# Zara Uses Operations Research to Reengineer

# Its Global Distribution Process

Felipe Caro

UCLA Anderson School of Management, Los Angeles, California 90095

fcaro@anderson.ucla.edu


Jérémie Gallien

MIT Sloan School of Management, Cambridge, Massachusetts 02142

jgallien@mit.edu


Miguel Díaz Miranda, Javier García Torralbo, Jose Manuel Corredoira Corras, Marcos Montes

Vazquez

Zara, Edificio Inditex, Avda de la Diputación, 15142 Arteixo, A Coruña, Spain

migueldm@inditex.com, javiergt@inditex.com, josemcc@inditex.com, marcosmv@inditex.com


José Antonio Ramos Calamonte

Prescia 5, 8izq 28028, Madrid, Spain

joseantonio_ramos_calamonte@carrefour.com


Juan Correa

Dell Computers, 10804 Catthorn Cove, Austin, TX 78759

jcorrea@sloan.mit.edu

August 3, 2009

**Abstract:** Overcoming significant technical and human difficulties, Zara recently deployed a new process relying extensively on sophisticated Operations Research models to determine every single shipment of inventory it sends from its two central warehouses to its 1,500 stores worldwide. By taking a retail size assortment view of the store inventory, the model incorporates the link between stock levels and demand while choosing store replenishment quantities. Through a rigorous controlled field experiment, the authors estimate that this new process has increased sales by 3-4%, corresponding to an estimated realized impact of approximately $233M and $353M in additional revenues for 2007 and 2008, respectively.

With more than 1,500 stores in 68 countries and € 6.26B in annual sales (2007), Zara is the flagship chain of the Inditex Group and one of the world's leading fashion retailer. Due to its impressive growth in recent years, it 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 at competitive prices.

To support this customer value proposition, Zara has developed an innovative and highly responsive design, production and distribution infrastructure which has been described in many press articles and case studies (e.g., Fraiman et al. 2002, Ghemawat and Nueno 2003). In particular, Zara's supply chain involves two primary warehouses located in Spain which periodically receive shipments of finished clothes from suppliers, and ship merchandise directly to every Zara store in the world twice a week (at the time of writing, Zara had opened two other warehouses in Madrid and León, Spain). This paper discusses the development, implementation and impact of a new process based on Operations Research and used by Zara on a daily basis since 2006 in order to determine these shipment quantities, i.e., the "blood stream" to its only sales channel. It constitutes a less technical and more practice-oriented companion paper to Caro and Gallien (2008).

The work presented here initiated in mid-2005 from a contact that Caro (UCLA) established with Ramos (working at Zara at the time). Shortly after, it evolved into a

collaboration of researchers Caro and Gallien (MIT) with Zara employees Ramos, García and Montes, under the executive supervision and sponsorship of Zara's CFO Miguel Díaz. The system implementation, performed under the overall responsibility of Inditex senior IT engineer Corredoira, also leveraged the six-month internship of (then) MIT graduate student Correa hosted by Inditex, which initiated its corporate partnership with MIT's Leaders For Manufacturing Program (Correa 2007).

In the following we discuss in more details the managerial problem we addressed, the solution we developed, and the impact of our work. This paper ends with a summary of our contributions to O.R. practice and some concluding remarks. We emphasize that all the financial impact estimations provided in this paper were performed by the authors. They do 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 involving flows from stores to designers (market information/customer desires), from designers to suppliers (production orders for new designs), from suppliers to warehouses (deliveries of finished clothes), and from warehouses to stores (outbound merchandise shipments). This last link is particularly critical, because it constitutes the "blood stream" of Zara's merchandise to its unique sales channel, and therefore directly affects its global revenues.

To distribute merchandise to its stores, Zara uses a supply chain consisting of two primary warehouses located in Spain which periodically receive shipments of finished clothes from suppliers, and ship replenishment inventory directly to every Zara store in the world twice a week. The key associated control challenge is to determine the exact number of units of each size (up to 8) of each article (up to 3,000 at any time) that should be included in each shipment to each store (more than 1,500). In addition to being critical as mentioned above, this control problem is particularly challenging, for the following reasons:

- As illustrated by Figure 1, most stores will only display merchandise for sale when the set of available sizes is complete enough – this is meant to achieve a balance between keeping

3

inventory displayed to generate sales and mitigating the impact of missing sizes on brand perception (driven by the negative experience of customers who have identified a specific article they would like to buy, perhaps after spending much time searching a crowded store, only to find out their size is no longer 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 the store runs out of one of the major sizes for a specific article, store associates then move all remaining inventory of that article from the display area to the backroom and replace it with a new article, thus effectively removing the incomplete article from customers' sight. In contrast, no such action is taken when the store runs out of one of the minor sizes. The incomplete article that was removed may return to the floor if missing sizes can be shipped again from the warehouse; otherwise, it is either transferred to another store where the sizes are consolidated, or remains in the backroom until clearance sales. 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 just described is not prescribed by any formal policy imposed upon store managers, rather it constitutes an observation of common store behavior that we validated empirically (see Caro and Gallien 2008).

- These shipment decisions must be determined in just a few hours after the relevant information (latest store inventory, sales history of the previous day) becomes available. This is because any further delay, in light of the 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 millions every week;

- The amount of relevant data (warehouse inventory, store inventory and store sales history for each article) is also 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., 4-5 weeks), so that the store assortment turns over much more frequently than with more traditional retailers;

| Remaining sizes | | | Action |
|---|---|---|---|
| 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.**

**Legacy Process**

Until 2006, Zara exclusively used the legacy process illustrated in Figure 2 (a) in order to generate all its store replenishment shipment decisions worldwide. As part of that process, each store manager would receive a weekly statement of the subset of articles available in the central warehouse for which he/she may request a shipment to his/her store. Note that this weekly statement (dubbed "the offer") would thus effectively implement any high-level assortment decision made by Zara's headquarters for that particular store. However, it 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 would then transmit back requested shipment quantities (possibly zero) for every size of every one of those articles. A team of employees at the warehouse would then aggregate and reconcile 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.

**(a) Legacy Process**                     **(b) New Process**
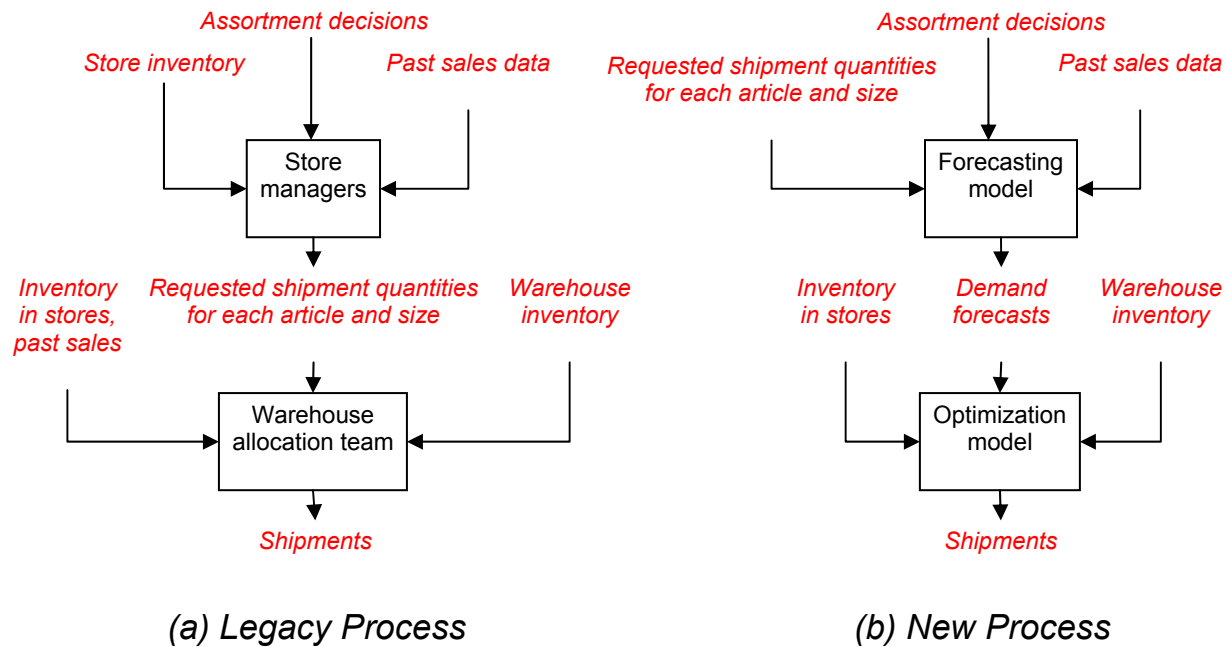
**Figure 2:** **In contrast with the legacy process, the new process relies on formal forecasting and optimization models to determine weekly shipments to stores (from Caro and Gallien 2008).**

The legacy process just described seems to have supported well and for many years the relatively small distribution network for which it had been originally designed. A key motivation for our project however was Zara's realization in 2005 that the recent growth of its network to more than a thousand stores might justify addressing several related improvement opportunities, and ultimately design a more scalable process. A first issue centered on the incentives of store managers, whose compensation and career promotion prospects are driven to a significant degree by the total sales achieved in their stores. We believe that as a consequence store managers frequently requested quantities exceeding their true needs, particularly when suspecting that the warehouse may not hold enough inventory of a top-selling article to satisfy all stores. Another issue is that store managers are responsible for a large set of tasks beyond determining shipment quantities, including building, sustaining and managing a team of several dozen sales associates in environments with high employee turnover, and are thus subject to important time pressures competing with their involvement in determining merchandise replenishment requests. Finally, we also believe that the very large

6

amount of data that the warehouse allocation team was responsible for reviewing (shipments of several hundred articles offered in several sizes to more than a thousand stores) also created significant time pressures which made it challenging to balance inventory allocations manually across stores and articles in a way that would globally maximize sales.

## The Solution

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

### Analytical Development

As seen in Figure 2 (b), the analytical development of the new process comprised two key steps: the *forecasting model* and the *optimization model*, which we now discuss in more details.

The forecasting model generates a prediction of the upcoming weekly demand for every size of every article in every store of Zara's network, and essentially relies on the standard methodology of regression analysis. While 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: (i) the objective and centralized time-series data of historical sales for the article, and (ii) the subjective and decentralized shipment request by the store manager, converted into a sales prediction by taking into account store inventory and existing guidelines about target sales coverage. The weight conferred to that 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. While the intuitive appeal of this interpretation helped

7

internal communications related to the forecasting model, an important remark is that this ultimate functional form was only identified through extensive empirical testing of many possible predictive equations. Another remark is that 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 to be discussed next. In particular, besides continuing to experiment with other functional forms and additional forecasting data sources (e.g., weather), Zara may in the long run consider introducing formal incentives for store managers to provide accurate forecasts, in addition to its more traditional sales-related incentives.

The first step of our optimization model development centered on the construction of 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 – the motivation for this work was to ultimately support an optimization model allowing to predict the impact of shipment decisions (affecting the inventory available at the beginning of every 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 considering the sales opportunities of different sizes of an article as independent Poisson processes. We refer the reader to Caro and Gallien (2008) for further technical details however, and focus here on the qualitative and managerial implications of this function, which are illustrated in Figure 3.
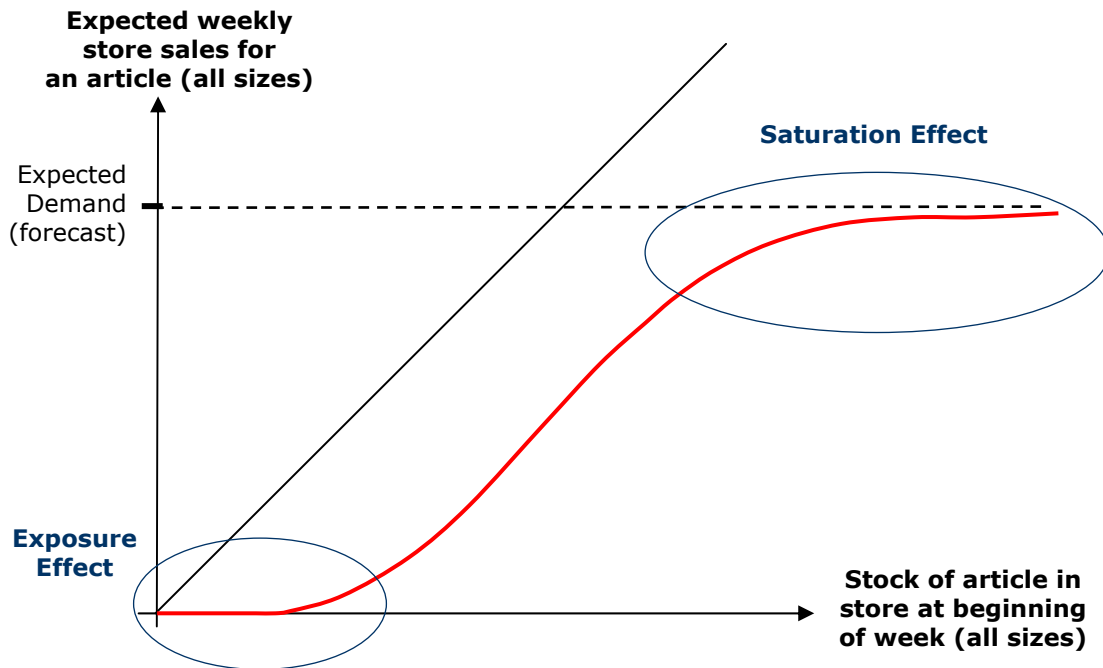
**Figure 3:** The inventory-to-sales function associated with each store captures both exposure and saturation effects.

A first feature of this inventory-to-sales function is that it takes into account the entire profile of inventory available across all sizes at the beginning of the week. In that sense, the representation seen 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 multi-dimensional). The fact that this function therefore jointly considers all sizes simultaneously (as opposed to each size independently) is motivated by the inventory display policies discussed in the Managerial Challenge section above and illustrated in Figure 1, and constitutes an important feature of our model: if different sizes were considered independently, nothing would prevent the model from shipping some units of a given size (e.g., XS) which would not even be displayed for sale in 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 same article at that store is not enough to trigger display. 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 sales the articles with a size profile that is complete enough. When considering the aggregate effect across all sizes of an

9

article, this gives rise to the *exposure effect* highlighted in Figure 3, whereby the model will (correctly) predict that no sales will occur in a particular store if the inventory available there is insufficient (i.e., does not include enough different sizes). Incidentally, other firms which do not employ similar size-based display policies may also observe the same effect due to the self-advertising function of inventory in the sales display area of a store, as discussed for example in Smith and Achabal (1998). In any case, this feature is critical from an inventory distribution standpoint, because when all the sales prediction for different stores are considered together, the exposure effect will push the optimization model to ship to those stores where it 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 units of inventory beyond a certain point (see highlighted area on the upper right of Figure 3). This is a much more classical feature for inventory distribution models (see e.g., Zipkin 2000), which tends 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 where it has the highest probability of selling. When considered jointly, the exposure and saturation effects just discussed enable the model to ship selectively to stores that it will bring above the exposure threshold, and properly balance inventory between them.

Finally, the last step of our model development work was to formulate a mixed-integer (MIP) program embedding piece-wise linear approximations of many independent inventory-to-sales functions each associated to a store in the network, and allowing to compute shipment quantities maximizing network-wide expected sales, subject to inventory availability constraints. The piece-wise linear approximations preserve the essential features of the inventory-to-sales function discussed above and made the model solvable with a 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 different article every week).

**Maximize**   P*NetworkStoreSales + K*FinalWarehouseStock

**Subject to**

Shipments <= InitialWarehouseStock

NetworkStoreSales = *Inv-to-Sales*(StoreInventory + Shipments)

FinalWarehouseStock = InitialWarehouseStock – Shipments

**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.**

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 (2008) for technical details. One of the most original aspects of this formulation is arguably the use of a control parameter we called the *aggressiveness factor* (denoted 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 the development of an optimization model explicitly capturing these issues presented significant analytical and data availability challenges, we instead enabled the user to affect the optimization model outcome according to a procedure that requires some subjective input but is specifically designed to take these issues into account. Specifically, our model's objective includes a standard first term capturing expected revenues for the article considered over the next week/replenishment cycle (this is the term P*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 K*FinalWarehouseStock in Figure 4). While 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 provided by the user, which can thus be interpreted 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 may 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 when it is desirable to free up some space in the warehouse for other articles.
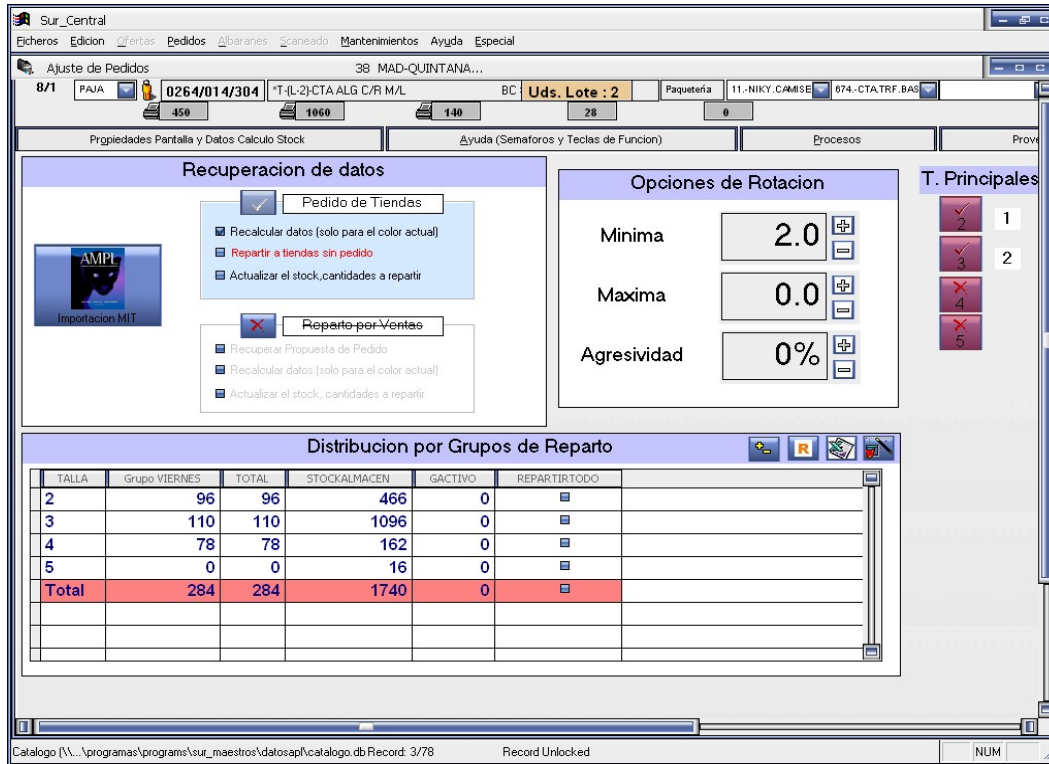


**Figure 5:** **Warehouse allocation employees use a new client application to run the optimization engine, perform what-if scenario analysis and visualize/modify the recommended shipments.**

**IT Implementation**

The IT implementation of the O.R. models described above was relatively challenging, in that it required establishing a dynamic access to several large live databases (store inventory and sales, warehouse inventory) in order to compute under very stringent time constraints a large quantity of 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 € 120M.

The core computational application was developed by co-authors Correa and Corredoira with AMPL and ILOG CPLEX. It involves database queries into Inditex's legacy IBM AS/400 system feeding a specific database (SQL Server). From a hardware standpoint, this application is currently executed on two dedicated Dell servers (CPU 3.20GHz, RAM 4GB, HD 80GB, OS Windows 2003 Server) located in the company's IT headquarters in La Coruña, each handling computations for about one half of the stores. The implementation also required the in-house development of a client application distributed on the PCs of the approximately 60 employees of the warehouse allocation team. This second application was developed with Visual FoxPro and provides an interface allowing these employees to request additional runs of the core computational application in order to perform what-if scenario analysis; visualize and manually modify any output of the optimization model; and finally communicate their chosen solution to the existing warehouse control systems 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 just described – see Figure 5 for a screen snapshot of this client application. Finally, the data communications infrastructure supporting these applications relies to date on Inditex's standard virtual private network (VPN), however a more efficient node-based infrastructure is under deployment.

**Implementation and Project Management**

Back in 2005 when the first discussions related to this project started, implementing O.R. models on a large scale at Zara to support core business decisions affecting the entire company's success seemed like a daunting task. This is because Zara is a fashion company with a culture strongly favoring human intuition, vision and judgment (as opposed to analytical methods) for decision-making purposes, and a history of success showing that these subjective or non-quantitative approaches can pay off when applied to many of the key decisions it faces (e.g., design of clothes). In fact, the awareness of O.R. within the company was essentially non-existent when the authors started collaborating on this project. The current state, whereby O.R. methods are trusted by some key executives and many regular employees to determine key operational decisions has required a substantial investment in communication and education

by all of the authors, as well as careful thinking about the design and management of this implementation project.

In particular, we believe that a couple of design choices were critical to our success in this environment. A first one was the role of the warehouse allocation team in the new process: It was 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 task more effectively. This orientation induced several technical choices, such as the development of a distributed client application allowing the warehouse employees to perform what-if analysis scenario defined for example by the aggressiveness factor $K$ discussed in the Analytical Development section, and to manually and freely modify all of the individual shipments computed by the optimization model before their physical execution. This last feature left the academic project team members uncomfortable at first, because of methodological concerns linked to the measurement of impact for the optimization model output. In actuality however, this choice was essential to secure the involvement of the warehouse allocation team, and to benefit from their considerable knowledge of the managerial challenge at stake, which proved invaluable when formulating the optimization model. In addition, it turns out that these concerns were not even founded, because after a live trial and debugging period of a couple of weeks, the warehouse allocation team members ended up almost never modifying the output of the optimization model directly (as opposed to changing the model's input and control parameters). In other words, allowing the warehouse team to have total flexibility over the output was essential to develop internal confidence in the model, though at the end this feature was hardly used.

A second important process feature from an implementation standpoint was the choice to leave the information interface with stores completely unchanged relative to the legacy process. That is, the input from store managers required by the forecasting model in the new process was in fact the requested shipment quantities that they were already providing as part of the legacy process (see Legacy Process and Analytical Development sections for background information and a description of how the forecasting model uses this information). While directly requesting demand forecasts from store managers may improve overall forecast accuracy in the long run, the practical difficulties and resource requirements associated with

14

managing a significant operational change affecting a global network of more than 1,500 stores prompted the team to not undertake this effort within the targeted project timeline. We note however that Zara may still leverage this opportunity to further improve the new process at any point in the future, and that more generally improvements of the forecasting procedure can be performed in a modular way without affecting the structure of the optimization model.

There are also several aspects of the management of this project which we believe were material success factors. A first important feature was the project team's organizational structure. Specifically, the work presented here initiated from a collaboration between academic researchers (Caro, Gallien) and industry practitioners (Ramos, García, Montes). While 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 enrolled in MIT's Leaders For Manufacturing (LFM) program provided an effective coordination mechanism and liaison between those two parts of the team. It also served as a catalyst for the model development work performed in preparation of that internship, which was to focus on the implementation of these models.
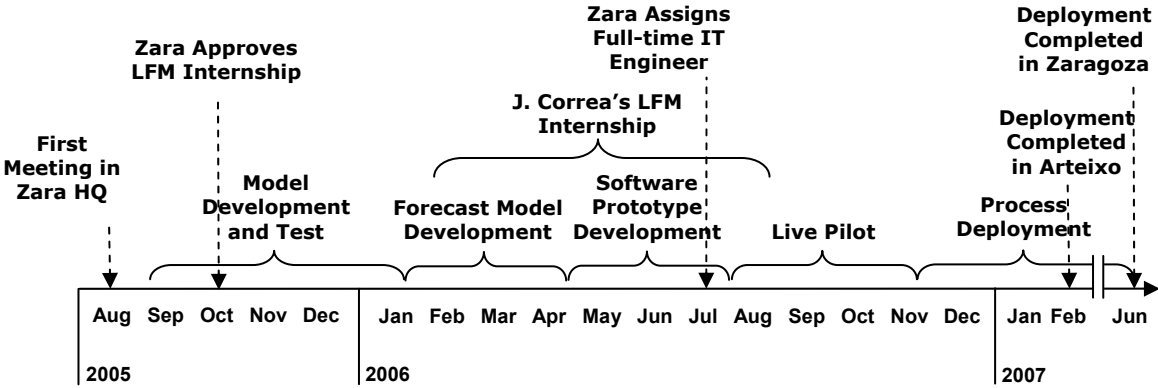


**Figure 6:    The project involved five main phases spanning almost two years.**

Another important catalyst was the live pilot experiment of the new process which the entire team planned early on as a critical project requirement, and was designed with three goals: (i) provide a convincing proof of feasibility for the new OR-based process; (ii) help identify and act upon improvement opportunities for the new process before its full-scale deployment;

15

and (iii) support a quantitative assessment of impact. In part because this live pilot was to affect the shipments of real merchandise to several hundred stores and also because it was to involve a friendly competition with the legacy process (see Impact section below for more details), it greatly helped define priorities and focus the energies of all team members. It also helped overcome the cultural barriers mentioned above and generate buy-in from important stakeholders within the company when subsequently planning and executing the full-scale deployment of the new process.

As seen in Figure 6, which shows a timeline of the main project phases and milestones discussed above, the complete development and implementation cycle for this work spanned just under two years, from the first project definition meeting (in August 2005) to the completion of full-scale deployment in the second global warehouse (in June 2007). The corresponding total cost of this project to Zara is estimated to range between $150,000 and $250,000, which excludes the labor costs of all contributing employees not specifically hired for this project.
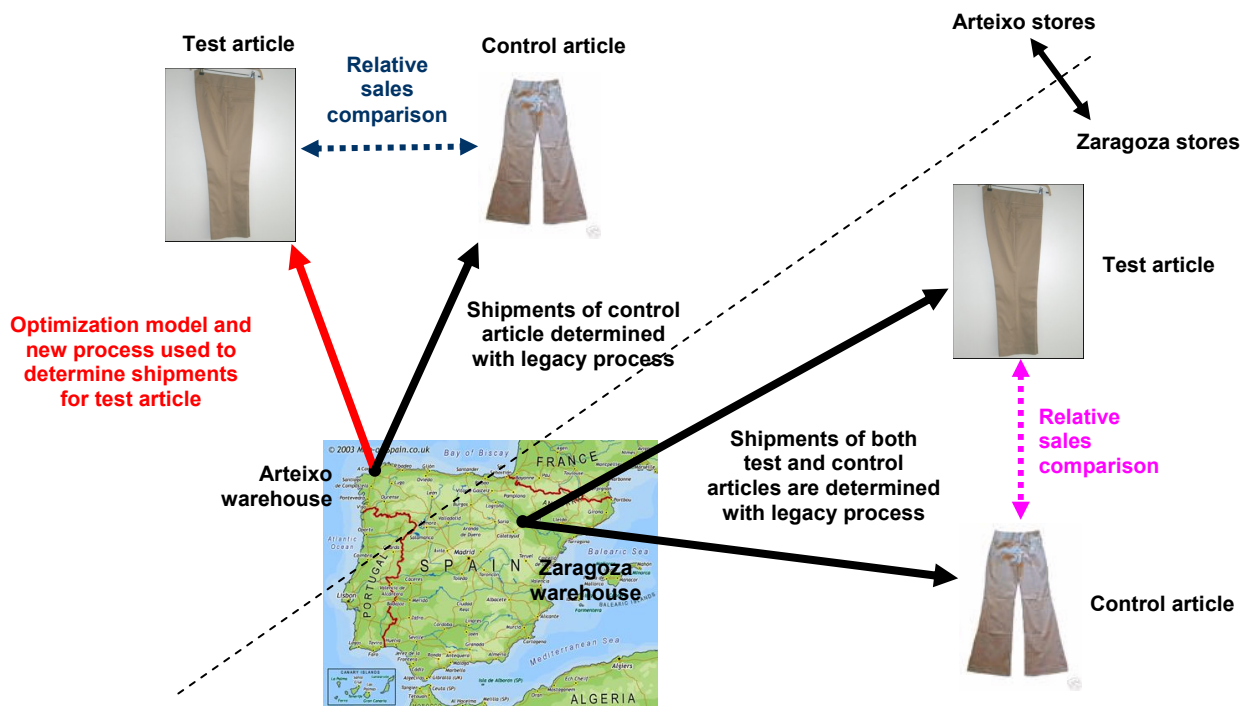
**Figure 7:** **The experimental design of the live pilot involves two dimensions of control for each test article (illustration provided here for one of the ten test articles and its paired control article).**

## Impact

### Measurement Methodology

The methodology followed to estimate the implementation impact involved a live pilot implementation experiment using two dimensions of control, as illustrated by Figure 7. A test group of ten articles was distributed using the optimization model to all the stores that exclusively receive shipments from Zara's first primary warehouse in Arteixo, Spain (this roughly represents half of all Zara stores worldwide). At the same time, a subset of ten "twin" control articles determined through a careful pair-wise matching with the test group was distributed to these same stores using the legacy manual process. This matching procedure was performed in order to enable a relative comparison between the test and the control group supporting an estimation of the specific impact of the new model while eliminating the impact of any other factors external to the model (which would affect both test and control groups) – this constitutes the first dimension of control mentioned above. The second dimension of control exploits the fact that the other half of Zara's stores worldwide only receives shipments from its second main warehouse in Zaragoza, Spain. For this second and relatively independent network of stores, shipments of both test and control groups of articles were determined using the legacy manual process for both groups. This experimental design was selected to enable an estimation of the error associated with the control-based impact estimation methodology described above: since for Zaragoza stores both test and control groups are distributed using the exact same procedure, any measured impact for this second half of the store network may only be attributed to estimation errors as opposed to the use of our new O.R.-based process.

### Financial Impact

The main results of the controlled live pilot experiment described above are summarized in Figure 8. These results are quite conclusive: the control-adjusted relative sales impact in Arteixo is positive for every single article, with a mean across articles of 4.1 percent (median 4.2 percent), whereas the corresponding estimation error calculated using data from Zaragoza is

17

centered around zero (mean and median across articles are 0.7 and -0.6 percent respectively). We also note here that the variability around those means that is seen across different articles in Figure 8 is easily explained by the measurement noise introduced for each individual article by the pair-wise matching procedure and possible forecasting errors. It is thus appropriate to focus indeed 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 offering. When (conservatively) subtracting the estimated average experimental error obtained from the Zaragoza data, these results therefore suggest that the new O.R.-based process increases sales during the selling season by three to four percent. This impact on revenue is easily explained by the model's ability, relative to the legacy process, of moving excessive inventory away from low-selling stores where it is not needed, and send it instead to high-performing stores where it thus reduces missed sales due to stock-outs. It is also explained by its ability to ship all sizes of an article to every store in a concerted manned, likewise sending inventory of specific sizes only where it is likely to sell.
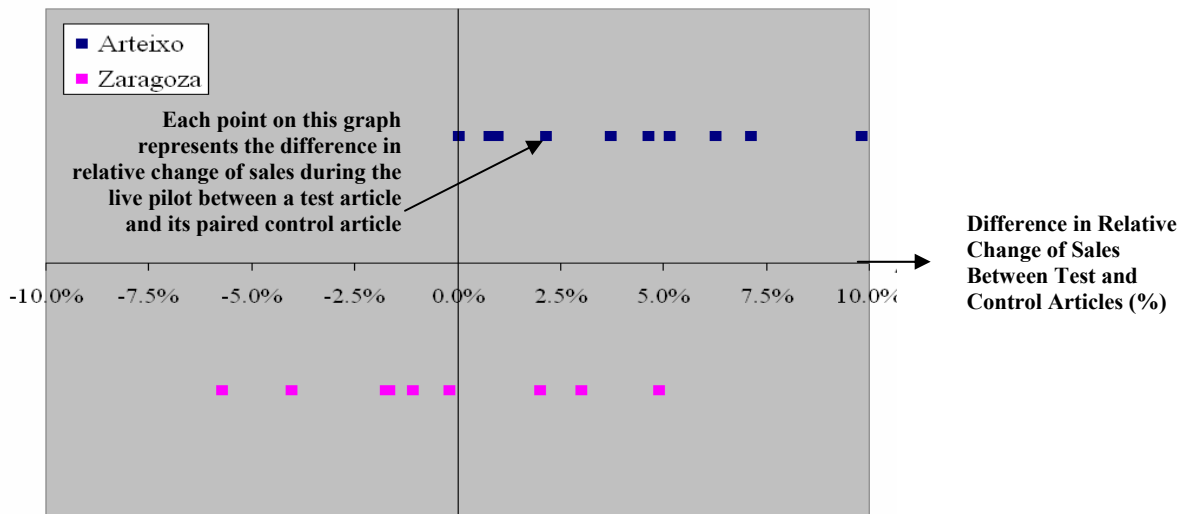


**Figure 8:** **The results of the live pilot experiment suggest that the new process increases sales by approximately three to four percent (adapted from Caro and Gallien 2008).**

18

From a financial standpoint, if the optimization model had been used for all of 2007 and 2008, this relative sales increase would have implied approximately $310M (2007) and $353M (2008) in additional revenue or $37.2M (2007) and $42.4M (2008) in additional net income, with both measures of impact predicted to continue growing at a rate of ten percent per annum in subsequent years. However, the full-scale deployment of the model started in late 2006 and was completed in June of 2007, so that the estimated actual realized impact of the model in 2007 is instead approx. $233M in additional revenue or $28M in additional net income (that is, 75 percent of the previous figures for 2007). In addition, the companion paper Caro and Gallien (2008) 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, even though it was initially planning on having to expand 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 activity: all members of this team have become enthusiastic users of the new tool, gratefully seeing their responsibility 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 due in part to the typical (non-quantitative) background of most employees in the fashion industry, and perhaps also because Zara owes much of its success to the unique intuition of its founder. Truth be told, we doubt that Zara will ever use O.R. models to help with several of its key challenges including anticipating volatile market trends, recruiting top designers and creating fashionable clothes, and it is not clear to us that it should. However, the work presented here did prompt a realization by many of Zara's key executives and employees that

for other processes involving large amounts of quantitative data, well-designed O.R. models will lead to better performance and more scalable operations. Indeed, Zara has initiated two additional major O.R. implementation projects in the areas of purchasing and pricing since the work described here was performed, and is now also actively seeking to recruit graduates with strong O.R. background and became a corporate partner with MIT's Leaders For Manufacturing Program as a result of the present project. In addition, the Inditex Group (which owns Zara) is planning to deploy the O.R.-based inventory distribution process described in this paper in some of its other retail chains such as Massimo Dutti.

## Summary of Contributions to O.R. Practice

To the best of our knowledge, this work constitutes the first reported application of O.R. to the retail business strategy of Fast Fashion adopted by companies that include Zara, H&M and Mango. Under this strategy, the life cycle of articles offered in stores only span a small fraction of a selling season (e.g., 5-6 weeks for Zara), so that the store assortment turns over much more frequently than with more traditional retailers, and frequent store visits are more appealing for customers. From a distribution standpoint, the warehouse inventory available for distribution to fast fashion stores thus tends to be scarcer than for traditional retailers, so that the inventory allocation problem is more difficult.

This work also appears to constitute the first described implementation of an inventory distribution model for an apparel retailer (fast fashion or not) 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 backroom whenever the combination of their sizes still available is not complete enough. While many retailers replenish their stores for each size independently (which can lead to useless shipments from a sales standpoint when major sizes are missing) or ship pre-determined size bundles (which make it difficult to balance the inventory available across sizes in stores), the presented model takes into account the entire inventory profile of all sizes offered, and computes coordinated shipment quantities for all these sizes at the same time (see Analytical Development section for more details).

20

Finally, we emphasize that even though our impact measurement methodology based on a controlled experiment is fairly common in other disciplines (e.g., medicine, social sciences), its application here to O.R. practice is noteworthy – this is because the impact of publicly described O.R. practice work is predominantly estimated through more questionable "before versus after" comparisons, which completely ignore the fact that many other factors besides the O.R. work being described may also be affecting the difference in performance observed in the "after" period.

## Concluding Remarks

Because of Zara's openness to the academic publication of this work, the O.R. models presented here may also impact other firms besides Zara. While this would clearly require some IT implementation and adaptation work, the inventory distribution model discussed above seems quite applicable in particular 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 still capture the notion of required store exposure inventory, also seems quite applicable to most of the retail industry beyond the apparel segment, and would likely constitute a significant improvement relative to the pervasive non-O.R. based methods and simple heuristics used by many firms when distributing scarce inventory to a network of stores (e.g., proportional rationing).

The work presented here has also been very formative for its academic authors (both junior faculty at major research universities over the duration of this project), and it has significantly impacted their research and teaching activities. In particular, the O.R. implementation presented here has already been the object of several masters, executive education and MBA course sessions taught at UCLA, MIT and Columbia University. Many lessons about the implementation of O.R. were also learnt by all team members, several of them being counter-intuitive at first. For example, it may seem important from an academic standpoint to design and study O.R. models capturing all of the relevant key input data (including forecast uncertainty and storage opportunity cost in our case). In practice however, our experience suggests that the implementation of such "comprehensive" models to the control of large-scale core processes in time-sensitive environments may present

insurmountable IT and data availability challenges, so that a hybrid approach combining a simpler but robust optimization model with carefully selected user inputs can be more effective.

In closing, we note that this paper presents a successful and high-impact application of O.R. techniques in the environment of fashion retailing, which is both highly visible to the general public and not currently perceived as a traditional application area for this discipline. We are therefore hopeful that our work will contribute to increase the awareness of O.R., 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 $\lambda_s$ 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

$$\tau_s(q_s) \triangleq \inf\{ t \geq 0: N_s(t) = q_s \}.$$

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

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

As described earlier, all the 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 next Jensen's approximation $E[\tau_{S^+} \wedge T] \approx \min_{s \in D} E[\tau_s \wedge T]$, and observe that

$$E[\tau_s \wedge T] = \sum_{k=1..q_s} \gamma(k, \lambda_s T) / \lambda_s \Gamma(k)$$

$$= \min_{i \in N} \{a_i(\lambda_s)(q_s - i) + b_i(\lambda_s)\}$$

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 $\Gamma$ and $\gamma$ are the Gamma function and the lower incomplete Gamma function, respectively. Our next approximation consists of only computing the minimum in the r.h.s. of the previous equality over the small finite subsets $N(\lambda_s)$ defined as

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

which are straightforward to compute numerically. Appropriate substitutions yield the final expression (see Caro and Gallien 2008 for a complete discussion):

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

**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);
- $\lambda_{sj}$ demand rate for size s in store j;

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

Decision Variables:

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

- $z_j$ correspond to the approximate expected sales across all sizes in each store j for the current period under consideration;

- $y_j$ are secondary variables representing the term $\min_{s \in S^+,\ i \in N(\lambda_{sj})} \{a_i(\lambda_{sj})(I_{sj}+x_{sj}-i)+b_i(\lambda_{sj})\}$

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

Constraints:

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

- $z_j \leq \left(\sum_{s \in S^+} \lambda_{sj}\right) y_j + \sum_{s \in S^-} \lambda_{sj} v_{sj}$ 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);

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

## Acknowledgements

## References

Caro, F., J. Gallien. 2008. Inventory management of a fast-fashion retail network. *Oper. Res.* Forthcoming.

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

Fraiman, N., M. Singh, L. Arrington, C. Paris. 2002. Zara. Columbia Business School Case. Columbia University, NY.

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).

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, NY.