Interfaces
Vol. 40, No. 1, January–February 2010, pp. 71–84
ISSN 0092-2102 | EISSN 1526-551X | 10 | 4001 | 0071

informs®

DOI 10.1287/inte.1090.0472

THE FRANZ EDELMAN AWARD
Achievement in Operations Research

Zara Uses Operations Research to Reengineer Its Global Distribution Process

Felipe Caro
UCLA Anderson School of Management, University of California, Los Angeles, Los Angeles, California 90095,
fcaro@anderson.ucla.edu

Jérémie Gallien
MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142,
jgallien@mit.edu

Miguel Díaz, Javier García, José Manuel Corredoira, Marcos Montes
Zara, 15142 Arteixo, La Coruña, Spain {migueldm@inditex.com, javiergt@inditex.com, josemcc@inditex.com, marcosmv@inditex.com}

José Antonio Ramos
Carrefour, 28028 Madrid, Spain, joseantonio_ramos_calamonte@carrefour.com

Juan Correa
Dell Computers, Austin, Texas 78759, jcorrea@sloan.mit.edu

Overcoming significant technical and human difficulties, Zara recently deployed a new process that relies extensively on sophisticated operations research models to determine each inventory shipment it sends from its two central warehouses to its 1,500 stores worldwide. By taking a retail size-assortment view of a store's inventory, the model incorporates the link between stock levels and demand to select store replenishment quantities. Through a rigorous, controlled field experiment, we estimate that this new process has increased sales by 3–4 percent; this corresponds to estimated profits of approximately $233 million and $353 million in additional revenues for 2007 and 2008, respectively.

Key words: retailing; fast fashion; inventory management; field experiment; sizes distribution.

Introduction

With more than 1,500 stores in 68 countries and €6.26 billion in 2007 annual sales, Zara is one of the world's leading fashion retailers and the flagship chain of the Inditex Group. Because of its impressive growth in recent years, Zara has also become one of the most recognized apparel brands worldwide (Helm 2008). This success is widely attributed to its fast-fashion business model, which involves frequent in-season assortment changes and ever-trendy items offered in appealing store environments and at competitive prices.

To support this customer-value proposition, Zara has developed an innovative and highly responsive design, production, and distribution infrastructure that many press articles and case studies (e.g., Fraiman et al. 2002, Ghemawat and Nueno 2003) have described. In particular, Zara's supply chain involves two primary warehouses in Spain that periodically receive shipments of finished clothes from suppliers and ship merchandise directly to each Zara store worldwide twice a week (at the time of this writing, Zara had opened two additional warehouses in Madrid and León, Spain). This paper discusses the development, implementation, and impact of a new OR-based process; since 2006, Zara has used it daily to determine these shipment quantities, i.e., the "bloodstream" to its only sales channel. It constitutes a less technical and more practice-oriented companion paper to Caro and Gallien (2009).

Our work began in mid-2005 from a contact that Caro (University of California, Los Angeles) established with Ramos (then working at Zara). It soon evolved into a collaboration of researchers Caro and Gallien (Massachusetts Institute of Technology, MIT) with Zara employees Ramos, García, and Montes under the executive supervision and sponsorship of Zara's CFO, Miguel Díaz. Inditex senior information technology (IT) engineer Corredoira had overall responsibility for the system implementation; the project also leveraged the six-month internship of (then) MIT graduate student Correa hosted by Inditex, which initiated a corporate partnership with MIT's Leaders for Manufacturing (LFM) Program (Correa 2007).

In this paper, we discuss the managerial problem we addressed, the solution we developed, and the impact of our work, and we end with a summary of our contributions to OR practice and some concluding remarks. We emphasize that the authors performed all the financial impact estimates discussed in this paper. They did not engage the responsibility of the Inditex Group, which advises that any forward-looking statement is subject to risk and uncertainty and could thus differ from actual results.

The Problem

Managerial Challenge

Zara's innovative business model is powered by a continuous cycle that involves flows from stores to designers (market information and customer desires), designers to suppliers (production orders for new designs), suppliers to warehouses (deliveries of finished clothes), and warehouses to stores (outbound merchandise shipments). This last link is particularly critical; it constitutes the bloodstream of Zara's merchandise to its unique sales channel and directly affects Zara's global revenues.

Remaining sizes			Action
S	M	L	Keep on display
S	M	L	Keep on display
	M	L	Keep on display
S	M		Keep on display
	M		Keep on display
S		L	Move to backroom
S			Move to backroom
		L	Move to backroom

Figure 1: For an article offered in three sizes with M (medium) as the only major size, any combination without size M is moved to the backroom.

To distribute merchandise to its stores, Zara uses a supply chain that consists of two primary warehouses in Spain. They periodically receive shipments of finished clothes from suppliers and ship replenishment inventory directly to each Zara store worldwide twice a week. The key associated control challenge is to determine the exact number of units of each size (up to eight) of each article (up to 3,000 at any time) that should be included in each shipment to each of its more than 1,500 stores. This control problem is particularly challenging for the following reasons:

- As Figure 1 illustrates, most stores only display merchandise for sale when the set of available sizes is complete enough. Their intention is to achieve a balance between keeping inventory displayed to generate sales and mitigating the impact of missing sizes on brand perception; this is driven by the negative feeling that customers experience when they have identified a specific article they would like to buy, perhaps after spending much time searching a crowded store only to learn that their size is not available. More specifically, store managers tend to differentiate between major sizes (e.g., S, M, L) and minor sizes (e.g., XXS, XXL) when managing in-store inventory. When a store runs out of a major size for a specific article, store associates move all remaining inventory of that article from the display area to the backroom and replace it with a new article, effectively removing the incomplete article from customer sight. In contrast, they take no such action when the store runs out of a minor size. They might return a previously removed article back to the floor if the missing sizes can be shipped again from the warehouse; otherwise, it is either transferred to another store where the sizes are consolidated or it remains in the backroom until the store has a clearance sale. Note that this store inventory display policy introduces significant shipment interdependencies across sizes; it may be pointless to ship some units of a given size (e.g., XS) if it is not accompanied by enough units of a major size (e.g., M) to trigger display. It is worth noting that the removal rule described above is not prescribed by any formal policy imposed upon store managers; it constitutes an observation of common store behavior that we validated empirically (Caro and Gallien 2009).

- These shipment decisions must be determined in only a few hours after the relevant information (e.g., current store inventory, previous-day sales history) becomes available. Any further delay, in light of warehouse processing times and transportation schedules, would effectively delay the replenishment of stores by one full day (this replenishment response time is particularly important to Zara's business model, which explains why direct shipments to all stores are sent by truck and air every week).
- The number of associated shipment decisions reaches several million each week.
- The amount of relevant data (warehouse inventory, store inventory, and store sales history for each article) is enormous.
- The available warehouse inventory is often limited. This follows from Zara's business model, whereby the store life cycle of articles typically spans only a small fraction of a selling season (i.e., five to six weeks), so that the store assortment turns over much more frequently than the assortment of more traditional retailers does.

Legacy Process

Until 2006, Zara exclusively used the legacy process, as Figure 2(a) shows, to generate all its store-replenishment shipment decisions worldwide. As part of that process, store managers received weekly statements showing the subset of articles available in the central warehouse for which they could request shipments to their stores. Note that these weekly statements (dubbed "the offer") would thus effectively implement any high-level assortment decisions made by Zara's headquarters for each store. However, a statement would not mention the total quantity of inventory available in the warehouse for each article listed. After considering the inventory remaining in their respective stores, store managers then transmitted back requested shipment quantities

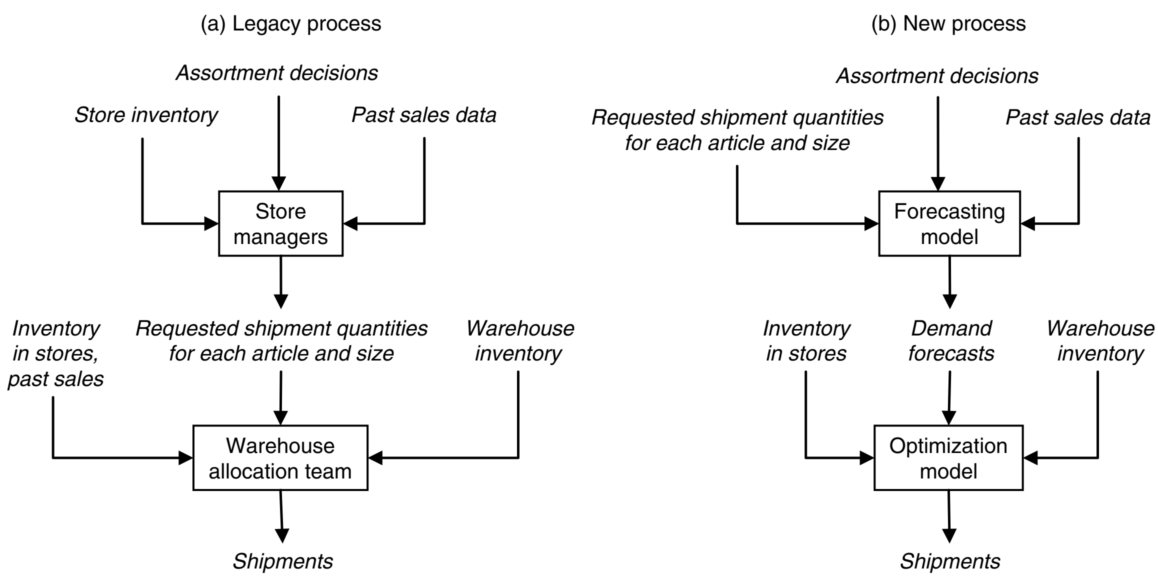

Figure 2: In contrast with the legacy process, the new process relies on formal forecasting and optimization models to determine weekly shipments to stores (Caro and Gallien 2009).

(possibly zero) for each size of each article. A team of employees at the warehouse then aggregated and reconciled the requests from all store managers by modifying (typically lowering) these requested shipment quantities so that the overall quantity shipped for each article and size was feasible in light of the remaining warehouse inventory.

For many years, the legacy process seems to have effectively supported the relatively small distribution network for which it had been designed originally. A key motivation for our project was Zara's realization in 2005 that the recent growth of its network to more than 1,000 stores might justify addressing several related improvement opportunities and ultimately designing a more scalable process. One issue centered on the incentives of store managers, whose compensation and career promotion prospects are driven significantly by the total sales achieved in their stores. We believe that this caused store managers to frequently request quantities exceeding their true needs, particularly when they suspected that the warehouse might not hold enough inventory of a top-selling article to satisfy all stores. In addition, store managers are responsible for many tasks beyond determining shipment quantities, including building, sustaining, and managing teams of several dozen sales associates in environments with high employee turnover; thus, they are subject to significant time pressures that compete with their involvement in determining merchandise replenishment requests. Finally, we also believe that the very large amount of data for which the warehouse allocation team was responsible for reviewing (i.e., shipments of several hundred articles offered in several sizes to more than 1,000 stores) created significant time pressures and made it challenging to balance inventory allocations manually across stores and articles in a way that would globally maximize sales.

The Solution

Figure 2(b) illustrates the structure of the new process we developed to help Zara compute its weekly store shipments. At a high level, it consists of using the shipment requests from store managers and past historical sales to build demand forecasts. It then uses the following as inputs to an optimization model: (1) these forecasts, (2) the inventory of each article and size remaining both in the warehouse and each store, and (3) the assortment decisions. The model has shipment quantities as its main decision variables and the maximization of global sales as its objective. In the sections below, we provide more details on our analytical work, the IT implementation, and the implementation and project management that were necessary to deploy this new process at Zara.

Analytical Development

As Figure 2(b) illustrates, the analytical development of the new process comprised two key steps: the *forecasting model* and the *optimization model.*

The forecasting model generates a prediction of the upcoming weekly demand for each size of each article in each store in Zara's network and essentially relies on the standard methodology of regression analysis. We refer the reader to Correa (2007) for further technical details and a complete definition. The most noteworthy feature of the implemented forecasting model may be its high-level structural form as a weighted linear combination of two primary sources of input data: (1) the objective and centralized time-series data of historical sales for the article, and (2) the subjective and decentralized shipment request by the store manager, converted into a sales prediction by considering store inventory and existing guidelines about target sales coverage. The weight conferred to the second input source by the least-square fitting procedure may thus be interpreted as a credibility measure (for forecasting purposes) of the store manager's input. Although the intuitive appeal of this interpretation helped Zara to communicate internally about the forecasting model, this ultimate functional form was only identified through extensive empirical testing of many possible predictive equations. In addition, our team viewed the forecasting model as a modular component of the process, in the sense that it can and should be improved relatively independently from the distribution optimization model that we discuss next. In particular, in addition to continuing to experiment with other functional forms and additional forecasting data sources (e.g., weather), Zara might in time consider introducing formal incentives for store managers to provide accurate forecasts, adding to its more traditional sales-related incentives.

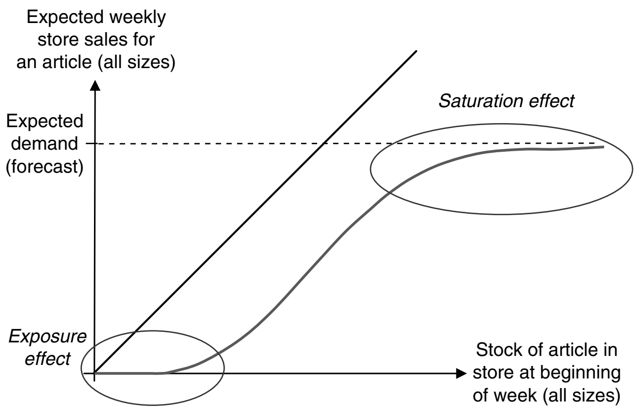

Figure 3: The inventory-to-sales function associated with each store captures both exposure and saturation effects.

The first step of our optimization-model development centered on constructing a predictive model for the expected upcoming weekly sales of all sizes of a given article in a given store, as a function of the relevant demand forecasts and the starting level of inventory in that store at the beginning of that week. We quickly started referring to this predictive model as the *inventory-to-sales* function; our motivation for this work was to ultimately support an optimization model allowing Zara to predict the impact of shipment decisions (affecting the inventory available at the beginning of each weekly replenishment period in each store) on global expected network sales. From a mathematical standpoint, this model resulted from the analysis of a stochastic model that considers the sales opportunities of different sizes of an article as independent Poisson processes. Caro and Gallien (2009) provide additional technical details; however, in this paper, we focus on the qualitative and managerial implications of this function (Figure 3).

A first feature of this inventory-to-sales function is that it considers the entire profile of inventory available across all sizes at the beginning of the week. In that sense, the representation in Figure 3, which suggests a one-dimensional, continuous-input variable (the sum of inventory across all sizes), is a simplification for exposition purposes of the actual underlying model (whose input is discrete and multidimensional). This function therefore jointly considers all sizes simultaneously (as opposed to each size independently); this is motivated by the inventory display policies discussed previously in the *Managerial Challenge* section and illustrated in Figure 1, and constitutes an important feature of our model: if the model considered different sizes independently, nothing would prevent it from determining that Zara should ship some units of a given size (e.g., XS) to a store that would not even display them for sale during the following week (let alone generate the sales predicted by the model), because the remaining inventory of a major size (e.g., M) of the article is not sufficient to trigger display at that store. In other words, the inventory-to-sales function we have constructed captures the dependencies across sizes that are introduced by Zara's policy of only displaying for sale the articles with a complete size profile. When considering the aggregate effect across all sizes of an article, this gives rise to the *exposure effect* (Figure 3), whereby the model will (correctly) predict that no sales will occur in a particular store if that store has insufficient available inventory (i.e., does not include enough different sizes). Incidentally, other firms that do not employ similar size-based display policies might also observe the same effect because of the self-advertising function of inventory in the store's sales display area (Smith and Achabal 1998). This feature is critical from an inventory distribution standpoint, because when all sales predictions for different stores are considered together, the exposure effect will push the optimization model to ship to those stores that can complete a full set of sizes (instead of scattering some limited inventory over the entire store network, which could leave many stores below the exposure threshold).

The second critical feature of the inventory-to-sales function is the *saturation effect*, which reflects the decrease of the marginal probability of sale as the store receives additional inventory units beyond a certain point (as highlighted in the upper right area of Figure 3). This is a much more classical feature for inventory distribution models (Zipkin 2000) that tend to balance the resulting shipments of merchandise across stores, so that each additional unit of available inventory is sent on the margin to the store at which it has the highest probability of selling. When considered jointly, the exposure and saturation effects enable the model to select stores that it will bring above the

$$\begin{array}{ll} \text{Maximize} & P * \textit{NetworkStoreSales} + K * \textit{FinalWarehouseStock} \\ \text{Subject to} & \textit{Shipments} \leq \textit{InitialWarehouseStock} \\ & \textit{NetworkStoreSales} \\ & \quad = \textit{Inv-to-Sales}\ (\textit{StoreInventory} + \textit{Shipments}) \\ & \textit{FinalWarehouseStock} = \textit{InitialWarehouseStock} - \textit{Shipments} \end{array}$$

Figure 4: Our single-period MIP model formulation involves a trade-off between sales from the current week and the value of inventory remaining in the warehouse.

exposure threshold, and properly balance inventory between them.

The last step of our model development work was to formulate a mixed-integer program (MIP) embedding piecewise-linear approximations of many independent inventory-to-sales functions, each associated with a store in the network, and allowing the computation of shipment quantities that maximize network-wide expected sales, subject to inventory availability constraints. The piecewise-linear approximations preserve the essential features of the inventory-to-sales function discussed above and make the model solvable with commercial MIP software. This model jointly computes shipment decisions of all sizes of a given article to all stores worldwide in any given week; however, for feasibility reasons, it ignores any dependencies between shipment decisions of different articles (i.e., Zara solves an instance of this optimization model for each article each week).

Figure 4 shows a high-level representation of the single-period MIP formulation, where *Inv-to-Sales* denotes the inventory-to-sales function described above; we refer the reader to the appendix and Caro and Gallien (2009) for technical details. One of the most original aspects of this formulation is arguably the use of a control parameter that we called the *aggressiveness factor* (denoted as K in Figure 4). In effect, this control is motivated by the fact that the model otherwise ignores the important issues of forecast uncertainty, the time horizon, and the opportunity cost of storage space at the warehouse. Because developing an optimization model that explicitly captures these issues presented significant analytical and data availability challenges, we enabled the user to affect the optimization-model outcome according to a procedure that requires some subjective input but is specifically designed to consider these issues. Specifically, our model's objective includes a standard first term that captures expected revenues for the article considered over the next week and replenishment cycle (denoted as $P * \textit{NetworkStoreSales}$ in Figure 4), and a second special term equal to the total inputted value for any inventory remaining in the warehouse after the shipments considered (denoted as $K * \textit{FinalWarehouseStock}$ in Figure 4). Although revenues in the first term are calculated from the known actual unit selling price P (assumed here to be constant across stores for expositional simplicity), remaining warehouse units are valued in the second term with the aggressiveness factor K, which the user provides and we can thus interpret as the unit value of articles left in the warehouse. A high value of K relative to the store selling price P results in "conservative" shipments, possibly appropriate shortly after a product introduction (when forecast uncertainty is high) or when the returns and transshipment costs associated with excessive inventory sent to low-selling stores might be particularly high. In contrast, a low relative value of K results in "aggressive" shipments, perhaps suitable when forecasts are deemed more reliable and (or) toward the end of the planned shelf life of an article, a time when freeing up some space in the warehouse for other articles is desirable.

IT Implementation

The IT implementation of the OR models described above was relatively challenging, in that it required establishing dynamic access to several large, live databases (store inventory, sales, and warehouse inventory) to compute, under very stringent time constraints, many decisions that are critical to the company's operation; in any typical week, Zara now solves approximately 15,000 instances of our large-scale MIP model to distribute six million units of stock valued at more than €120 million.

Coauthors Correa and Corredoira developed the core computational application using AMPL and ILOG CPLEX. It involves database queries into Inditex's legacy IBM AS/400 system, which feeds a specific database (SQL Server). From a hardware standpoint, this application is currently executed on two dedicated Dell servers (CPU 3.20 GHz, RAM 4 GB, HD 80 GB, OS Windows 2003 Server) located in Zara's IT headquarters in La Coruña, Spain; each handles computations for about half of Zara's stores. The

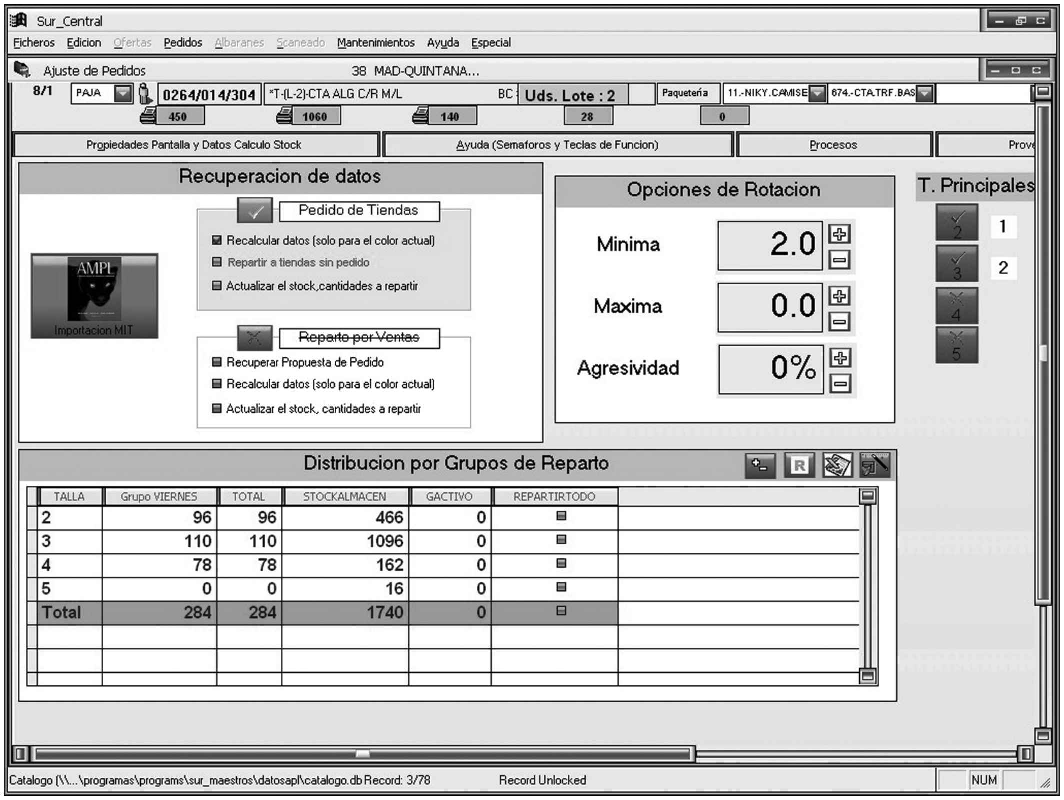

Figure 5: Warehouse allocation employees use a new client application to run the optimization engine, perform what-if scenario analyses, and visualize and modify the recommended shipments.

implementation also required the in-house development of a client application, which we distributed on the PCs of the warehouse allocation team's approximately 60 employees. This second application was developed using Visual FoxPro and provides an interface that allows these employees to request additional runs of the core computational application to perform what-if scenario analyses, visualize, and manually modify any output of the optimization model, and communicate their chosen solutions to the existing warehouse control systems; thus, they are effectively implementing the physical picking, sorting, packing, and truck-loading operations corresponding to the shipments determined by the warehouse allocation team using the decision support system. Figure 5 shows a screen snapshot of this client application. Finally, the data-communications infrastructure that supports these applications relies, to date, on Inditex's virtual private network (VPN); however, a more efficient node-based infrastructure is being deployed.

Implementation and Project Management

In 2005, when the first project-related discussions started, implementing OR models on a large scale at Zara to support core business decisions, which would affect the entire company's success, seemed like a daunting task. Zara is a fashion company with a culture that strongly favors human intuition, vision, and judgment (as opposed to analytical methods) for decision making. It has a history of success in showing that these subjective or nonquantitative approaches can pay off when applied to many of the key decisions it faces (e.g., design of clothes). OR awareness within Zara essentially did not exist when the authors started collaborating on this project. Bringing Zara to its current state, in which many employees and some key executives trust OR methods to determine key operational decisions, required a substantial investment in communication and education by all the authors, and careful thinking about the design and management of this implementation project.

In particular, we believe that some design choices were critical to our success in this environment. One was the role of the warehouse allocation team in the new process. We clearly established and communicated very early on that the forecasting and optimization models were not meant to replace that team, but rather to assist its members in performing their tasks more effectively. This orientation induced several technical choices, such as developing the distributed client application to allow the warehouse employees to perform what-if analysis scenarios defined, for example, by the aggressiveness factor K discussed in the *Analytical Development* section, and to manually and freely modify all the individual shipments computed by the optimization model before their physical execution. The academic project team members were initially uncomfortable with this last feature, because of methodological concerns linked to measuring impact for the optimization model output. However, we realized that this choice was essential to secure the warehouse allocation team's involvement and to allow us to benefit from the team's considerable knowledge of the managerial challenge at stake; this knowledge proved invaluable to us as we formulated the optimization model. In addition, we recognize now that these concerns were unfounded; after a live trial and debugging period of a few weeks, the warehouse allocation team members rarely (if ever) modified the optimization model's output directly (as opposed to changing the model's input and control parameters). In other words, allowing them to have total flexibility and control over the output was essential to develop their internal confidence in the model, although they ultimately rarely used this feature.

From an implementation standpoint, a second important process feature was our choice to leave the information interface with stores as it was in the legacy process. That is, the store manager input that the forecasting model required in the new process was the requested shipment quantities that the managers were already providing as part of the legacy process (the *Legacy Process* and *Analytical Development* sections provide background information and a description of how the forecasting model uses this information). Although directly requesting demand forecasts from store managers might improve overall forecast accuracy in the long run, the practical difficulties and resource requirements associated with managing such a significant operational change, which would affect a global network of more than 1,500 stores, prompted the team not to undertake this effort within the targeted project timeline. We note, however, that Zara might still leverage this opportunity to improve the new process in the future and that more general forecasting-procedure improvements can be performed in a modular way without affecting the optimization model's structure.

We believe that several aspects of managing this project were material success factors. The first was the project team's organizational structure. Specifically, our work began with collaboration between academic researchers (Caro and Gallien) and industry practitioners (Ramos, García, and Montes). Although such collaborations generally have the potential to be powerful and mutually enriching, they also present intrinsic challenges because of possible differences in objectives, incentives, and time horizons. In our case, we observed that the jointly supervised six-month internship at Zara of (then) graduate student Correa, who was enrolled in MIT's LFM program, provided an effective coordination mechanism and liaison between the two parts of the team. It also served as a catalyst for the model development work performed in preparation for that internship, which was to focus on implementing these models.

Another important catalyst was the live pilot experiment of the new process, which the entire team planned as a critical project requirement and was designed with three goals: (1) provide a convincing proof of feasibility for the new OR-based process, (2) help identify and act upon improvement opportunities for the new process before its full-scale deployment, and (3) support a quantitative impact assessment. Partly because the live pilot would affect the shipments of real merchandise to several hundred stores and would also involve a friendly competition with the legacy process (the *Impact* section provides details), it helped greatly to define priorities and focus the energies of all team members. It also helped to overcome the cultural barriers mentioned above and

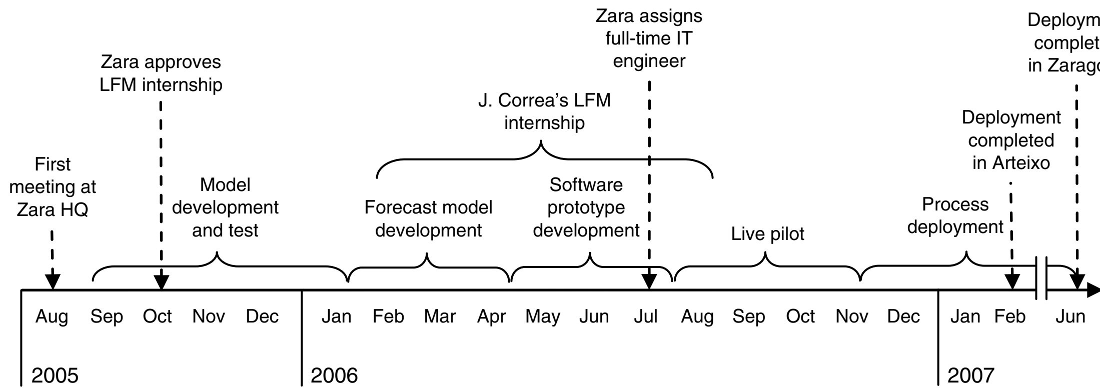

Figure 6: The project involved five main phases spanning almost two years.

generate buy-in from important stakeholders within the company when the team subsequently planned and executed the full-scale deployment of the new process.

The complete development and implementation cycle for this work spanned just under two years (Figure 6), and its corresponding total cost to Zara is estimated to range between $150,000 and $250,000, excluding the labor costs of all contributing employees not specifically hired for this project.

Impact

Measurement Methodology

The methodology to estimate the implementation impact involved a live pilot implementation experiment

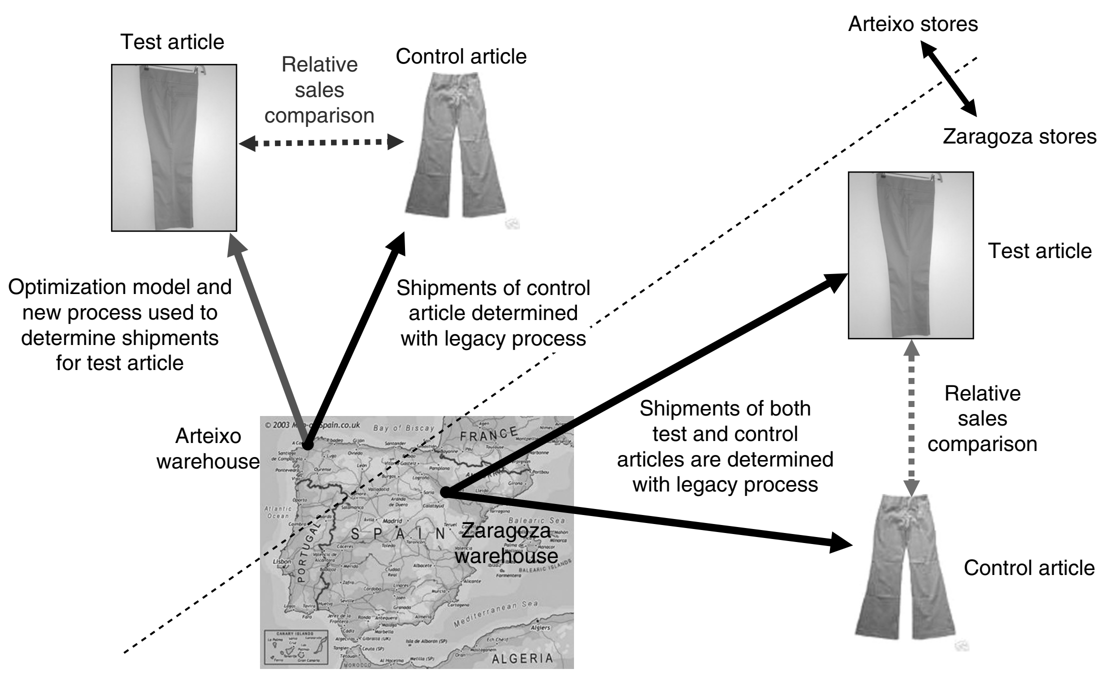

Figure 7: The experimental design of the live pilot involves two dimensions of control for each test article; the illustration shows 1 of the 10 test articles and its paired control article.

using two dimensions of control (Figure 7). Using the optimization model, we distributed a test group of 10 articles to all stores that exclusively receive shipments from Zara's first primary warehouse in Arteixo, Spain (this represents approximately half of all Zara stores worldwide). Simultaneously, we distributed a subset of 10 "twin" control articles, which we determined through a careful pairwise matching with the test group, to the same stores using the legacy manual process. This matching procedure enabled us to do a relative comparison between the test and the control group and thus supported estimating the specific impact of the new model and eliminated the impact of any other factors external to the model (which would affect both the test and control groups); this constitutes the first dimension of control. The second dimension of control exploits the fact that the other Zara stores worldwide receive only shipments from its second main warehouse in Zaragoza, Spain. For this second and relatively independent network of stores, we determined shipments of both test and control group articles using the legacy manual process. We selected this experimental design to enable us to estimate the error associated with the control-based impact-estimation methodology described above. For Zaragoza stores, we distributed both test and control group articles using the same procedure; any measured impact for this second half of the store network could only be attributed to estimation errors, as opposed to the use of our new OR-based process.

Financial Impact

Figure 8 summarizes the main results of the controlled live pilot experiment described above. These results are conclusive: The control-adjusted relative sales impact in Arteixo is positive for every article, with a mean across articles of 4.1 percent (median 4.2 percent); the corresponding estimation error calculated using data from Zaragoza is centered around zero (mean and median across articles are 0.7 and −0.6 percent, respectively). We also note here that the variability around the means that we see across different articles in Figure 8 is easily explained by the measurement noise introduced for each individual article by the pairwise matching procedure and by possible forecasting errors. Therefore, focusing on the averages of these measurements across articles, which are indicative of the model's overall impact on the entire range of Zara's product offerings, is appropriate. When (conservatively) subtracting the estimated average experimental error obtained from the Zaragoza data, these results show that the new OR-based process increases sales during the selling season by 3–4 percent, as seen in Figure 8. This impact on revenue is easily explained by the model's ability,

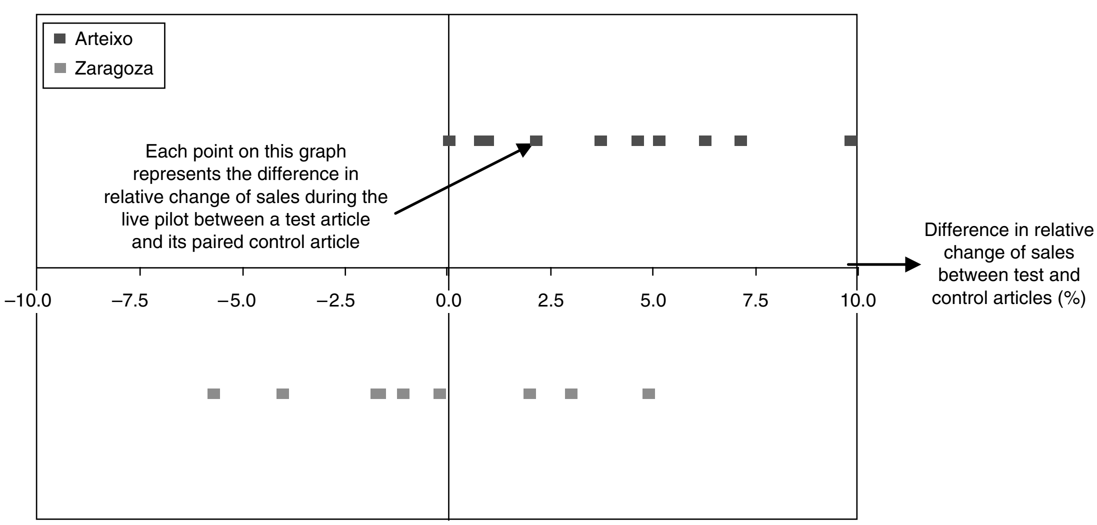

Figure 8: The results of the live pilot experiment suggest that the new process increases sales by approximately 3–4 percent (adapted from Caro and Gallien 2009).

relative to the legacy process, to determine that Zara should move excessive inventory away from low-selling stores where it is not needed, send it to high-performing stores in which it reduces the number of sales missed because of stock-outs, and ship all sizes of an article to each store in a concerted manner, likewise sending inventory of specific sizes only to stores in which it is likely to sell.

From a financial standpoint, if Zara had used the optimization model for all of 2007 and 2008, this relative sales increase would have implied approximately $310 million (2007) and $353 million (2008) in additional revenue or $37.2 million (2007) and $42.4 million (2008) in additional net income, with both measures of impact predicted to continue growing at a rate of 10 percent per annum in subsequent years. However, the model's full-scale deployment started in late 2006 and was completed in June of 2007; therefore, the estimated, actual realized impact of the model in 2007 is instead approximately $233 million in additional revenue or $28 million in additional net income (that is, 75 percent of the previous figures for 2007). The companion paper, Caro and Gallien (2009), presents some quantitative evidence that the model also reduced the transshipments between stores and increased the time spent by articles on display within their life cycle; however, the financial implications of these observations are harder to estimate.

Organizational Impact

Zara was able to maintain its warehouse inventory allocation team at its early 2007 staffing level of approximately 60 individuals worldwide, although it was initially planning on expanding that team because of its projected 10–12 percent sales growth per annum. More importantly, the optimization model has had a significant impact on the daily lives of these employees by enriching their professional responsibilities: all team members have become enthusiastic users of the new tool, gratefully seeing their responsibilities shift from repetitive manual data entry to exception handling, scenario analysis, and process improvement.

Cultural Impact

This project has also had some cultural impact at Zara, a company where many favor human intuition, vision, and judgment (as opposed to analytical methods) when making decisions. This is partly because the typical background of most fashion-industry employees is not quantitative and perhaps also because Zara owes much of its success to the unique intuition of its founder (Ghemawat and Nueno 2003). We doubt that Zara will ever use OR models to help with several of its key challenges, including anticipating volatile market trends, recruiting top designers, and creating fashionable clothes. Moreover, it is not clear to us that the company should. However, the work presented here did prompt a realization by many of Zara's key executives and employees that for processes involving large amounts of quantitative data, well-designed OR models will lead to better performance and more scalable operations. Because of this work, Zara has initiated two additional major OR implementation projects on purchasing and pricing. It is now also actively seeking to recruit graduates with strong OR backgrounds and to become a corporate partner with MIT's LFM Program. In addition, the Inditex Group is planning to deploy the OR-based inventory distribution process described in this paper in some of its other retail chains, such as Massimo Dutti.

Summary of Contributions to OR Practice

To the best of our knowledge, this work constitutes the first reported application of OR to the fast-fashion retail-business strategy, as adopted by companies that include Zara, H&M, and Mango. Under this strategy, the life cycle of merchandise in these stores spans only a small fraction of a selling season (e.g., five to six weeks for Zara). Therefore, their store assortments turn over much more frequently than those of traditional retailers, customers find frequent store visits to be more appealing, and the warehouse inventory available for distribution to these stores tends to be scarcer than for traditional retailers, making the inventory allocation problem more difficult.

This work also appears to constitute the first described implementation of an inventory distribution model for an apparel retailer (fast-fashion or other), which specifically captures the dependencies across sizes introduced by store inventory display policies. Specifically, many retailers, including Zara, remove some articles from display and put them into the store's backroom whenever the combination of their sizes still available is not complete enough. Although many retailers replenish their stores for each size independently (sometimes leading to useless shipments from a sales standpoint if major sizes are missing) or ship predetermined size bundles (making it difficult to balance the inventory available across sizes in stores), our model considers the entire inventory profile of all sizes offered and computes coordinated shipment quantities for all these sizes simultaneously; the *Analytical Development* section provides more detail.

Finally, we emphasize that although our impact-measurement methodology based on a controlled experiment is fairly common in other disciplines (e.g., medicine and social sciences), its application to OR practice is noteworthy, because the impact of publicly described OR practice work is predominantly estimated through more questionable "before versus after" comparisons that completely ignore the many other factors that are not related to OR, but could also be affecting differences in performance observed in the "after" period.

Concluding Remarks

Because of Zara's openness to the academic publication of this work, the OR models presented might also impact other companies. Although the inventory distribution model discussed above would clearly require some IT implementation and adaptation work, it seems particularly applicable to the many apparel retailers facing the challenge of coordinating shipment decisions across different sizes because of store display policies. A simpler version of this model, which would not capture sizes but would capture the notion of required store-exposure inventory, also seems applicable to most of the retail industry beyond the apparel segment and would likely constitute a significant improvement relative to the pervasive non-OR-based methods and simple heuristics that many firms use when distributing scarce inventory to a network of stores (e.g., proportional rationing).

Our work has also been very formative for its academic authors (both were junior faculty at major research universities during this project), and it has significantly affected their research and teaching activities. In particular, this OR implementation has already been the object of several masters, executive education, and MBA course sessions taught at UCLA, MIT, and Columbia University. All team members also learned many lessons, several of which were initially counterintuitive, about implementing OR. For example, from an academic standpoint, designing and studying OR models that capture all the relevant key input data (including forecast uncertainty and storage opportunity cost in our case) might seem important. In practice, however, our experience suggests that implementing such "comprehensive" models to control large-scale core processes in time-sensitive environments could present insurmountable IT and data-availability challenges; thus, a hybrid approach combining a simpler but robust optimization model with carefully selected user inputs might be more effective.

In closing, we note that this paper presents a successful and high-impact application of OR techniques in a fashion-retailing environment, which is both highly visible to the general public and not currently perceived as a traditional application area for this discipline. We are hopeful that our work will contribute to increasing OR awareness and help improve perceptions of its applicability.

Appendix

Single-Store Inventory-to-Sales Model

Consider an article offered in a set of sizes $S = S^+ \cup S^-$, where S^+ denotes the major sizes (e.g., $\{S, M, L\}$) and S^- the minor sizes (e.g., $\{XS, XL\}$). Sale opportunities for each size $s \in S$ are assumed to be independent across sizes and follow a Poisson process with rate λ_s and cumulative counting measure $\{N_s(t), t \geq 0\}$, where t denotes the time elapsed since the last replenishment (i.e., $N_s(t)$ is the random number of sale opportunities for size s that occurred between 0 and t). Let q_s represent the inventory level of size s immediately after replenishment at time 0; the virtual stockout time $\tau_s(q_s)$ can be defined for every size $s \in S$ as

$$\tau_s(q_s) \stackrel{\Delta}{=} \inf\{t \geq 0\colon N_s(t) = q_s\}.$$

Likewise, the earliest time at which one of the major sizes runs out, from an initial profile $\mathbf{q}$ of inventory across sizes and assuming no replenishment occurs, can be expressed as

$$\tau_{S^+}(\mathbf{q}) \stackrel{\Delta}{=} \min_{s \in S^+} \tau_s(q_s).$$

As we described above, all inventory is removed from customer view as soon as one of the major sizes runs out at any point between successive replenishments. Under that policy, the (random) total number of sales in a replenishment period can be expressed as

$$G(\mathbf{q}) \triangleq \sum_{s\in S^+} N_s(\tau_{S^+} \wedge T) + \sum_{s\in S^-} N_s(\tau_{S^+\cup\{s\}} \wedge T),$$

where $T > 0$ denotes the time between consecutive replenishments (one week for Zara) and $a \wedge b \triangleq \min(a, b)$. Applying Doob's optional sampling theorem, the expectation $g(\mathbf{q}) \triangleq E[G(\mathbf{q})]$ can be expressed as

$$g(\mathbf{q}) = \lambda_{S^+}E[\tau_{S^+} \wedge T] + \sum_{s\in S^-} \lambda_s E[\tau_{S^+\cup\{s\}} \wedge T],$$

$$\text{where } \lambda_{S^+} \triangleq \sum_{s\in S^+} \lambda_s.$$

For any subset of sizes $D \subset S$, we apply Jensen's approximation, $E[\tau_{S^+} \wedge T] \approx \min_{s\in D} E[\tau_s \wedge T]$, and observe that

$$E[\tau_s \wedge T] = \sum_{k=1..qs} \gamma(k, \lambda_s T)/\lambda_s \Gamma(k)$$
$$= \min_{i\in N}\{a_i(\lambda_s)(q_s - i) + b_i(\lambda_s)\},$$

where $a_k(\lambda_s) \triangleq \gamma(k, \lambda_s T)/\lambda_s\Gamma(k)$ and $a_\infty(\lambda_s) \triangleq 0, b_i(\lambda_s) \triangleq \sum_{k=1..i-1} a_k(\lambda_s)$ for $i \geq 1, b_0(\lambda_s) \triangleq 0$, and $b_\infty(\lambda_s) \triangleq T$, and Γ and γ are the Gamma function and the lower incomplete Gamma function, respectively. Our next approximation consists of computing only the minimum on the right side of the previous equality over the small finite subsets $N(\lambda_s)$ defined as

$$N(\lambda_s) \triangleq \{i \in N \cup \{\infty\}\colon b_i(\lambda_s) \approx 0, 0.3T, 0.6T, 0.8T, 0.9T, T\},$$

which are straightforward to compute numerically. Appropriate substitutions yield the final expression:

$$g(\mathbf{q}) \approx \lambda_{S^+} \min_{s\in S^+,\, i\in N(\lambda s)} \{a_i(\lambda_s)(q_s - i) + b_i(\lambda_s)\}$$
$$+ \sum_{s\in S^-} \lambda_s \min_{s'\in S^+\cup\{s\},\, i\in N(\lambda s')} \{a_i(\lambda_{s'})(q_{s'} - i) + b_i(\lambda_{s'})\}.$$

Caro and Gallien (2009) provide a complete discussion.

Network Sales-Optimization Model

Input Data

Set of sizes: $S = S^+ \cup S^-$ partitioned into major sizes S^+ and regular sizes S^- (index s);

Set of stores: J (index j);

W_s: Inventory of size s available in the warehouse;

I_{sj}: Inventory of size s available in store j;

P_j: Selling price in store j;

K: Aggressiveness factor (value of inventory remaining in the warehouse after the current shipments);

λ_{sj}: Demand rate for size s in store j; and

$N(\lambda_{sj})$: Approximation set for size s in the inventory-to-sales function approximation for store j.

Decision Variables

$x_{sj} \in N$: Shipment quantity of each size $s \in S$ to each store $j \in J$ for the current replenishment period;

z_j: The approximate expected sales across all sizes in each store j for the current period under consideration;

y_j: Secondary variables representing the term

$$\min_{s\in S^+,\, i\in N(\lambda sj)} \{a_i(\lambda_{s'j})(I_{s'j} + x_{s'j} - i) + b_i(\lambda_{s'j})\}; \quad \text{and}$$

v_{sj}: Secondary variables representing the term

$$\min_{s'\in S^+\cup\{s\},\, i\in N(\lambda s'j)} \{a_i(\lambda_{s'j})(I_{s'j} + x_{s'j} - i) + b_i(\lambda_{s'j})\}.$$

Objective

Max $\sum_{j\in J} P_j z_j + K(\sum_{s\in S}(W_s - \sum_{j\in J} x_{sj}))$.

Constraints

$\sum_{j\in J} x_{sj} \leq W_s$ for all $s \in S$ (warehouse inventory availability constraint);

$z_j \leq (\sum_{s\in S^+} \lambda_{sj})y_j + \sum_{s\in S^-} \lambda_{sj} v_{sj}$ for all $j \in J$ (primary inventory-to-sales function implementation constraint);

$y_j \leq a_i(\lambda_{sj})(I_{sj} + x_{sj} - i) + b_i(\lambda_{sj})$ for all $j \in J$, $s \in S^+$, and $i \in N(\lambda_{sj})$ (secondary inventory-to-sales function implementation constraint);

$v_{sj} \leq \{a_i(\lambda_{sj})(I_{sj} + x_{sj} - i) + b_i(\lambda_{sj})\}$ for all $j \in J$, $s \in S^-$, $i \in N(\lambda_{sj})$ (secondary inventory-to-sales function implementation constraint);

$v_{sj} \leq y_j$ for all $j \in J$, $s \in S^-$ (secondary inventory-to-sales function implementation constraint); and

$x_{sj} \in N$; $z_j, y_j \geq 0$; $v_{sj} \geq 0$ (nonnegativity and integer constraints).

Acknowledgments

We thank the Zara Distribution Team—in particular, Juan Quintela, Jesús González, and Iván Escudero; the Inditex Communications Department—in particular, Jesús Echevarria and Amaya Guillermo; Anthony Placet from the MIT Sloan Office of Communications; Donald Rosenfield and Jonathan Griffith from the MIT Leaders For Manufacturing Program; all the volunteers who made the 2009 Franz Edelman Competition possible, and in particular Coaches Layek Abdel-Malek, Ananth Iyer, Yoshi Ikura, and Chair Srinivas Bollapragada; the 2009 Franz Edelman Competition sponsors, Georgia Institute of Technology, Procter & Gamble, and SAS. F. Caro and J. Gallien would also like to thank faculty colleagues at the UCLA Anderson School of Management and MIT Sloan School of Management.

References

Caro, F., J. Gallien. 2009. Inventory management of a fast-fashion retail network. *Oper. Res.*, ePub ahead of print August 12, http://or.journal.informs.org/cgi/content/abstract/opre.1090.0698v1.

Correa, J. 2007. Optimization of a fast-response distribution network. Unpublished master's thesis, Massachusetts Institute of Technology, Cambridge.

Fraiman, N., M. Singh, L. Arrington, C. Paris. 2002. Zara. Columbia Business School case, Columbia University, New York.

Ghemawat, P., J. L. Nueno. 2003. ZARA: Fast fashion. Harvard Business School Multimedia Case 9-703-416, Harvard University, Boston.

Helm, B. 2008. Best global brands. *BusinessWeek* (September 18) 52.

Smith, S. A., D. D. Achabal. 1998. Clearance pricing and inventory policies for retail chains. *Management Sci.* **44**(3) 285–300.

Zipkin, P. H. 2000. *Foundations of Inventory Management*. McGraw-Hill, New York.