

Demand Forecasting of Retail Sales Using Data Analytics and Statistical Programming

Panagiota LALOU

National Technical University Athens, School of Mechanical Engineering, Athens, Greece

Stavros T. PONIS

*National Technical University Athens, School of Mechanical Engineering, Athens, Greece
staponis@central.ntua.gr*

Orestis K. EFTHYMIUO

Trade Logistics S.A.

Abstract. *Forecasting the demand of network of retail sales is a rather challenging task, especially nowadays where integration of online and physical store orders creates an abundance of data that has to be efficiently stored, analyzed, understood and finally, become ready to be acted upon in a very short time frame. The challenge becomes even bigger for added-value third party logistics (3PL) operators, since in most cases and demand forecasting aside, they are also responsible for receiving, storing and breaking inbound quantities from suppliers, consolidating and picking retail orders and finally plan and organize shipments on a daily basis. This paper argues that data analytics can play a critical role in contemporary logistics and especially in demand data management and forecasting of retail distribution networks. The main objective of the research presented in this paper is to showcase how data analytics can support the 3PL decision making process on replenishing the network stores, thus improving inventory management in both Distribution Centre (DC) and retail outlet levels and the workload planning of human resources and DC automations. To do so, this paper presents the case of a Greek 3PL provider fulfilling physical store and online orders on behalf of a large sporting goods importer operating a network of 129 stores in five different countries. The authors utilize the power of 'R', a statistical programming language, which is well-equipped with a multitude of libraries for this purpose, to compare demand forecasting methods and identify the one producing the smallest forecast error.*

Keywords: Demand Forecasting, Decision Making, Data Analytics, Statistical Programming, Retail Sales, Third Party Logistics (3PL) Operators

Please cite the article as follows: Lalou, P., Ponis, S.T. and Efthymiou, O. K., (2020) "Demand Forecasting of Retail Distribution Networks Using Data Analytics and Statistical Programming", Vol. 15, No. 2, pp. 186-202. DOI: 10.2478/mmcks-2020-0012.

Introduction

In the era of Industry 4.0, more and more decreasing fulfillment order cycle times, call companies for a paradigm shift towards utilizing new generation technologies including warehouse automations to increase fulfillment speed and stay competitive (Efthymiou and

Ponis, 2019; Barreto et al., 2017). Therefore, it is critical to achieve a sustainable and more efficient end-to-end supply chain management (SCM), by utilizing big data analytics technologies to handle massive amount of emerged data and enable informed decision-making. Demand forecasting is a critical input for all levels of decision making -strategic, tactical and operational. In the case of retail distribution networks, demand forecasting is characterized by increased complexity due to short delivery time windows and other sales related internal and external factors (Aksoy et al., 2012). This -inherent to the problem's nature- complexity has further increased in the last ten years due to the rapid expansion and wide use of the e-commerce sales channel on top of the physical one demanding a holistic approach for demand forecasting, order management and inventory replenishment of retail outlets. Needless to say, that currently amidst the unstable period of Covid-19 pandemic with some retail networks shutting down while others stressed to their limits, demand uncertainty is in every managerial agenda all over the world.

In this turbulent environment the role of 3PL operators has become more critical than ever and supply chains rely on them for facing massive shifts in supply and demand and the expected aftershocks for demand planning departments, which call for more efficient forecasting processes and a flexible and highly responsive 3PL delivery network. The objective of this paper is to provide downstream supply chain stakeholders, i.e. manufacturers, wholesalers and 3PL operators with an analytics decision support framework, which will enable them to evaluate and balance operation and capital costs between Distribution Center (DC) automations and stores, by forecasting final customer demand and defining store-level inventory management policies, instead of using a pure pull store replenishment strategy. In that framework, manufacturers and wholesalers are challenged to reconsider their forecasting processes, decide upon and utilize appropriate and informative datasets, analyze them using data analytics tools and provide useful reports for supporting inventory replenishment decisions. At the same time, 3PL operators which have been assigned with demand forecasting on behalf of a client will also gain a significant advantage when forecasting demand using point-of-sale data, as demand forecasts affect their fundamental activities -such as receiving and storing inbound quantities from suppliers, and picking and allocating orders to scheduled shipments. The demand forecasting methodological approach that is developed in this study aims, primarily, to reduce the workload on automations and human resources, which are both involved in fulfillment process at the DC, while maintaining high customer service levels. In doing so, it presents the case of Greek 3PL provider fulfilling physical store and online orders on behalf of a large sporting goods importer operating a network of 129 stores in five different countries. Utilizing historical order data from the importer, the proposed approach is implemented in the case company leading to the selection of the most suitable forecasting model and its application for forecasting SKU order quantities for each store of the network. Finally, simulations are deployed, to quantify the impact of the proposed framework at DC's automations workload, for the selected sample of SKUs and evaluate customer facing service level.

The paper is organized in five sections including this introduction. In the second section, a literature review is presented studying existing efforts in demand forecasting for retail distributed networks followed by a discussion on the different approaches identified. Next, in section three the methodological steps followed, preparing the case company for applying the proposed solution are presented. The solution of the specific problem is

described in details, in section four. In the last section of the paper and after a short summary of the work presented in the previous sections, the findings and conclusions derived from applying the proposed demand forecasting approach are outlined.

Literature Review

Forecasting sales is an integral part of any supply chain, directly and significantly affecting the wellbeing and the prosperity of manufacturers and wholesalers around the world. Indeed, according to Geurts and Kelly (1986), accurate sales forecasts can help management select appropriate levels of aggregate inventory investments and provide essential inputs to many decision activities in functional areas such as marketing, sales, production, purchasing, as well as finance and accounting (Fildes et al., 2019). Alas, achieving sales forecasts' accuracy in practice is a very difficult and complex task, unfortunately tightly interwoven with demand variability and its inevitable nature. In that sense demand forecasting in practice, most of the times refers to the capability of the analyst to predict a range of the expected demand variability instead of an exact demand point. Understandably enough, forecast error is inevitable in sales forecasting and although efforts should be put in reducing it, managers have to understand that sales forecasts must be considered in the light of both the purpose of the sales forecast and the special difficulties inherent in sales forecasting (Sartorius and Mohn, 1976).

Sales forecasting at the retail level, although abiding with the aforementioned general characteristics they do have certain particularities, creating specific business challenges for demand planners. First, one has to understand that forecasts are made at the product and store level in essence feeding an aggregate plan that has to be put in place in order to integrate different demand signals and ignite upstream supply chain processes. Second, retailers in many occasions are autonomous or semi-autonomous organizations with their own strategic mindset and management decision making, which sometimes impedes demand forecasting processes and creates bias and errors, as in the case of poorly or uncommunicated promotional activities at the store level. Finally, as in the business case described in this paper, one has to take into account the immense stress inadequate retail sales forecasting imposes to the downstream delivery network not only to facilities, automations and transportation capacities but also to the workforce of the involved facilities. This is a harsh reality, whether the retailer runs its own distribution network or has outsourced it to a 3PL operator as in the business case presented in this paper or uses a hybrid approach, with many products being supplied from the retailer's own distribution centers but others in a direct store delivery fashion (Fildes et al., 2019). The complexity of the retail sales forecasting problem is further amplified by the emergence of e-commerce and unforeseen and unexpected events as already mentioned in the introduction.

This very demanding business environment has led demand planners and sales executives to the realization that traditional approaches combining product and market expertise, intuition and instincts with the support of traditional legacy system functionalities will not suffice in this new era where e-commerce prevails and advances in technology and data collection systems are simultaneously flooding business systems with high volumes of transactional data populating corporate databases (Boone et al., 2019). Nowadays, the need for fast, reliable and improved, in terms of error reduction, retail sales forecasting systems is an imperative. Traditionally, sales forecasting is implemented with the use of well-established methods such as moving or weighted average, exponential

smoothing applied on time-series with trend and seasonal patterns (Gardner, 1985) and statistical time series analysis models such as ARIMA (Autoregressive Integrated Moving Average) identifying autocorrelations in the data (Box et al., 2015). Although these methods have been applied for decades producing satisfactory results, the inherent linearity of the approaches comes in contradiction with the type of relationship between model variables in real life, which is most of the times non-linear. This deficiency ignited an abundance of scientific publications aiming to address this issue, with the most common approach being that of utilizing the power of Artificial Neural Networks (ANNs). ANN models are usually non-linear and as such they give more efficient results in classification, sample recognition, and forecasting problems than linear models (Aslanargun et al., 2007). This realization inevitably created a large stream of publications comparing ANNs and traditional linear methods for sales forecasting. Alon et al., (2001) for example, by comparing Artificial Neural Networks (ANNs) and traditional methods, concluded that in the case of forecasting aggregate retail sales and on average, ANNs fare favorably in relation to the more traditional linear statistical methods, followed by ARIMA over exponential smoothing. The same result is confirmed by empirical studies, such as those of Kohzadi et al., (1996) and Kuo and Reitsch (1995).

In the next years, research has focused on improving the application and results of ANNs mostly in terms of result accuracy and forecast error reduction. In that direction, Zhang and Qi (2005) attempted to enhance forecasting accuracy of ANNs by applying data curation techniques such as detrending and deseasonalization without being able to dramatically reduce forecasting errors. Au et al., (2008), further studied the use of ANNs by applying Evolutionary Computation (EC) techniques in searching for the ideal ANN structure. By comparing the performance of the EC enhanced models the authors find that the proposed algorithms are better than the traditional seasonal ARIMA models for products with features of low demand uncertainty and weak seasonal trends. Finally, in an attempt to rip the benefits of both worlds, i.e. linear and non-linear models, research efforts turned towards hybrid models for demand forecasting (Aslanargun et al, 2007). Many hybrid models are recorded in literature combining ARIMA and ANNs such as the ones by Aburto and Weber (2007) leading to improved forecasting accuracy for a supermarket retail chain or enhancing ANNs with more search and learning capabilities integrating Genetic Algorithms (GAs) and back-propagation to generate more accurate forecasts on the expense of execution speed (Wong and Guo, 2010). To address the speed issue, several researchers proposed the use of the Extreme Learning Machine (ELM) computing framework, which provides faster learning speeds and has gained significant adoption, especially in the apparel/fashion industry (Fildes et al., 2019).

Summarizing the current status quo of forecasting retail sales, one can identify the contemporary demand planner amidst an abundance of models and solutions to a very difficult problem with almost never a 'correct' solution, which at some points can be overwhelming. Practice shows that most of the times managers tend to use methods and tools they deeply understand and are easy to use, therefore traditional univariate time series analysis methods despite being linear and not able to efficiently grasp underlying patterns, are still widely used in the industry. More complex, intelligent, non-linear methods utilizing the power of machine learning algorithms and tools are used less often and most of the times they are packaged in larger software solutions providing a user interface managers can understand and use by providing input data and parameters

without stumbling on the complexity of the underlying solution engine. Furthermore Seeger et al., (2016) note the strain such algorithms put in computer systems, which have to process and train data sets consisting of thousands of constantly changing SKUs for hundreds of retail stores. The operation and maintenance of such non-linear forecasting models is very demanding and a strong barrier for their adoption by smaller organizations, having limited resources and personnel expertise.

This paper proposes a methodological approach, based on data analytics and statistical programming, to the retail sales forecasting problem that can provide managers with a useful tool enhancing the effectiveness of their traditional sales forecasting toolset and approaches. The proposed approach is tested in a 3PL operating in the apparel industry, which is notorious for offering products of high complexity (color-size assortments), short SKU life (at least two selling seasons within a fiscal year) and high demand volatility. The methodology approach and problem solution described in the next section is relatively simple to implement, explain and introduce at the management level, does not put strain in the organization's computing infrastructure and has proven to work well in tandem with traditional forecasting tools, managers use in their everyday operations.

Problem Description

The methodology proposed in this paper involves a series of steps leading to the selection of the most suitable demand forecasting approach for a 3PL offering demand forecasting and replenishment services for a large apparel retailer. The 3PL operator has a central automated Distribution Center, serving 129 retail stores in five different countries. A brief description of the existing replenishment process at place, follows in the next subsection.

The Store Replenishment Process

In this first methodological step, an analysis of existing replenishment procedures took place. In the case of the examined retailer, store replenishment orders are triggered by daily sales of the stores communicated to the central DC. Specifically, replenishment orders restock the total quantity sold at a store at the end of each business day. The replenishment order is sent to the DC by electronic means, i.e. VPN connectivity between the store's Retail Management System (RMS) and the 3PL's ERP (Enterprise Resource Planning) system. Store orders are divided into two categories: a) initial and b) repetitive replenishment orders. Initial orders relate to new goods receipts from suppliers. Around 40% of the quantity of initial orders is distributed directly to stores after goods receipt in the DC in order to fill their shelves with new arrivals, and covers usually around 3 months of sales. This process is presented in Figure 1.

At the DC level, received store orders generate internal replenishment orders sent to the shop floor for execution by the Warehouse Management System of the 3PL. Order picking follows, collecting products from an overstock box area. Quantities are then loaded through a decanting operation into a Goods-to-Person picking automation system in order to be executed. The automation system is comprised by three towers of automated racking with a total capacity of 22.032 box (totes) storage locations. Each tower has seventeen (17) shuttles for the horizontal movement of totes within the aisles and two lifts for collecting them from the various levels and bringing them to the conveying system. The conveying

system, which is deployed on two levels, connects the automated racking to the pick & pack stations. The system has five (5) pick & pack stations and each of them has thirty (30) stores locations. Therefore, up to a hundred and fifty (150) store orders can be open for picking at any given time. The picking operation is performed by Pick-to-Light buttons located on every store location marking the picking item and its respective picking quantity. The light indication is switched off when the employee has finished picking the suggested quantity and presses the confirmation button. Finally, ready-made boxes continue their route to be grouped by store, loaded on pallets and transported to stores.

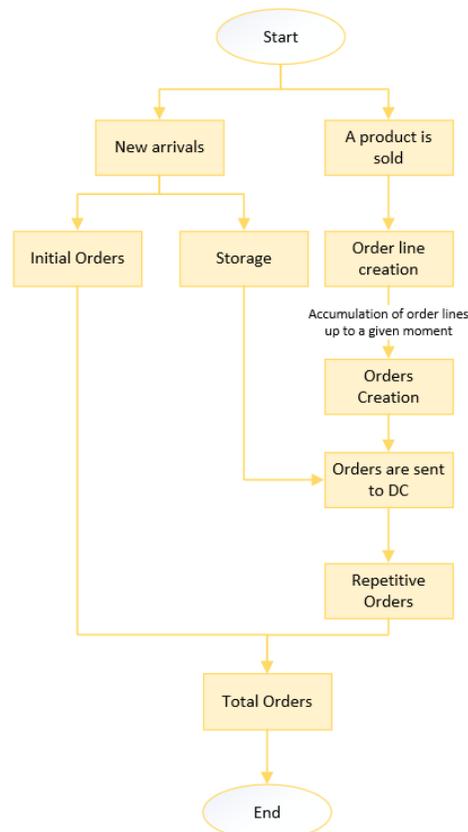


Figure 1. The Store Replenishment Process

Problem statement and Initial Analysis

Products in the apparel industry are seasonal with a short life span, as already mentioned in the previous section. In the studied case, products are sold through two seasons, i.e. Spring – Summer and Fall – Winter. The SKUs can be taxonomized in three categories, i.e. textile, footwear and accessories. The company purchases the total quantity for each SKU once a year, three to six months before every season, and stores it the 3PL's DC. The second step of the proposed approach imposes an SKU profiling process in order to identify possible routes and SKU areas with the potential for large improvements. Therefore, an analysis of more than 8 million order lines of historical order data took place. Results revealed that a relatively small percentage of SKUs, i.e. 0,09% accounted for 6.14% of total replenishment order lines. Further analysis, according to the Pareto principle, showed that only 1.000 SKUs accounted for the 20.86% of total order replenishment lines,

as shown in Figure 2 and Table 1, in the next page. This, in essence, means that by analyzing a relatively small percentage of SKUs, the authors could produce significant benefits at a larger organizational scale. Finally, analysis revealed that more than 80% of total orders are single item orders, while approximately only 3% of all orders have more than three units. Undoubtedly, this is the result of the pull replenishment model the apparel company follows, generating daily replenishment orders mirroring daily sales, as described in the previous section.

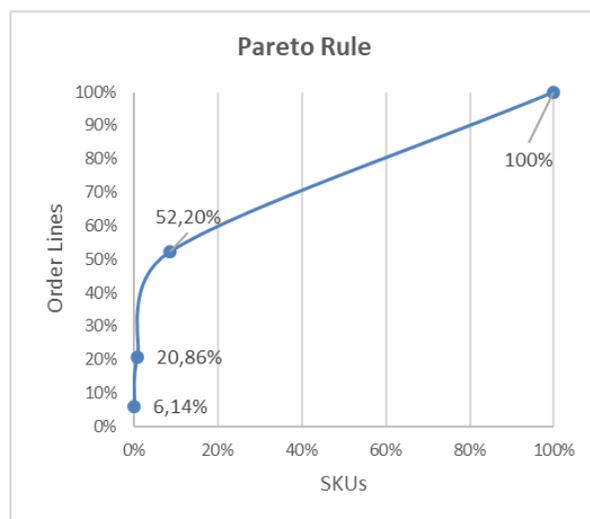


Figure 2. SKU Profiling – Pareto Analysis

Table 1. SKU vs. Order Lines Analysis

Number of codes	Percentage over total number of codes	Number of matching lines	Percentage over total lines
100	0.09%	531,167	6.14%
1000	0.85%	1,806,018	20.86%
10.000	8.5%	4,519,125	52.20%

Cost analysis – identification of relevant costs

In the next step, problem formulation is ignited beginning from the development of its objective function. In doing so, relevant costs are identified. Cost of purchasing and maintaining inventory and transportation costs from the DC to the stores are not relevant in this study. It is assumed that at some point all products will eventually be transported to the stores at the same total cost, and as already mentioned, purchase of inventory happens once before each season for each SKU. Therefore, in this study, two cost elements are taken into account and participate in the model. The cost of order handling ($\frac{D}{Q} \cdot A$), which is the cost of picking an order line at the DC and the cost of maintaining inventory in the picking automation ($\frac{Q}{2} \cdot c \cdot I$), (Equation (1), Table 2).

$$K = \frac{D}{Q} \cdot A + \frac{Q}{2} \cdot c \cdot I \quad (1)$$

Table 2. Explanation of symbols of equation (1).

Symbol	Explanation
K	Total cost
A	Ordering cost
Q	Replenishment Order Quantity
SS	Safety Stock
c	Unit cost
I	Excess Holding Cost

Optimal Order Quantity, known as Economic Order Quantity (EOQ), is exported by deriving equation (1) with respect to Q. Equation (2), which is exported from the prior procedure, is denoted by Q*.

$$Q^* = \sqrt{\frac{2 \cdot D \cdot A}{c \cdot I}} \quad (2)$$

The outcome of replacing Replenishment Order Quantity (Q) with EOQ (Q*) - with reference to equation (2) - in Cost Equation (1), is the cost equation of EOQ (3); the cost of EOQ is denoted by K*.

$$K^* = \frac{D}{\sqrt{\frac{2 \cdot D \cdot A}{c \cdot I}}} \cdot A + \left(\frac{\sqrt{\frac{2 \cdot D \cdot A}{c \cdot I}}}{2} \right) \cdot c \cdot I \Rightarrow K^* = \sqrt{\frac{2 \cdot D \cdot A}{c \cdot I}} + \sqrt{\frac{2 \cdot D \cdot A}{c \cdot I}} \Rightarrow$$

$$K^* = \sqrt{2 \cdot c \cdot I \cdot D \cdot A} \quad (3)$$

It is critical to find a comparative correlation between simplified handling cost (K) and EOQ cost (K*). By dividing equation (1) by equation (3), and replacing the EOQ one (2), the following equation (4) is obtained.

$$\frac{K}{K^*} = \frac{\frac{D}{Q} \cdot A + \frac{Q}{2} \cdot c \cdot I}{\sqrt{2 \cdot c \cdot I \cdot D \cdot A}} \Rightarrow$$

$$\frac{K}{K^*} = \frac{D \cdot A}{Q \cdot \sqrt{2 \cdot c \cdot I \cdot D \cdot A}} + \frac{c \cdot I \cdot Q}{2 \cdot \sqrt{2 \cdot c \cdot I \cdot D \cdot A}} \Rightarrow$$

$$\frac{K}{K^*} = \frac{Q^*}{2 \cdot Q} + \frac{Q}{2 \cdot Q^*} \quad (4)$$

Therefore, in case the reorder quantity is not equal to optimum one, but it is the optimum doubled, say $2 \cdot Q^*$, equation (4) is modified as follows:

$$\frac{K}{K^*} = \frac{Q^*}{2 \cdot 2 \cdot Q^*} + \frac{2 \cdot Q^*}{2 \cdot Q^*} \Rightarrow \frac{K}{K^*} = 1.25$$

Following the same logic, in case the reorder quantity is not equal to optimum one, but it is the optimum tripled, say $3 \cdot Q^*$, equation (4) is modified as follows:

$$\frac{K}{K^*} = \frac{Q^*}{2 \cdot 3 \cdot Q^*} + \frac{3 \cdot Q^*}{2 \cdot Q^*} \Rightarrow \frac{K}{K^*} = 1.67$$

It is obvious that there is an increase up to 25% in handling costs in the case the reorder quantity is double the optimal one, and equivalently, an increase of 66.67% in the case the reorder quantity is three times the optimal one. Still, inventory handling costs represent only a small portion of the total cost, between two and three percent. On the contrary, missed sales due to unavailability, result in higher profit loss. Once again it has to be noted that inventory handling cost in this case study, does not include storage cost, which is not related to the DC's picking automation. The working hypothesis is that storage costs in stores and in the DC are approximately the same for the company as they belong to the same group. Consequently, inventory management costs associated with the overall effort, include the costs of storage inventory in the AS/RS automation and the picking cost, per order line.

Valuation of forecasting models

Two parameters are involved in the selection and formulation of demand forecasting models. The cost of implementing a forecasting model, and the costs incurred by the forecasting errors it produces. The cost of implementing a sophisticated and complex demand forecasting model for low value, and / or low profit margin products may be much higher than maintaining excessive inventory. On the other hand, in high value and / or high profit margin products, where inventory costs are high, and lost sales lead to high profit losses, it is worth investing resources in creating an advanced forecasting model. The factors mentioned earlier, determine the complexity and resources that should be invested prior to selecting, or formulating forecasting models. The optimal area is found by minimizing the sum of costs of these two factors, implementation costs, and forecasting errors costs (Chambers et al., 1971). For that reason in this case, relatively simple forecasting models are examined capable of keeping the forecasting errors within a tolerable range, as presented in the next section.

Solution Approach

A dynamic problem that every distribution center has to solve, is keeping the balance between inventory costs and the service level it provides. The possibility of lost sales led to increased inventory levels, which resulted in excessive inventory capital investments and immense expenditure for their storage and maintenance. To address this problem, technologies have been developed in recent years, which enable strategic decisions to be coordinated with tactical ones regarding inventory management, to achieve a healthy balance. Advanced data analytics tools for visualizing and presenting data could facilitate this process and lead to more efficient decision making (Vassakis et al., 2018).

In the case presented in this paper, the next step of the proposed approach was to identify suitable datasets and curate them in order to obtain a form and structure appropriate for subsequent analysis. Next, the curated datasets were examined in order to identify root causes of the problem, by utilizing analytical methods with the help of statistical programming in the R language programming environment. By utilizing data analytics methods to define inventory management policies, a deeper understanding of the characteristics and trends of demand is feasible, for both central storage units and different sales channels. It is possible to identify, which SKUs are moving too slow or are inactive intermittently, and the sales or volume contribution of SKUs that are active, without requiring time-consuming and complicated procedures. Finally, simulation of alternative inventory replenishment scenarios can be used for quantifying their impact on logistics operations.

Forecasting model determination

The case company, as mentioned earlier, does not use a forecasting model. Instead it utilises a pull approach replenishing stores by servicing the sales quantities of the previous business day. Inevitably, as SKU profiling proves, a great strain is imposed to the DC operated by the 3PL services provider. The absence of any forecasting model in the As-Is situation led to the decision to initiate a forecasting model suitability analysis by first assessing simple linear techniques such as Moving Average. Moving Average models calculate a forecast based on a certain number of past data. The most commonly used Moving Average methods are Simple Moving Average (SMA) and Exponential Moving Average (EMA) (Silver et al., 1998). To evaluate a forecast at computational level, forecast metrics are used, and in more detail error measurements. A forecast error is defined as the difference between the actual demand value and the forecasted one, (equation 5). Equation 5 is explained in Table 3.

$$e_i = A_i - F_i \quad (5)$$

Table 3. Explanation of symbols in equation (5).

Symbol	Explanation
e	Error term of observation i
A	Actual value of observation i
F	Forecasted value of observation i
I	Observation counter

In this study, we used two established forecast error metrics, i.e. Root Mean Square Error (RMSE) (equation 6) and relative Root Mean Square Error (rRMSE) (equation 7). To determine which demand forecasting model has a better fit to the demand patterns at hand, for each product code and for every store, a program was developed in the R language programming environment. Firstly, code – store combinations, which had observations in the selected time period, were selected for further analysis. Observations for every combination were aggregated over time and their outliers were removed. Next, demand forecasting was carried out using Moving Average models. The symbols used in Equations 6 and 7, are explained in Table 4.

$$\sqrt{\frac{\sum_{i=0}^n e_i^2}{n}} \quad (6)$$

$$\frac{\sqrt{\frac{\sum_{i=0}^n e_i^2}{n}}}{\mu} \quad (7)$$

Table 4. Explanation of symbols in equations (6) and (7).

Symbol	Explanation
e	Error term of observation i
n	Total number of observations
μ	Mean value of observations

Two forecasting methods were evaluated, i.e. Exponential Moving Average (EMA) and Simple Moving Average (SMA), using bi-weekly data aggregations. The results show that forecasts generated using the EMA method had better performance in terms of bias and accuracy in contrast to those produced by the SMA method, measured by lower RMSE, rRMSE, and their relative forecast error distribution. A representative example of forecast error distribution using Simple Moving Average (SMA) and Exponential Moving Average (EMA) methods, and bi-weekly data aggregation for a code-store combination is demonstrated in Figure 3. Forecasts generated using EMA and SMA are unbiased since their error distributions are symmetrical to zero, and additionally EMA forecast is better than SMA forecast since EMA's forecast error distribution has more observations near zero and has narrower range compared to SMA's forecast error distribution.

Additionally, in order to evaluate forecast methods, the Diebold-Mariano test was utilized to examine the predictive accuracy of pairs of competing forecast methods (Diebold and Mariano, 1995). The null hypothesis of Diebold-Mariano test states that the outputs of the two forecasting methods do not have significant difference, while rejection of null hypothesis means statistically significant different level of accuracy for the two forecasts. Application of this test to all code-store combinations that were examined for this case,

using “forecast” library in R environment, resulted in rejection of the null hypothesis of Diebold-Mariano test (alternative hypothesis: two-sided), meaning that EMA and SMA forecasts were statistically significant different. This led to selection of the best-performing forecast model based on RMSE and rRMSE values.

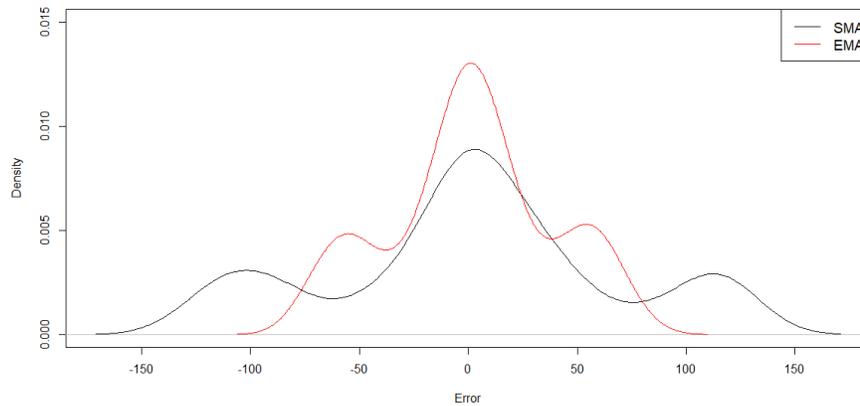


Figure 3. Error distribution by using SMA and EMA methods (weekly data aggregations).

The analysis revealed that by aggregating observations over time, errors are getting lower. For instance, two cases of inventory replenishment frequency were examined, i.e. weekly and biweekly replenishment. Results showed that the average rRMSE value for the case of inventory replenishment once a week was 34.76%, while extending the time period to two weeks provided a significantly lower rRMSE of 18.74%. The proposed demand forecasting process is presented in Figure 4.

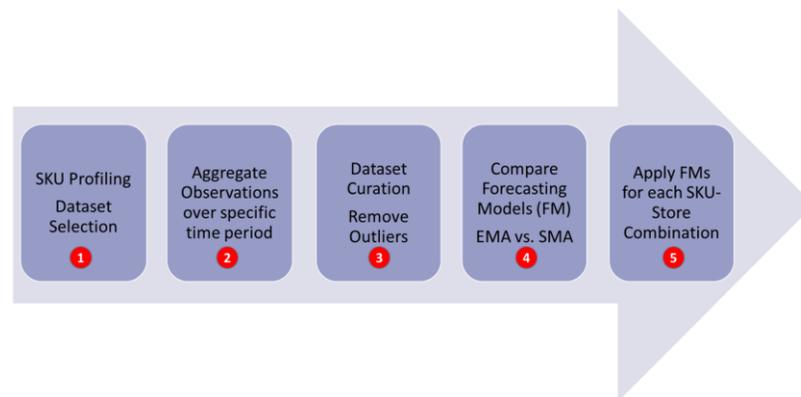


Figure 4. The Proposed Demand Forecasting Approach

Forecasting model determination

To evaluate the proposed forecasting models and replenishment strategies, a simulation model was developed. Data was retrieved from 102 store-code combinations that are classified as fast moving. These combinations are active for at least 35 weeks within a fiscal year. Various combinations of minimum stock in stores and replenishment quantities were simulated in order to evaluate results on DC automation workload and remaining inventory in the stores at the end of each season. Two scenarios were generated for simulation with a moving average of three weeks demand. In the first scenario, one-week demand of safety

stock is maintained at the store level, and replenishment quantity from DC is of up to two weeks of demand as cycle stock (Figure 5). In the second scenario, one-week demand of safety stock is maintained at the store level, and replenishment quantity is of up to three weeks of demand as cycle stock (Figure 6). Figures 5 and 6 below, display the inventory levels for these two scenarios for 42 days of actual sales for the first three store-code combinations.

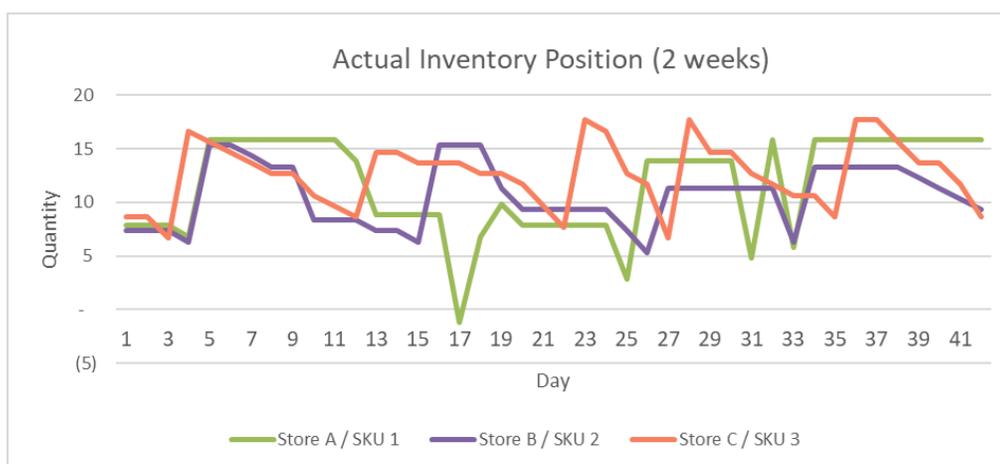


Figure 5. Inventory Levels – Scenario 1

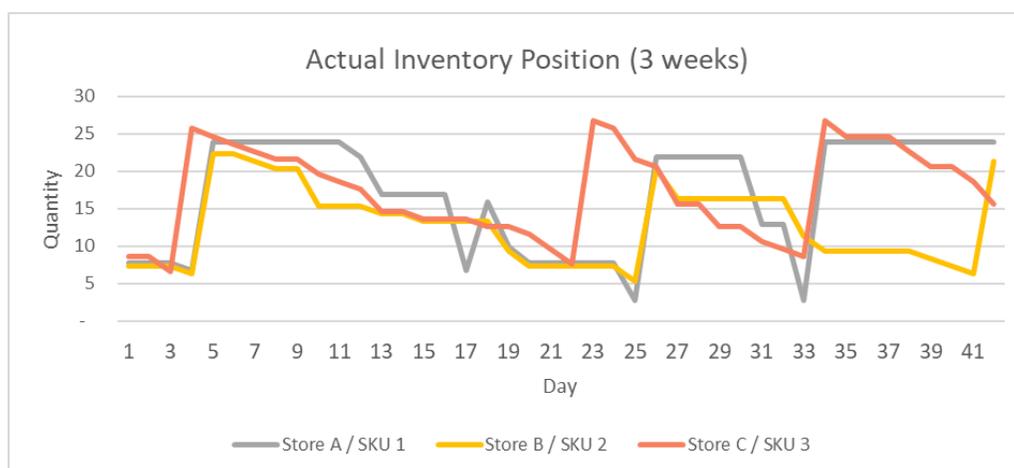


Figure 6. Inventory Levels – Scenario 2

Inventory Position was calculated using equation (8), explained in Table 5.

$$IP = \text{Initial Inventory} - \text{Sales} + \text{Received} \quad (8)$$

Table 5. Explanation of symbols in equation (8).

Symbol	Explanation
IP	Inventory Position
Initial	Inventory at the beginning of current period

Inventory	
Sales	Number of sales during current period
Received	Number of received units during current period

For the calculation of inventory at the beginning of each period, equation (9) is used, explained in Table 6.

$$\text{Initial Inv.}_T = \text{Initial Inv.}_{T-1} - \text{Sales}_{T-1} + \text{Received}_{T-1} \quad (9)$$

Table 6. Explanation of symbols in equation (9).

Symbol	Explanation
Initial Inv. _T	Inventory at the beginning of current period
Initial Inv. _{T-1}	Inventory at the end of last period
Sales _{T-1}	Number of sales during last period
Received _{T-1}	Number of received units during last period

To place an order for an item in a store, Inventory Position should be lower than the Reorder Point. If this condition is satisfied, then an order is placed equal to: Maximum Inventory Level minus Inventory Position. If not, then the order variable of the current period equals to zero. Therefore, a positive value of order variable is translated into an order line for the DC. In the first scenario, with one-week demand of safety stock at the store level, and replenishment quantity from DC of up to two weeks of demand as cycle stock, the number of total order lines equals to 436. Given that the total number of order lines equals to 1,226, a decrease of 64.44% is observed. In the second scenario, with one-week demand of safety stock at the store level, and replenishment quantity from DC of up to three weeks of demand as cycle stock, the number of total order lines equals to 288. Given that the total number of order lines equals to 1,226, a decrease of 76.51% is observed.

Conclusions

The main objective of this study was to investigate the benefits and added value of applying data analytics and statistical programming in the demand forecasting process of an apparel retailer operating a network of 129 stores in five countries. Demand forecasting, inventory management and replenishment has been assigned by the apparel manufacturer to a large 3PL operator. The study was conducted in the 3PL's DC whose operations are continuously strained by the existing pull replenishment system, specifically its Automated Storage and Retrieval System (AS / RS, Order Storage and Retrieval - OSR) which is responsible for decanting and picking. Data analysis revealed that over 80% of all order lines had only one item. This type of very fragmented store replenishment procedure results in low productivity and capacity of the picking automation system. The reason is that for a single order line to be fulfilled, a system tote has to be retrieved from the automated racking, move to the picking station and then the picker has to take the piece and place it on an order carton. Each picking station has a length of about 7.5 meters and therefore the picker has to move from source location to target location back and forth for each pick. This movement takes on average about 15 seconds per pick and it remains the same whether the

picker moves/places one piece, two pieces or more (up to his manual carrying capacity). The only work procedure that adds picking time per piece is labeling, which needs to be done for every single piece.

Therefore, one realizes that by increasing pieces per order line, capacity of the automation system increases at the same time. With an average capacity of the system being 23,000 pieces picked per day in two shifts and with the current ratio of pieces per order line, the number will be substantially higher if this ratio increases. The increased system capacity will then result in less operating hours needed leading to reduced operating costs such as salaries, or equipment maintenance. On the same note, a higher throughput capacity of the system means that its exploitation is more efficient leading to reduced capital expenditures and need for immediate additional investments. Therefore, aggregation of order lines for repetitive ones over a short period of time, is strongly recommended.

Literature review along with analysis and processing of company's data using the R language programming environment showed that simple forecasting models, moving average models in this case, provide a relatively small forecasting error. In more detail, the average rRMSE value for the case of weekly inventory replenishment was 34.76% and for the case of biweekly inventory replenishment was substantially lower, i.e. 18.74%. The inventory replenishment simulation using time series forecasting methods for a representative sample of code - store combinations showed that the load of Distribution Center automations was significantly reduced. Specifically, its results indicated a significant reduction of replenishment lines that DC would have to fulfill, which is an important indicator of DC's workload. More precisely, two indicative scenarios were developed for a dataset of 102 fast moving store-SKU combinations. In the first scenario, with one-week demand of safety stock at the store level, and replenishment quantity from DC of up to two weeks of demand as cycle stock, reduction of automation's workload could be up to 64.44%. In the second scenario, with one-week demand of safety stock at the store level, and replenishment quantity of up to three weeks of demand as cycle stock, reduction of automation's workload could be up to 76.51%. However, it is worth mentioning that service level of the stores for these codes did not fall below 99.70%. Using the methods proposed in this study, significant potential of automation and human workforce could be released as explained above. Resources being released could be utilized to serve more stores, or the fast-growing e-commerce channel, instead of proceeding to potential expansion of warehouse space or automation systems to meet its needs. Consequently, the analysis presented in this paper triggers the need for investigating further the consequences of modifying the current store replenishment model, since this decision is expected to have a significant impact on DC operations.

The research presented in this study is limited by the small size of SKUs participating in the analysis (approx. 1%). Although these SKUs represent a significant percentage of the sum of line orders in the dataset, the analysis should be expanded to include, if possible, the complete product range of the manufacturer. Furthermore, other factors could be potentially added to the replenishment algorithm, such as trends for specific high selling dates like Black Friday, or seasonal sales, or even include weather conditions as parameters affecting sales for the different regions that stores are located, e.g. ordering of items which could be potentially sold such as jackets in early autumn or before a cold weather spell. Finally, further research efforts should focus in the expansion of the proposed approach to

include micro-fulfillment centers, e.g. Pick-Up Points (PUPs) in the problem formulation, since their carrying inventory is becoming substantial, due to the rapidly increasing e-commerce sales.

Acknowledgements

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK-01168).

References

- Aburto, L., & Weber, R. (2007). Improved supply chain management based on hybrid demand forecasts. *Applied Soft Computing*, 7(1), 136-144. <https://doi.org/10.1016/j.asoc.2005.06.001>
- Aksoy, A., Ozturk, N., & Sucky, E. (2012). A decision support system for demand forecasting in the clothing industry. *International Journal of Clothing Science and Technology*, 24(4), 221-236. <https://doi.org/10.1108/09556221211232829>
- Alon, I., Qi, M., & Sadowski, R. J. (2001). Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods. *Journal of retailing and consumer services*, 8(3), 147-156. [https://doi.org/10.1016/S0969-6989\(00\)00011-4](https://doi.org/10.1016/S0969-6989(00)00011-4)
- Aslanargun, A., Mammadov, M., Yazici, B., & Yolacan, S. (2007). Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting. *Journal of Statistical Computation and Simulation*, 77(1), 29-53. <https://doi.org/10.1080/10629360600564874>
- Au, K. F., Choi, T. M., & Yu, Y. (2008). Fashion retail forecasting by evolutionary neural networks. *International Journal of Production Economics*, 114(2), 615-630. <https://doi.org/10.1016/j.ijpe.2007.06.013>
- Barreto, L., Amaral, A., & Pereira, T. (2017). Industry 4.0 implications in logistics: an overview. *Procedia Manufacturing*, 13, 1245-1252. <https://doi.org/10.1016/j.promfg.2017.09.045>
- Boone, T., Ganeshan, R., Jain, A., & Sanders, N. R. (2019). Forecasting sales in the supply chain: Consumer analytics in the big data era. *International Journal of Forecasting*, 35(1), 170-180. <https://doi.org/10.1016/j.ijforecast.2018.09.003>
- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.
- Chambers J., Mullick K. & Smith D., (1971), *How to Choose the Right Forecasting Technique*. [online]. Available at: <https://hbr.org/1971/07/how-to-choose-the-right-forecasting-technique> [accessed at 27.11.2019].
- Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1), 134-144. <https://doi.org/10.1198/073500102753410444>
- Efthymiou, O. K., & Ponis, S. T. (2019). Current Status of Industry 4.0 in Material Handling Automation and In-house Logistics. *International Journal of Industrial and Manufacturing Engineering*, 13(10), 1370-1374. <https://doi.org/10.5281/zenodo.3566333>
- Fildes, R., Ma, S., & Kolassa, S. (2019). Retail forecasting: Research and practice. *International Journal of Forecasting* (in press). <https://doi.org/10.1016/j.ijforecast.2019.06.004>

- Gardner Jr, E. S. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*, 4(1), 1-28. <https://doi.org/10.1002/for.3980040103>
- Geurts, M. D., & Kelly, J. P. (1986). Forecasting retail sales using alternative models. *International Journal of Forecasting*, 2(3), 261-272. [https://doi.org/10.1016/0169-2070\(86\)90046-4](https://doi.org/10.1016/0169-2070(86)90046-4)
- Kohzadi, N., Boyd, M. S., Kermanshahi, B., & Kaastra, I. (1996). A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing*, 10(2), 169-181. [https://doi.org/10.1016/0925-2312\(95\)00020-8](https://doi.org/10.1016/0925-2312(95)00020-8)
- Kuo, C., & Reitsch, A. (1995). Neural networks vs. conventional methods of forecasting. *The Journal of Business Forecasting*, 14(4), 17.
- Vassakis, K., Petrakis, E., & Kopanakis, I. (2018). Big data analytics: Applications, prospects and challenges. In: Skourletopoulos G., Mastorakis G., Mavromoustakis C., Dobre C., Pallis E. (eds) *Mobile Big Data. Lecture Notes on Data Engineering and Communications Technologies*, vol 10. (pp. 3-20). Springer. https://doi.org/10.1007/978-3-319-67925-9_1
- Sartorius, L. C., & Mohn, N. C. (1976). *Sales forecasting models: a diagnostic approach* (No. 69). Georgia State University Press.
- Seeger, M. W., Salinas, D., & Flunkert, V. (2016). Bayesian intermittent demand forecasting for large inventories. In: *Advances in Neural Information Processing Systems* (pp. 4646-4654).
- Silver Ed. A., Pyke D. F. & Peterson R., (1998). *Inventory Management and Production Planning and Scheduling*, 3, pp. 86-98. Wiley.
- Takahashi, K., & Nakamura, N. (2004). Push, pull, or hybrid control in supply chain management. *International Journal of computer integrated manufacturing*, 17(2), 126-140. <https://doi.org/10.1080/09511920310001593083>
- Wong, W. K., & Guo, Z. X. (2010). A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. *International Journal of Production Economics*, 128(2), 614-624. <https://doi.org/10.1016/j.ijpe.2010.07.008>
- Zhang, G. P., & Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160(2), 501-514. <https://doi.org/10.1016/j.ejor.2003.08.037>