baita et al. 1998, Christiansen 1999, Lawrence 1977).

In the fashion retail industry, assort-packing and distribution decisions have been made simultaneously, because of the nature of fashion market; a variety of items must be provided to multiple stores during a relatively short period, such as during the week before the selling season begins. The integration of assort packing and distribution has been considered recently in the literature under the name of *pack optimization* or *prepack optimization* (Chettri and Sharma 2008; Wang 2010; Vakhutinsky 2011; Vakhutinsky et al. 2011; Hoskins et al. 2014; McMains et al. 2014a, b; Pratt 2014; Fischetti et al. 2015). Chettri and Sharma (2008) first introduced the overall decision-making process involved in the packing and distribution tasks. In patents (Vakhutinsky 2011; Vakhutinsky et al. 2011; McMains et al. 2014a, b; Pratt 2014), leading business solution and software companies have proposed various heuristic approaches to the assort-packing and distribution problem based on their clients' needs, such as iteratively solving the problem in which part of the solution is fixed or some constraints are relaxed. Wang (2010) formulated the model for assort-packing and distribution optimization with inner packs and outer packs, which minimize the handling costs and mismatches between the supply and demand, assuming that the available pack configurations are given. Wang suggested a dynamic programming approach and two heuristic methods for the problem. Hoskins et al. (2014) provided a constraint programming model and a mixed-integer linear programming model for the assort-packing and distribution problem. They also suggested solving it by iteratively finding better box configurations using a metaheuristic algorithm and evaluating fitness values by solving assignment problems with the configurations. Fischetti et al. (2015) iteratively solve a restricted problem of packing and distribution, which they obtain by fixing a subset of variables based on the problem substructure. They found that the number of stores significantly affects the computation time; therefore, they first address the subset of stores and then iteratively add stores. They measured the performance of this method using the data in Hoskins et al. (2014).

Figure 4. (Color online) Each Bar Graph Indicates the Average Quantity of Each Size From the Historical Assort-Packing Data for the Four Product Types in 2013

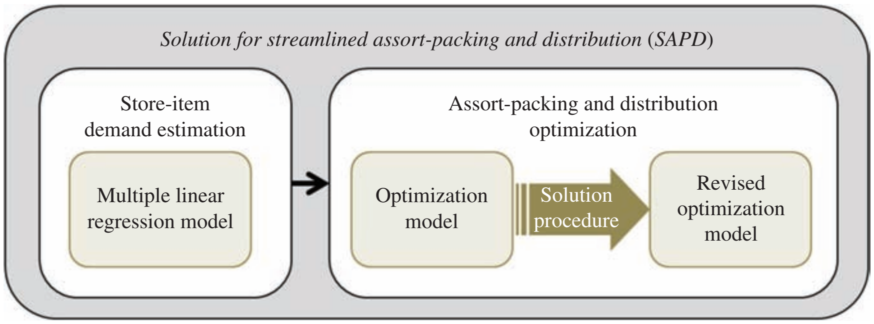


K/S has a much larger number of stores than the numbers used in the previous studies. However, it made its SAPD decisions manually and on an ad hoc basis. The company split the manual procedure into two main stages: (1) determining a set of different box configurations, and (2) assigning the box configurations to stores. Based on our research, the planning managers made approximately three to seven box configurations per article. In addition, most stores received one box per article for the convenience of the store managers who had to handle the hundreds of different articles arriving each season. Figure 4 shows the average distribution of item (size) quantity in a single box, as it was configured previously. The distribution in a box tends to be similar to the demand distribution of the items. Additionally, we found a nonnegligible number of boxes that contained only major sizes (e.g., S, M, L). The planning mangers told us that those boxes were for stores with low sales. Once the managers determined a set of box configurations, they ranked each store's score based on the estimated store-item demand and assigned the box configurations to each store. Stores with a higher rank were assigned larger boxes, which contained more items. Because the decisions on box configurations and their distribution were separate and made manually, the stores frequently received more or fewer items than they needed. Managers had no suitable tool and insufficient time to calculate the difference between the distribution amount and the demands of each store, and could not revise and adjust the decision for every item and store one at a time.

Solution for SAPD

In this section, we describe the analytical development of the new decision-making process for SAPD. Figure 5

Figure 5. (Color online) Our Approach Consists of Two Main Parts: (1) Store-Item Demand Estimation and (2) Assort-Packing and Distribution Optimization



illustrates the overall procedure of our two-part solution approach. In estimating the store-item demand, we aim to estimate the demand for each item in each store using multiple linear regression. The results of the demand estimation provide input for the assort-packing and distribution optimization. We solve an optimization model for assort packing and distribution with a solution procedure, which enables us to treat the industry-size instance in a timely manner. In the next three sections, we present the details of demand estimation, the optimization model, and the solution procedure, respectively.

Store-Item Demand Estimation

Store-item demand estimation is not the primary focus of this paper; we use the standard demand-estimation methods that the retail industry uses widely. Therefore, we briefly describe the approach in this section. Readers who are interested in the details of the approach should refer to Lee (2015) and Jang (2015).

The underlying concept behind store-item demand estimation is that we estimate the store demand for an aggregated level of product (article) and then evaluate the store demand for a lower level of product (item). Furthermore, we do not estimate the absolute demand quantity, but rather the sales contribution of each store.

In estimating the store-item demand, we try to find how many of each item should be supplied to each store. Note that because of the long manufacturing lead time at K/S, the company's marketing department determines the production quantity of each SKU six to eight months before the beginning of the selling season. Therefore, store-item demand estimation in this project is to estimate the store-level distribution quantity for the given produced SKUs—not to predict the absolute demand or production quantity for a particular item for the target selling season. For example, suppose that 6,000 pieces of a winter jacket, Article No. 10234, have been produced for initial distribution, and the quantity for each size, XS, S, M, L, and XL, is 900, 1,200, 1,800, 1,500, and 600, respectively. We must estimate how many of each item (size) needs to be supplied to each store for the given production quantity.

Directly estimating the store-item demand is often challenging because the sales amount for a given item in a given store is small. For example, store-item sales quantities of between 1 and 10 account for 96 percent of the total sales for K/S. Therefore, the common practice in the general retail industry is to estimate the demand for an aggregated (higher) level of products in each store first and then use that value to calculate the demand for the SKU level in each store (Williams and Waller 2011, Fisher and Vaidyanathan 2014, Chien et al. 2015). For aggregated level-of-demand estimation, we predict the *sales contribution* of each store for the given article. Using the aforementioned example, if we estimate that the sales contribution of Store A is five percent for Article No. 10234, then 300 articles are sent to Store A. Then, for a lower level of demand estimation, we make the following assumption: the size distribution of store demand follows the size distribution of overall demand, which is equivalent to production quantity. In the previous example, the size distribution of the production quantity is 15 percent (XS), 20 percent (S), 30 percent (M), 25 percent (L), and 10 percent (XL). We assume that the store-level size distribution also follows this size proportion. Specifically, if Store A is supposed to receive 300 articles of

Article 10234 based on the estimated sales contribution, then it will receive each item quantity in proportion to the size distribution, such as 45 (XS), 60 (S), 90 (M), 75 (L), and 30 (XL). We should note that the size distribution differs across articles, because the production quantity for each size differs across articles.

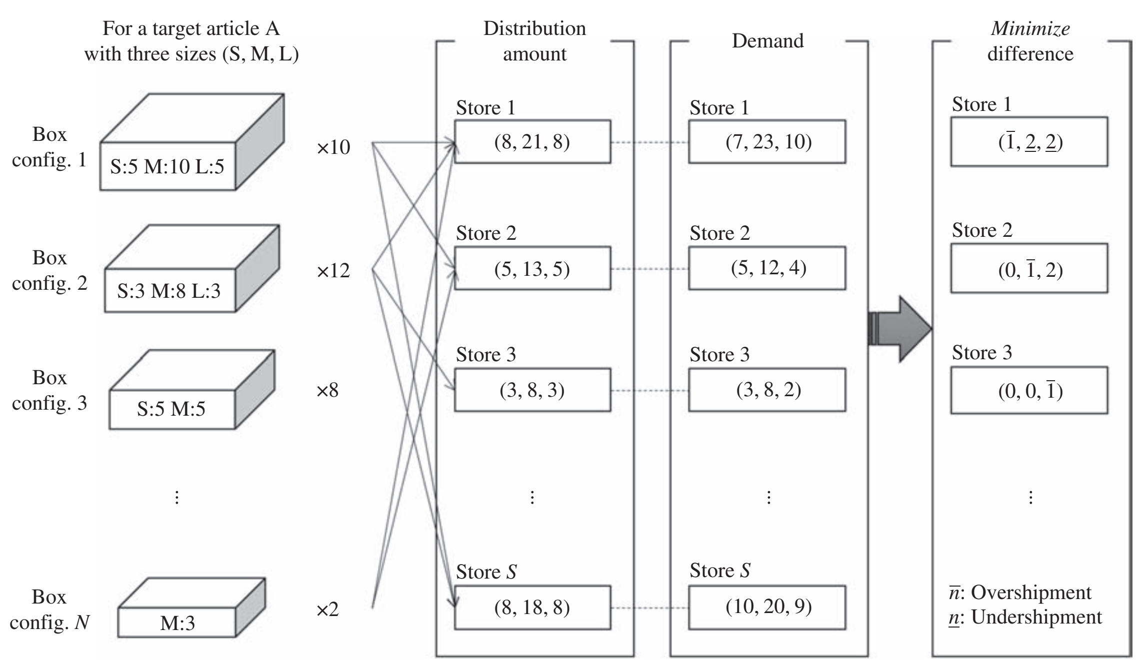
To estimate the sales contribution of each store for a target article, we used multiple linear regression. We consider two main types of input variables: (1) historical sales of each store, and (2) characteristics of each store. The sales of similar articles during the previous year and the total sales performance for the past three months are considered. The previous year's performance is used to reflect the sales with the same seasonal characteristics, whereas the sales performance of the past three months is used to reflect the recent sales patterns of the store. The store characteristics include store location, sales ratio of new and existing customers, and sales ratio of VIP and normal customers. These variables for the multiple linear regression model were selected quantitatively by a statistical variable selection method and qualitatively by the decision of managers in the K/S sales and planning team.

Optimization Model

We develop an optimization model for the assort-packing and distribution decision. Figure 6 describes the assort-packing and distribution optimization problem.

To solve the assort-packing and distribution optimization problem, the following managerial constraints at K/S must be considered. First, the number of different box configurations that can be used for packing is restricted because OEM suppliers perform the assort packing. Traditionally, K/S has used three to seven box configurations per article. The number of different box configurations is determined based on negotiations with the OEM suppliers; therefore, in the optimization problem, we use this number from the K/S planning mangers as a given parameter. Second, the total number of boxes to be used for assort packing and distribution is limited. Because the operational cost for assort packing depends on the number of boxes used, the total number of boxes is restricted. Moreover, as we previously mention, the store managers prefer to receive one box per article because single-box packing for one article is easy to handle and makes managing their inventory convenient. Therefore, in the

Figure 6. For Each Article, We Determine a Set of Box Configurations for Use in the Distribution Process and Assign the Box Configurations to Each Store, While Minimizing the Total Difference Between Distribution Amounts and Demands of Stores, That Is, Total Overshipment and Undershipment



optimization problem, we let the total number of boxes per article be mostly equal to or slightly more than the total number of stores. With this approach, most stores are likely to receive one box per article. Third, each store must receive at least one box, unless the store does not receive any products of the article. Fourth, the total number of products a box can contain is restricted. Specifically, boxes with different total numbers of products are available, and the possible total numbers are given as a set of possible box capacities. For example, if possible box capacities are given as $\{2, 3, 4, 5\}$ for Article A, then the available box configuration is the one with its total number of products equivalent to one of 2, 3, 4, or 5. The number of boxes with a specific capacity is unlimited. The set of possible box capacities is determined based on the physical sizes of the article. In the optimization model, we do not consider the geometric conditions inside the box. Finally, the total distribution quantity of each item cannot exceed the production quantity of each item for initial distribution.

Due to the discrete nature of assort packing and distribution, over- and undershipment may be inevitable; however, K/S wanted to minimize these mismatches between distribution amount and store demand. This situation can be modeled as an optimization problem as follows: Find a best set of box configurations and simultaneously determine its distribution to stores, while minimizing the total over- and undershipment and satisfying the aforementioned constraints. Appendix A provides the details of the mathematical formulation.

Solution Procedure

An important factor in solving our optimization model for practical instances is that many box configurations are possible; thus, the problem is typically computationally intractable. Standard solvers require a significant amount of time to converge because of their high computational and memory requirements. The problem is especially challenging because K/S must make its assort-packing and distribution decisions as quickly as possible; six planning managers must make these SAPD decisions for more than 800 articles during one week prior to each selling season. A rough estimate of the maximum allowable time to determine SAPD for an article, including any review or revision of the decision, is (6 managers × 5 weekdays × 8 hours per day × 60 minutes per hour)/800 articles = 18 minutes per article). We found that the standard solvers, such as SAS/OR and IBM ILOG CPLEX, could not solve the K/S optimization problem even after we ran them for three days of CPU time for relatively small-size instances. Therefore, we devised a solution procedure to find a good, feasible solution in a timely manner.

The key concept underlying our solution procedure is to reduce the number of possible box configurations by providing in advance a set of *reasonable* box configurations to the model, which we expect will give a good, feasible solution. That is, instead of finding a solution among many possible box configurations, we first provide a set of reasonable box configurations and let the optimization select the best configuration within the set. Appendix B describes the alternative formulation for the original optimization model in which a set of reasonable box configurations is given in advance. Jang (2015) presents a comparison of the solutions from the original and the alternative optimization models, thus validating our approach.

To generate the set of reasonable box configurations, we incorporate the following general logic into a solution procedure. First, we create an initial set of box configurations that meet the following two conditions: (1) the configurations satisfy the capacity constraint, and (2) the configurations do not to exceed the maximum demand for each item. As an example, suppose that the possible box capacity is $\{2, 3, 4, \ldots, 10\}$ for an article with three sizes—small, medium, and large, denoted by (S, M, L). Let the maximum demand for each size be $(3, 5, 2)$. That is, in each box, the total number of items should be between 2 and 10 and the number of each size should be less than or equal to $(3, 5, 2)$. Then, configurations such as $(2, 5, 4)$ and $(3, 5, 3)$ cannot be in the candidate set because they violate both conditions.

Next, we select a set of reasonable box configurations from the initial set of box configurations based on the following conjecture: if a store receives one box per article, it will be the optimal box configuration whose size distribution is similar to the size distribution of the production quantity. This conjecture builds on the key assumption of our demand estimation procedure—that the size distribution of each store's demand is identical to the size distribution of the overall demand, which is equivalent to production quantity. The box configurations whose size distribution is similar to that of the production quantity would be more beneficial in minimizing over- and undershipment than the other configurations. For example, suppose we have two box configurations, A and B, each of which contains three different items. Let the items' quantities be $(2, 5, 3)$ in A and $(4, 4, 2)$ in B. Then, the size distributions of each box configuration are $(0.2, 0.5, 0.3)$ and $(0.4, 0.4, 0.2)$, respectively. If the production quantity for the three items is $(300, 1{,}100, 600)$, whose size distribution is $(0.15, 0.55, 0.3)$, then the size distribution of A is more like the size distribution of the production quantity than that of B. That is, if we use a similarity measure as the sum-of-squared difference (SSD), then the SSD from the size distribution of production quantity is 0.005 for A and 0.095 for B. Therefore, we consider A to be the more reasonable (beneficial) box configuration than B.

In Figure 7, we summarize the steps taken to generate a set of reasonable box configurations for the example of an article with three different sizes (S, M, L) with the production quantity (300, 500, 200) (Figure 7). In this example, we set $h = 1$ and select the box configuration with the most similar size distribution to that of the production quantity for each box capacity. For reference, we found with some experimentation that when $h = 3$, the solution procedure provides a good solution within a reasonable computation time. Finally, taking the selected set of reasonable box configurations

Figure 7. In This Example, the Article Is Available in Three Sizes (S, M, L) and the Production Quantity Is (300, 500, 200)

Parameters

- Possible box capacities: {2, 3, 4, …, 10}
- Maximum demand for each size: (3, 5, 2)
- Production quantity for each size: (300, 500, 200)
- $h = 1$

(1) Initial set of box configurations

S	M	L	Total # of products
0	0	2	2
0	1	1	2
0	2	0	2
1	0	1	2
1	1	0	2
2	0	0	2
0	1	2	3
0	2	1	3
0	3	0	3
⋮	⋮	⋮	⋮
3	5	2	10

(2) Size distribution and SSD

S	M	L	Total # of products	SSD
0.00	0.00	1.00	2	0.98
0.00	0.50	0.50	2	0.18
0.00	1.00	0.00	2	0.38
0.50	0.00	0.50	2	0.38
0.50	0.50	0.00	2	0.08
1.00	0.00	0.00	2	0.78
0.00	0.33	0.67	3	0.34
0.00	0.67	0.33	3	0.14
0.00	1.00	0.00	3	0.38
⋮	⋮	⋮	⋮	⋮
0.30	0.50	0.20	10	0.00

(3) Final set of box configurations

S	M	L	Total # of products
1	1	0	2
1	1	1	3
1	2	1	4
1	3	1	5
2	3	1	6
2	4	1	7
2	4	2	8
3	4	2	9
3	5	2	10

Notes. First, we generate an initial set of box configurations for each box capacity. As the first table shows, six different box configurations are possible for the box capacity 2: $(0, 0, 2)$, $(0, 1, 1)$, $(0, 2, 0)$, $(1, 0, 1)$, $(1, 1, 0)$, and $(2, 0, 0)$. Second, given the initial enumeration of the box configurations for each box capacity, we evaluate the size distribution of each box configuration and its SSD from the size distribution of the production quantity (Table 2). Third, we select the box configurations that have the smallest SSD for each box capacity (Table 3).

as input parameters, we solve the alternative optimization formulation with a standard solver.

Experimental Results and Analysis of the Effect on Sales

We conducted two types of experiments to validate the efficiency and effectiveness of the proposed method for SAPD: computational experiments and onsite pilot experiments. Based on the results of the computational experiments, K/S decided to conduct the onsite pilot experiments on its products from the 2015 summer season to investigate the actual effect of our proposed method. We also conducted a simulation analysis to identify the effect of our proposed method on sales. These experiments and analyses led to the actual implementation of the proposed method in the K/S internal system. Jang (2015) includes an additional numerical experiment on the performance of the solution procedure.

Computational Experiments

The goal of the computational experiments is to answer the following question: *Does the proposed method satisfy the estimated demand better than the manual method K/S used*? As test articles, we selected five historical articles from 2014, that is, Articles 1–5, to test various production volumes. We retrieved the records of the manually determined box configurations and distributions from 2014 and compared them to those generated by the proposed optimization method. Specifically, we used the 2014 store-item demand-estimation data. From the 2014 assort-packing and distribution records, we evaluated how the number of distributed products deviated from the store-item demand for all items and stores (i.e., *total deviation*). We also ran the optimization using the estimated store-item demand and then evaluated the total deviation. Finally, we compared these two total deviations for each article.

Table 3 shows the percentage difference in total deviation between the manual and proposed methods for the five different test articles, which we calculated using this formula:

$$(\textit{total deviation of manual method} \\ - \textit{total deviation of proposed method}) \\ \cdot (\textit{total deviation of manual method})^{-1} \times 100\%$$

Table 3. Results of the Computational Experiment Comparing the Manual and Proposed Methods

	Article 1 (%)	Article 2 (%)	Article 3 (%)	Article 4 (%)	Article 5 (%)
Percentage difference	6.42	20.39	27.35	30.82	42.97

Note. Each percentage value indicates the amount by which the proposed method reduces the total deviation when we compare it to the manual method.

Table 3 shows how our proposed method improves (i.e., decreases) the total deviation when we compare it to the manual method. Note that the articles on the x axis are arranged based on the total production quantity. That is, an article with higher production volume is placed on the right; Article 1 therefore has the lowest volume of production and Article 5 has the highest among the five articles. As we show in the graph, the proposed method outperforms the manual method for all five articles. The larger the production volume of an article, the higher the benefit from the proposed method because the assort combination for an article with a low production volume might be trivial and the manual box configurations can generate a good solution. In contrast, for an article with high production quantity, the combination of the box configurations becomes more complicated, and finding a good solution using the manual approach becomes impractical.

On-Site Pilot Experiments

Based on the computational results, which clearly demonstrate the effectiveness of our solution procedure, the executives and the managers at K/S decided to conduct an on-site pilot experiment using actual products. In February 2015, the SAPD for the articles of the 2015 summer season were made based on the output from our proposed solution method. The goal of the on-site pilot experiments was to compare the manual method and the proposed one in terms of satisfying the actual demand of the overall stores.

The test articles were selected by the K/S sales and planning team using the following two criteria. First, the style and color of each article should be basic, not trendy, to reduce the influence of the product characteristics. Second, the total distribution quantity of each article should be sufficient to provide statistical significance in the results of the experiment. For an article with a small distribution quantity, stockout might occur in some stores early in the selling season due to the small amount of inventory received. In this case, transshipments between stores might occur early in the season; consequently, measuring the effect of the two methods on initial distribution would become difficult. Based on these criteria, K/S selected two articles for testing, Articles 1 and 2.

We divided the stores into two groups, Group A and Group B, of equal size with respect to their historical sales to ensure that the two groups would have comparable net sales. For Article 1, stores in Group A received the products based on the manual method, and stores in Group B received the products based on the proposed method. For Article 2, we selected Group B to use the manual method and Group A to use our proposed method.

For the performance evaluation of the two methods, we use the ratio between the distribution quantity to a store and the midseason leftover quantity in the store. Note that using conventional performance metrics, such as *sales quantity*, might be ambiguous for evaluating the performance of the assort-packing and distribution activity, because the sales are affected not only by the assort-packing and distribution activity but also by other activities such as transshipment between stores and replenishment policies. Considering that the goal of the initial distribution is to best allocate the products so that the stores hold a suitable level of inventory for their demand, the KAIST and the Kolon Analytics teams focused on observing the correlation between the distribution amount and the on-hand inventory level for each store at a point sometime in the middle of the sales season. This point, which is the evaluation date, is selected before product replenishment starts. Thus, the effects of the replenishment policy and transshipments are excluded. The underlying premise of the test is as follows. Consider the ratio of the on hand inventory level to the distributed quantity for a given article at a store. If the articles are distributed appropriately to the stores at the beginning of the season, the ratios should be similar over the stores at this point. In contrast, if the distribution is not done properly, there will be discrepancies in the ratios among the stores—some stores will have insufficient inventory, and others will have too much inventory. For our experiment, we selected the April 30, 2015 as the evaluation date; by this date, nearly one-third of the selling season was over and replenishment had not yet started.

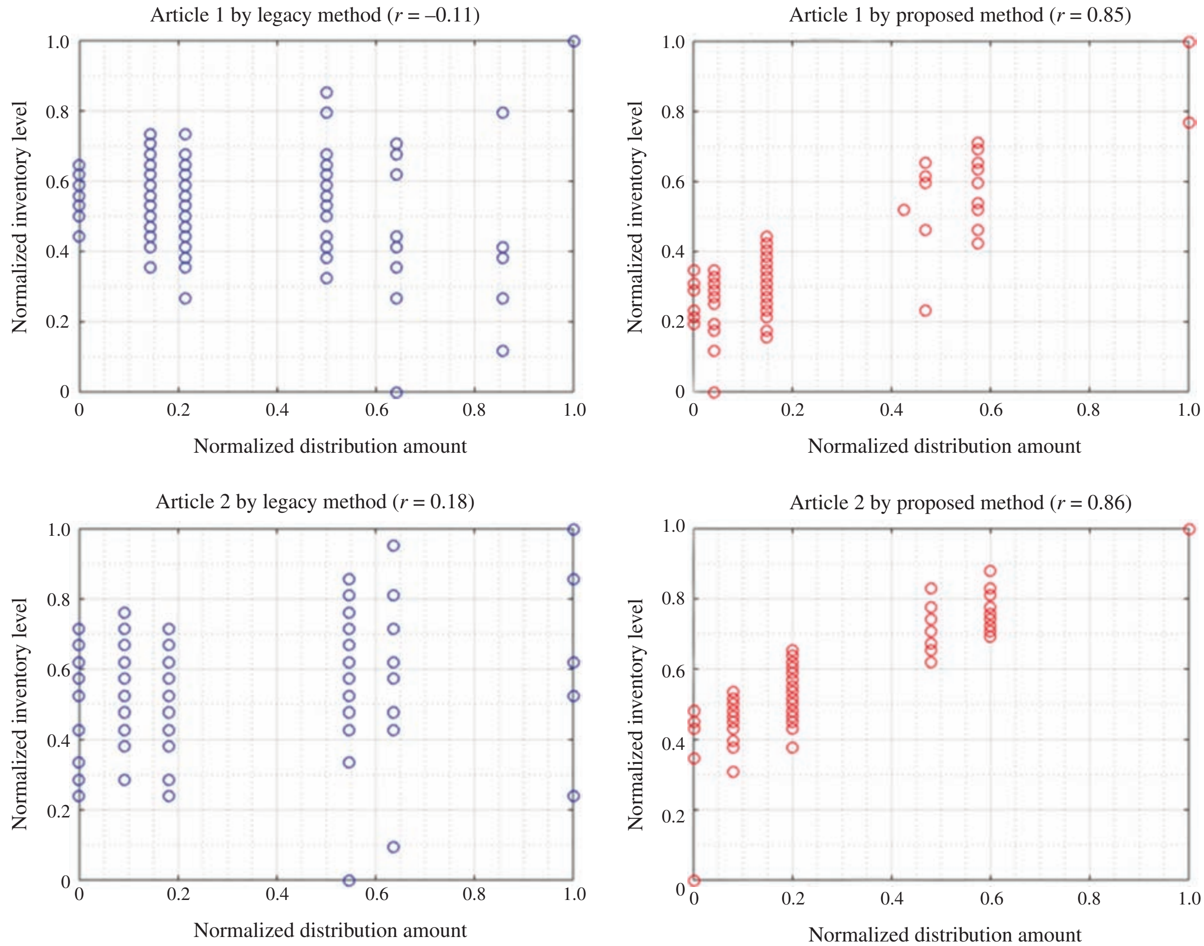
Figure 8 summarizes the comparison of the correlations between the on hand inventory level and the distribution amount for every store by the two methods. The results clearly imply that the proposed method allocates the inventory more appropriately to the store demands than the manual method does. Specifically, the Pearson correlation coefficient (r) for each method on each test article is given in the title of the graphs. The r values for the proposed method are much higher than those for the manual method, and the graphs also show the strong correlation between distribution amount and inventory level for the proposed method, whereas they show no apparent correlation for the manual method.

Analysis of the Effect on Sales

To verify and validate the proposed method with randomly generated demand, we ran a simulation in which we modeled the sales process of a selling season. We developed the simulation model developed using ARENA simulation software, and implemented the sales process, which includes activities such as initial distribution of inventory, demand generation, and sales. The key performance indicator is sales volume (i.e., the number of sales in a selling season). We compared the sales volumes, resulting from the initial distribution decision, using the proposed and the manual methods. We tested 10 articles from 160 stores that K/S selected from the fall and winter season of 2014. The daily demand of each article in each store was generated by a Poisson distribution with the mean of the daily sales volume of each article obtained from historical transaction data. For each article, we ran the simulation for 153 days (from August to December), which is the period when the sales were mainly occurring, with the initial distribution quantity for the fall-winter season. The number of replications for each article was 100, and we applied the same set of demand scenarios to both methods. Therefore, we had 100 paired observations (proposed and manual) for each article.

Table 4 shows the sample mean and sample standard deviation of the sales volume from each method. We denote the sales volume from the proposed method

Figure 8. (Color online) In the Plots, Each Dot Represents a Single Store with the Normalized Initial Distribution Amount and On-Hand Inventory Level at the Evaluation Date



Notes. The quantities are normalized between 0 and 1 to cover the actual numbers in the plot. In each graph, the x axis corresponds to the normalized initial distribution amount, and the y axis corresponds to the normalized inventory level that is on hand at the evaluation date. Then, we can evaluate the ratio for each store from the x and y values. With our premise, we can say that the distribution is good if the x and y values have a high linear correlation.

by PM, the sales volume from the manual method by MM, and the difference between the sales volumes from the two methods by D = PM − MM. To determine if the proposed method improves sales volume, we test the null hypothesis that the true mean difference is zero. As we demonstrate in Table 4, the results show statistically significant differences between the sales volumes from each method. In most cases (except Article 8), the proposed method outperforms the manual one; all have 99 percent confidence intervals with positive values. That is, we see statistically significant increases in the sales volumes when we use the proposed method. This suggests that the new system for SAPD has a positive effect on K/S sales, consistent with the results of the on-site pilot experiment.

When we reported the results of the computational and on-site pilot experiments and the simulation analysis to the business executives at K/S, they became convinced of the reliability of the proposed method. Management therefore agreed to implement our SAPD solution into the K/S IT system for daily use.

Implementation and Organizational Effects

Implementation

In May 2015, the Kolon Industries IT department implemented the proposed method into the K/S internal

Table 4. We Performed the Paired *T*-Test to Compare the Sales of the Manual Method (MM) and Proposed Method (PM)

	Manual method (MM)		Proposed method (PM)		Difference (D = PM – DM)			
Article	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. error	99% maximum error	99% confidence interval
1	426.37	18.46	434.13	20.11	7.76	0.73	1.92	[5.84, 9.68]
2	288.43	12.98	292.40	16.21	3.97	1.17	3.07	[0.90, 7.04]
3	8.05	1.42	9.55	1.86	1.50	0.09	0.25	[1.25, 1.75]
4	11.94	2.03	15.01	2.60	3.07	0.13	0.34	[2.73, 3.41]
5	19.19	3.85	22.62	4.53	3.43	0.15	0.38	[3.05, 3.81]
6	17.94	3.20	20.35	3.95	2.41	0.16	0.42	[1.99, 2.83]
7	148.2	9.39	158.78	11.24	10.58	0.45	1.19	[9.39, 11.77]
8	4.14	1.62	3.87	1.50	−0.27	0.05	0.13	[−0.40, −0.14]
9	107.52	5.71	112.48	6.45	4.96	0.88	2.32	[2.64, 7.28]
10	25.26	2.70	32.69	3.47	7.43	0.21	0.55	[6.88, 7.98]

Notes. The sales difference is defined as D = PM – MM. Note that all the sales except Article 8 improve when we use the proposed method.

system, based on the prototype system developed by the KAIST team. The database queries and the procedure for generating input data were programmed using SAS, and the computational procedure for solving the optimization model was developed using SAS/OR. The input and output interfaces were designed with a Microsoft Excel spreadsheet using the SAS add-in for Microsoft Office. This allows managers to load and modify the input data, perform the optimization for SAPD, and read and modify the output. Figure 9 shows a screenshot of the user interface. System design and implementation were completed in July 2015.

Effect on Organization

The new method improves the efficiency of the decision-making process. It significantly reduces the time and effort needed to make assort-packing and distribution decisions, which can be burdensome for planning managers, because the managers must make these decisions for more than 800 articles within a short period prior to each selling season. The method allows the managers to focus on additional analytical and productive tasks, such as sales monitoring and what-if analysis. More importantly, the managers feel comfortable with the new method because it reflects their empirical knowledge and gives logical and consistent output.

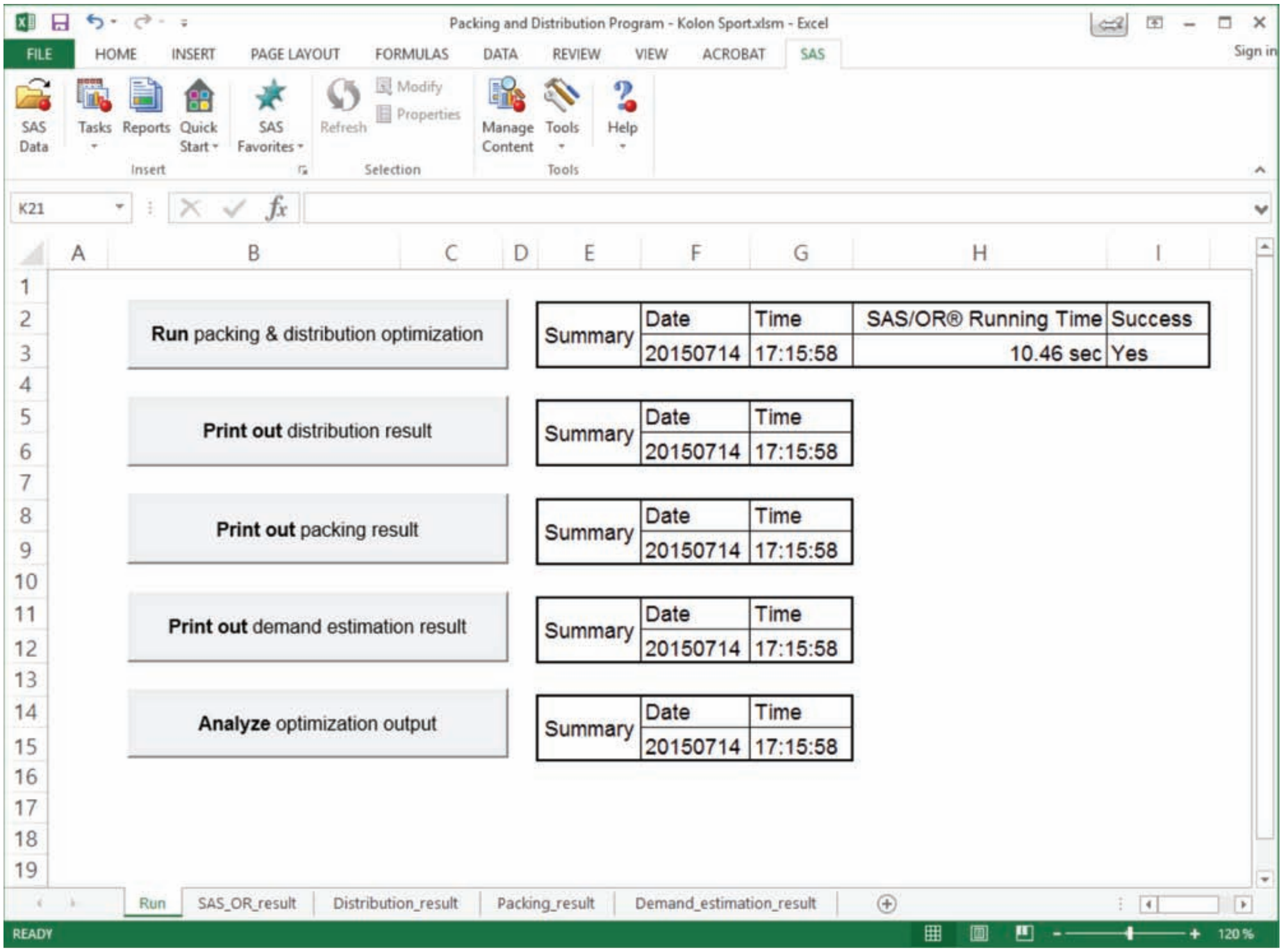
The series of analytical processes undertaken throughout the project has provided guidance on the future direction of the big data analytics team at Kolon Industries. Beyond the data analysis, the team now places renewed importance on identifying the problem, developing the decision support method, and implementing it in the internal system to derive specific effects. In addition to planning the deployment of the optimized assort-packing and distribution process to the other fashion brands of Kolon Industries, the team is beginning to expand the project to the upstream and downstream assort process of packing and distribution to systemize the overall decision-making process using business analytics techniques.

Summary and Concluding Remarks

In this paper, we demonstrate how K/S improved its decision-making process for SAPD and discuss its effects. We describe the business analytics project jointly conducted by Kolon Industries and KAIST. During this project, we successfully developed a system to assist in the decision-making process for assort packing and distribution at K/S.

The Kolon analytics team estimates that the project eventually improved sales by 8–10 percent. Specifically, in the verification letter, the Vice President of Kolon Industry confirmed that the proposed SAPD method contributed to the 8–10 percent sales increase it saw in the fall and winter seasons of 2015 and improved the under- and over-distribution rates by 10–15 percent.

Figure 9. (Color online) Screenshot of the User Interface of the New System for SAPD



We estimate that costs related to the man-hour working time needed to determine assort packing and distribution and to the additional box usage to fulfill the unmet demand of the stores will be significantly reduced.

Our future studies include evaluating the cost impact of the practical constraints in the model (e.g., preferring one box per article in each store). Through detailed cost and sensitivity analyses on those practical considerations, we believe that we can enhance the solution for SAPD. Additionally, in the project, the number of total box configurations used for packing was limited and given by the OEM supplier. Without this limit, we could achieve more accurate delivery; however, Kolon Industry would be required to pay a higher price to the OEM supplier. Therefore, there is a tradeoff between the benefit and cost of having flexibility in the distribution. The same applies to the limit for the total number of boxes. We suggested this research topic to Kolon Industry for a follow-up project.

Appendix A. Optimization Model for Assort Packing and Distribution

In this appendix, we present an optimization model formulation of the assort-packing and distribution problem. The goal of the optimization model is to determine (1) how many products of each item should be put into each box configuration, and (2) how many of each box configuration should be distributed to each store, while minimizing total over-and undershipments to stores.

Index Sets

$\mathcal{I}$ = Items indexed with i.
$\mathcal{S}$ = Stores indexed with s.
$\mathcal{B}$ = Box configuration ID indexed with b. $|\mathcal{B}| = N$, where N is the maximum number of different box configurations that can be used for packing.

Note that all these sets are given.

Parameters

d_{is} = Demand quantity for item i in store s.
K = Set of possible box capacities. A capacity value in this set is denoted by $k \in K$.
M = Maximum number of boxes.
p_i = Production quantity for item i.

Variables

x_{bs} = Number of boxes for box configuration ID b distributed to store s.
y_{ib} = Number of products for item i in box configuration ID b.
t_{bk} = 1 if box configuration ID b has its capacity value with k and 0 otherwise.
o_{is} = Overshipment of item i to store s.
u_{is} = Undershipment of item i to store s.

The optimization model for assort packing and distribution, denoted by (OPT-0), is then

$$\text{(OPT-0)} \quad \text{Minimize} \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}}(u_{is}+o_{is}) \tag{A.1}$$

subject to

$$\sum_{b\in\mathcal{B}} y_{ib}x_{bs} - o_{is} + u_{is} = d_{is},\ i\in\mathcal{I},\ s\in\mathcal{S}, \tag{A.2}$$

$$\sum_{i\in\mathcal{I}} y_{ib} = \sum_{k\in\mathcal{K}} k t_{bk}, \quad b\in\mathcal{B}, \tag{A.3}$$

$$\sum_{k\in\mathcal{K}} t_{bk} = 1, \quad b\in\mathcal{B}, \tag{A.4}$$

$$\sum_{b\in\mathcal{B}}\sum_{s\in\mathcal{S}} x_{bs} \le M, \tag{A.5}$$

$$\sum_{b\in\mathcal{B}} x_{bs} \ge 1, \quad s\in\mathcal{S}, \tag{A.6}$$

$$\sum_{s\in\mathcal{S}}\sum_{b\in\mathcal{B}} y_{ib}x_{bs} \le p_i, \quad i\in\mathcal{I}, \tag{A.7}$$

$$x_{bs} \ge 0, \quad \text{integer } b\in\mathcal{B},\ s\in\mathcal{S}, \tag{A.8}$$

$$y_{ib} \ge 0, \quad \text{integer } i\in\mathcal{I},\ b\in\mathcal{B}, \tag{A.9}$$

$$t_{bk}\in\{0,1\}, \quad b\in\mathcal{B},\ k\in\mathcal{K}, \tag{A.10}$$

$$o_{is},\ u_{is} \ge 0, \quad i\in\mathcal{I},\ s\in\mathcal{S}. \tag{A.11}$$

The objective function (A.1) minimizes the sum of the over- and undershipment for all items in all stores. Constraint (A.2) computes over- and undershipment by comparing the distribution amount and the demand. If the distribution amount of an item shipped to a store is greater than its demand, overshipment occurs, and we compute the overshipment as the distribution amount minus the demand. If the distribution amount of an item shipped to a store is smaller than its demand, undershipment occurs, and we compute the undershipment as the demand minus the distribution amount. Constraint (A.3) is used to ensure the box configuration has a possible capacity. Constraint (A.4) allows each box configuration to have only one of the possible capacities. Constraint (A.5) bounds the total number of boxes. Constraint (A.6) lets each store receive at least one box. Constraint (A.7) is used to restrict the total distribution quantity of each item such that it does not exceed the production quantity of that item. Constraints (A.8)–(A.11) represent the conditions on the decision variables.

As a result of the product of the two decision variables in constraint (A.2) and (A.7), the model becomes nonlinear. It can be linearized by decomposing each x_{bs} variable into its binary expansion (Hoskins et al. 2014, Fischetti et al. 2015). The linearized model, (OPT-1), is given as follows:

$$\text{(OPT-1)} \quad \text{Minimize} \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{I}}(u_{is}+o_{is}) \tag{A.12}$$

subject to

$$x_{bs} = \sum_{l} 2^l v_{bsl}, \quad b\in\mathcal{B},\ s\in\mathcal{S}, \tag{A.13}$$

$$v_{bsl}\in\{0,1\}, \quad b\in\mathcal{B},\ s\in\mathcal{S},\ l=0,\ldots,L, \tag{A.14}$$

$$\sum_{b\in\mathcal{B}}\sum_{l} 2^l w_{bisl} - o_{is} + u_{is} = d_{is},\ i\in\mathcal{I},\ s\in\mathcal{S}, \tag{A.15}$$

$$w_{bisl} \le \bar{Y} v_{bsl} \quad b\in\mathcal{B},\ i\in\mathcal{I},\ s\in\mathcal{S},\ l=0,\ldots,L, \tag{A.16}$$

$$w_{bisl} \le y_{ib}, \quad b\in\mathcal{B},\ i\in\mathcal{I},\ s\in\mathcal{S},\ l=0,\ldots,L, \tag{A.17}$$

$$w_{bisl} \ge y_{ib} - \bar{Y}(1-v_{bsl}), \quad b\in\mathcal{B},\ i\in\mathcal{I},\ s\in\mathcal{S},\ l=0,\ldots,L, \tag{A.18}$$

$$w_{bisl} \ge 0, \quad b\in\mathcal{B},\ i\in\mathcal{I},\ s\in\mathcal{S},\ l=0,\ldots,L, \tag{A.19}$$

$$\sum_{i\in\mathcal{I}} y_{ib} = \sum_{k\in\mathcal{K}} k t_{bk}, \quad b\in\mathcal{B}, \tag{A.20}$$

$$\sum_{k\in\mathcal{K}} t_{bk} = 1, \quad b\in\mathcal{B}, \tag{A.21}$$

$$\sum_{b\in\mathcal{B}}\sum_{s\in\mathcal{S}} x_{bs} \le M, \tag{A.22}$$

$$\sum_{b\in\mathcal{B}} x_{bs} \ge 1, \quad s\in\mathcal{S}, \tag{A.23}$$

$$\sum_{s\in\mathcal{S}}\sum_{b\in\mathcal{B}}\sum_{l} 2^l w_{bisl} \le p_i. \quad i\in\mathcal{I}, \tag{A.24}$$

$$x_{bs} \ge 0, \quad \text{integer } b\in\mathcal{B},\ s\in\mathcal{S}, \tag{A.25}$$

$$y_{ib} \ge 0, \quad \text{integer } i\in\mathcal{I},\ b\in\mathcal{B}, \tag{A.26}$$

$$t_{bk}\in\{0,1\}, \quad b\in\mathcal{B},\ k\in\mathcal{K} \tag{A.27}$$

$$o_{is},\ u_{is} \ge 0, \quad i\in\mathcal{I},\ s\in\mathcal{S}. \tag{A.28}$$

Note that this linearized model is still intractable for industry-size instances as the number of variables and constraints significantly increases due to linearization. Exact algorithms or previously suggested iterative computational approaches would require excessive computational resources to cope with the large-size instances. Therefore, we developed a solutions-based approach to obtain a good, feasible solution for the industry-size problem in a timely manner.

Appendix B. Alternative Optimization Model for Assort-Packing and Distribution

In this appendix, we present an alternative optimization approach developed to solve the assort-packing and distribution problem with industry-size instances in a timely manner. Next, we show the reformulation of the optimization model described in Appendix A and its solution method.

We first change the original formulation using the following concept: instead of determining how many products of each item should be put into each box configuration in the model, we provide in advance a set of feasible box configurations to the model and let the model select the box configurations in the given set. That is, we change the model to determine which box configurations are used among the feasible ones and how many of them are distributed to each store. Therefore, we devise a method to select a set of *reasonable* box configurations that we expect will give a good solution among all possible box configurations. In steps 1–5, we describe the method for generating set $\mathcal{R}$ of reasonable box configurations. The notations and indices in the alternative optimization model follow those defined for the original formulation.

1. Create set $\mathcal{P}$ of initial box configurations such that it satisfies the following conditions: (a) the number of products in the box configurations must be equal to one of the values in $\mathcal{K}$, and (b) it does not to exceed the maximum demand for each item.

2. Make subset $\mathcal{P}_k \subset \mathcal{P}$ for $k \in \mathcal{K}$, such that the capacity value of every box configuration in subset $\mathcal{P}_k$ is k: $\mathcal{P}_k = \{P_k(1), P_k(2), \ldots, P_k(G)\}$, where $P_k(\cdot)$ denotes the box configuration in $\mathcal{P}_k$ and $G = |\mathcal{P}_k|$.

3. For every box configuration $P_k(\cdot)$ in $\mathcal{P}_k$, compute the sum-of-squared difference (SSD) between the size distribution of the box configuration and the size distribution of production quantity, which is denoted by $SSD(P_k(\cdot))$.

4. For each $\mathcal{P}_k$, select the first h smallest values of $SSD(P_k(\cdot))$ and store the corresponding box configurations in set $\mathcal{R}_K$.

5. Make set $\mathcal{R}$ of reasonable box configurations by $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \cdots \cup \mathcal{R}_{|k|}$.

Using this set $\mathcal{R}$, the alternative optimization model selects the best set of box configurations in $\mathcal{R}$ and assigns the box configurations to the stores. We denote by c_{ri} the number of products for item $i \in \mathcal{J}$ in box configuration $r \in \mathcal{R}$.

For the decision variables, we denote by x_{rs} the number of box configurations $r \in \mathcal{R}$ distributed to store $s \in \mathcal{S}$. Binary variable z_r is used to indicate whether a box configuration $r \in \mathcal{R}$ is selected. Finally, if the quantity of item $i \in \mathcal{J}$ shipped to store $s \in \mathcal{S}$ is more (less) than d_{is}, overshipment o_{is} (undershipment u_{is}) occurs, respectively.

The alternative optimization model for assort packing and distribution, denoted by (OPT-2), is then

$$\text{(OPT-2)} \qquad \text{Minimize} \sum_{s\in\mathcal{S}}\sum_{i\in\mathcal{J}}(u_{is}+o_{is}) \tag{B.1}$$

subject to

$$\sum_{r\in\mathcal{R}} c_{ri}x_{rs} - o_{is} + u_{is} = d_{is},\ i\in\mathcal{J},\ s\in\mathcal{S}, \tag{B.2}$$

$$\sum_{r\in\mathcal{R}} Z_r \le N, \tag{B.3}$$

$$\sum_{r\in\mathcal{R}}\sum_{s\in\mathcal{S}} x_{rs} \le M, \tag{B.4}$$

$$\sum_{r\in\mathcal{R}} x_{rs} \ge 1, \quad s\in\mathcal{S}, \tag{B.5}$$

$$\sum_{r\in\mathcal{R}}\sum_{s\in\mathcal{S}} c_{ri}x_{rs} \le p_i, \quad i\in\mathcal{J}, \tag{B.6}$$

$$\sum_{s\in\mathcal{S}} x_{rs} \le Mz_r, \quad r\in\mathcal{R}, \tag{B.7}$$

$$\sum_{s\in\mathcal{S}} x_{rs} \ge z_r, \quad r\in\mathcal{R}, \tag{B.8}$$

$$x_{rz} \ge 0, \quad \text{integer } r\in\mathcal{R},\ s\in\mathcal{S}, \tag{B.9}$$

$$z_r \in \{0,1\}, \quad r\in\mathcal{R}, \tag{B.10}$$

$$o_{is},\ u_{is} \ge 0, \quad i\in\mathcal{J},\ s\in\mathcal{S}. \tag{B.11}$$

References

Baita F, Ukovich W, Pesenti R, Favaretto D (1998) Dynamic routing-and-inventory problems: A review. *Transportation Res. A* 32(8): 585–598.

Bortfeldt A, Wäscher G (2013) Constraints in container loading—A state-of-the-art review. *Eur. J. Oper. Res.* 229(1):1–20.

Caro F, Gallien J (2010) Inventory management of a fast-fashion retail network. *Oper. Res.* 58(2):257–273.

Caro F, Gallien J, Díaz M, García J, Corredoira JM, Montes M, Ramos JA, Correa J (2010) Zara uses operations research to reengineer its global distribution process. *Interfaces* 40(1):71–84.

Chandra P, Fisher ML (1994) Coordination of production and distribution planning. *Eur. J. Oper. Res.* 72(3):503–517.

Chettri IS, Sharma D (2008) Pre pack optimization: Increasing supply chain efficiency. Accessed March 21, 2017, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.6116&rep=rep1&type=pdf.

Chien YH, Wang P, Li Y (2015) Classifying, clustering, and grouping demand series. U.S. Patent 20,150,302,432, filed May 5, 2011, issued October 22, 2015.

Christiansen M (1999) Decomposition of a combined inventory and time constrained ship routing problem. *Transportation Sci.* 33(1):3–16.

Dyckhoff H (1990) A typology of cutting and packing problems. *Eur. J. Oper. Res.* 44(2):145–159.

Federgruen A, Zipkin P (1984) A combined vehicle routing and inventory allocation problem. *Oper. Res.* 32(5):1019–1037.

Financial Supervisory Service of Korea (2015) Quarterly report of Kolon Industries (in Korean). Accessed November 26, 2015, http://dart.fss.or.kr/.

Fischetti M, Monaci M, Salvagnin D (2015) Mixed-integer linear programming heuristics for the prepack optimization problem. *Discrete Optim.* 22(A):195–205.

Fisher M, Vaidyanathan R (2014) A demand estimation procedure for retail assortment optimization with results from implementations. *Management Sci.* 60(10):2401–2415.

Hoskins M, Masson R, Melançon GG, Mendoza JE, Meyer C, Rousseau LM (2014) The prepack optimization problem. Simonis H, ed. *Integration of AI and OR Techniques in Constraint Programming* (Springer International Publishing, Berlin), 136–143.

Jang YJ (2015) KOLON-KAIST Innovative project: Big data and business analytics application research. Report, FnC Organization, Kolon Industries, Inc., Gwacheon, Korea.

Kolon Industries (2015) Financial Information: Condensed financial statements (consolidated). Accessed November 26, 2015, http://www.kolonindustries.com/Eng/Service/service02.asp.

Kunz GI (2009) *Merchandising: Theory, Principles and Practice*, 3rd ed. (Fairchild, New York).

Lawrence MJ (1977) Practice of management science—An integrated inventory control system. *Interfaces* 7(2):55–62.

Lee JY (2015) Heuristic approach for prepack optimization of product distribution. Master's thesis, Korea Advanced Institute of Science and Technology, Daejeon, Korea.

Lodi A, Martello S, Monaci M (2002) Two-dimensional packing problems: A survey. *Eur. J. Oper. Res.* 141(2):241–252.

McMains TD, Burkhardt T, Pratt RW (2014a) Systems and methods for determining pack allocations. U.S. Patent 8,688,497, filed January 10, 2011, issued April 1, 2014.

McMains TD, Burkhardt T, Pratt RW (2014b) Systems and methods for determining pack allocations. U.S. Patent 8,788,315, filed January 10, 2011, issued July 22, 2014.

Pratt RW (2014) Computer-implemented systems and methods for pack optimization. U.S. Patent 8,812,338, filed April 29, 2008, issued August 19, 2014.

Samsung Fashion Research Institute (2014) Market report: Prospects and strategies of apparel goods in 2014. Accessed November 26, 2015, http://www.samsungdesign.net/.

Silver EA (1981) Operations research in inventory management: A review and critique. *Oper. Res.* 29(4):628–645.

Vakhutinsky A (2011) Retail pre-pack optimization system. U.S. Patent 20120284071, filed May 5, 2011, issued November 8, 2012.

Vakhutinsky A, Subramanian S, Popkov Y, Kushkuley A (2011) Retail pre-pack optimizer. U.S. Patent 20120284079, filed May 5, 2011, issued November 8, 2012.

Wang P (2010) Pack optimization problem: Models and solution methods. Doctoral dissertation, North Carolina State University, Raleigh, NC.

Wäscher G, Haußner H, Schumann H (2007) An improved typology of cutting and packing problems. *Eur. J. Oper. Res.* 183(3): 1109–1130.

Williams BD, Waller MA (2011) Top-down versus bottom-up demand forecasts: The value of shared point-of-sale data in the retail supply chain. *J. Bus. Logist.* 32(1):17–26.

Verification Letter

Jae-Hyuk Jang, Vice President, Kolon Industries, Kolon Tower, 11, Kolon-ro, Gwacheon-si, Gyeonggi-do do, South Korea, writes:

"As the Vice President of Kolon Industries, Inc. (http://www.kolonindustries.com/Eng/), and leader of the Big Data Team for Kolon, I am writing to verify the company's actual use of and resulting benefits from the project described in the manuscript entitled 'Business Analytics for Streamlined Assort-Packing and Distribution (SAPD) of Fashion Goods at Kolon Sport' submitted by Young Jae Jang and Shin Woong Sung.

"Kolon Industries is the top fashion company in South Korea, and Kolon Sport is the company's signature brand in outdoor fashion. In 2014, the company created a Big Data Team to transform itself into a data-oriented analytical organization. The team is responsible not only for data analysis and the development of algorithms, but also for providing managers and the organization as a while with strategic directions and a data-oriented culture to remain competitive in the big data era.

"The Kolon-KAIST project described has played a critical role in helping the Big Data Team to achieve its mission. The KAIST team worked with members of the Big Data Team to identify problems in the distribution process; analyze a large amount of data to identify sale patterns; create a demand forecasting method; build an optimization algorithm to distribute fashion goods more effectively; and eventually deliver a solution for streamlining Kolon's packing and distribution (SAPD) processes.

"The KAIST team also worked with store managers and sales merchandisers to better understand issues at the store level, and then devoted significant effort to convincing them of the benefits of using the proposed SAPD method, which renders the distribution process more logical and effective. Previously, the distribution method varied, depending on who was making the decisions. Although a simple rule was in place to guide SAPD decisions, it was based on experience rather than fact-based reasoning. Since implementation of the proposed SAPD method, however, the entire process has become more logical and consistent. Also, it used to take several days to finalize SAPD decisions for a season, whereas now it takes several hours. In sum, the entire SAPD process has undergone significant improvement since the proposed method was implemented.

"Although the manuscript focuses on optimization modeling and algorithm development, the authors also discuss the significant work done on demand forecasting. Unfortunately, this component of their work cannot be explained in depth in the manuscript owing to the company's internal data policy and confidentiality agreements.

"Kolon Sport estimates that the proposed SAPD method contributed to the 8%–10% sales increase it saw in the fall and winter seasons of 2015 and improved the under- and over-distribution rates by 10%–15%. These results have been validated with various departments and management levels, and the method has received positive feedback.

"Throughout the project, the KAIST team helped Kolon Industries to better understand the business analytics concept and provided us with insight into how to create value from data and how to deliver value and results. Although value is not quantifiable, I believe that our new understanding of the business analytics concept will prove an important asset for Kolon Industries, helping us to remain competitive in the era of big data."

Shin Woong Sung is a PhD candidate in the Department of Industrial and Systems Engineering, Korea Advanced Institute of Science and Technology (KAIST). She holds a BS and an MS in the Department of Industrial and Systems Engineering from KAIST. Her research interests are in the application of operations research techniques to retail supply chain management, including such topics as assortment planning, distribution planning, and transshipment. She was awarded an Honorable Mention for SAS and INFORMS Analytics Society Student Analytical Scholar Competition in 2013.

Young Jae Jang received his PhD degree in mechanical engineering from Massachusetts Institute of Technology (MIT) in 2007 and a double MS degree in mechanical engineering and operations research from MIT in 2001. He received a BS degree in aerospace engineering in 1997 from Boston University. He is an associate professor in the Industrial and Systems Engineering Department at the Korea Advanced Institute of Science and Technology (KAIST), South Korea. He is also affiliated with the Cho Chun Shik Graduate School of Green Transportation at KAIST.

Jung Hoon Kim is a senior data analyst in the Data Insight team at Kolon Industries FnC Organization. He is responsible for supporting agile and accurate decision making on the fashion retail operations by providing insights drawn from many data analytics projects throughout the value chain of fashion industry. He received his MA in statistics from Yonsei University in South Korea. He has 13 years of experience in data analysis, focused on risk management and customer relationship management in finance, telecommunications, logistics, and fashion industries.

Juyeong Lee is a data scientist at SK Telecom, a leading telecommunication company in South Korea. He received a master's degree in industrial and systems engineering from Korea Advanced Institute of Science and Technology (KAIST). During his time at KAIST, he was actively involved in research relating to optimization modeling/algorithms and also large-scale machine learning techniques.