Expert Systems With Applications 76 (2017) 140–151

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Cluster-based hierarchical demand forecasting for perishable goods

Jakob Huber[a,*], Alexander Gossmann[b], Heiner Stuckenschmidt[a]

[a] University of Mannheim - Institute for Enterprise Systems, Schloss, 68131 Mannheim, Germany
[b] OPAL - Operational Analytics GmbH, Julius-Hatry-Straße 1, 68163 Mannheim, Germany

ARTICLE INFO

Article history:
Received 12 August 2016
Revised 23 January 2017
Accepted 24 January 2017
Available online 28 January 2017

Keywords:
Predictive analytics
Decision support
Perishable fresh food
Hierarchical demand forecasting
Cluster analysis

ABSTRACT

Demand forecasting is of particular importance for retailers in the context of supply chains of perishable goods and fresh food. Such goods are daily produced and delivered as they need to be provided as fresh as possible and quickly deteriorate. Demand underestimation and overestimation negatively affect the revenues of the retailer. Stock-outs have an undesired impact on consumers while unsold items need to be discarded at the end of the day. We propose a DSS that supports day-to-operations by providing hierarchical forecasts at different organizational levels based on most recent point-of-sales data. It identifies article clusters that are used to extend the hierarchy based on intra-day sales pattern. We apply multivariate ARIMA models to forecast the daily demand to support operational decisions. We evaluate the approach with point-of-sales data of an industrialized bakery chain and show that it is possible to increase the availability while limiting the loss at the same time. The cluster analysis reveals that substitutable items have similar intra-day sales pattern which makes it reasonable to forecast the demand at an aggregated level. The accuracy of top-down forecasts is comparable to direct forecasts which allows reducing the computational costs.

1. Introduction

Retailers offering perishable fast-moving consumer goods face the challenge to provide the right quantity of an article in their store. Perishable goods are typically delivered several times per week (Van Donselaar, van Woensel, Broekmeulen, & Fransoo, 2006) and can only be sold for at most a few days as the freshness of such products decreases rapidly. Hence, items that are not sold by the end of the day are waste and lead to loss. On the other hand, running out-of-stock (OOS) leads to revenue loss as the customers cannot buy the item they are looking for. A retailer can increase its revenue by increasing the availability of the articles while limiting the waste. A common problem of retailers related to the ordering of fresh food is that the order quantities are often determined by store managers based on their experience (Van Donselaar et al., 2006; Van Donselaar, Gaur, Van Woensel, Broekmeulen, & Fransoo, 2010). Recent developments in the area of large scale data analysis provide new opportunities for optimizing sales planning of perishable goods by providing demand-driven short-term forecasts.

The benefits of expert systems in the context of supply chain management depend on the age of the dependent fact data as its value decreases between the occurrence of the respective business event and the executed action (Hackathorn, 2004; Watson, 2009). Traditional business intelligence (BI) (Chaudhuri, Dayal, & Narasayya, 2011) systems are too slow at gathering data that is relevant for short-term or day-to-day decisions (Hahn & Packowski, 2015; Sahay & Ranjan, 2008). However, this is necessary in the context of sales planning for perishable goods which are daily produced and delivered. Traditional BI systems access the data warehouse rather than the operational databases which are optimized for online transaction processing (OLTP) and contain the most recent data. This separation was necessary as the requirements of online analytic processing (OLAP) (e.g. filtering, aggregation, drill-down, pivoting) are different from OLTP. Hence, an ETL (extract transform load) process is necessary for replicating the data into the data warehouse that is accessible by BI systems. In the last decade, the developments of database technology led to column-based in-memory databases that are capable of efficiently handling OLTP as well as OLAP queries (Plattner, 2009; Sikka et al., 2012). A column-based data organization enables efficient data compression and fits the requirements of OLAP queries as they often depend only on a limited number of columns. Maintaining the data in-memory allows using data structures that are not suitable for disk based databases and reduces the latency which makes real-time analytics possible. Due to these advantages, in-memory based databases become more popular in supply chain management. The

* Corresponding author.
E-mail addresses: jakob@informatik.uni-mannheim.de (J. Huber), alexander.gossmann@opal-analytics.com (A. Gossmann), heiner@informatik.uni-mannheim.de (H. Stuckenschmidt).

http://dx.doi.org/10.1016/j.eswa.2017.01.022
0957-4174/

largest class of benefits of real-time BI are related to enhanced operational decisions (Sahay & Ranjan, 2008).

Demand forecasting for perishable goods is an application scenario that can be enhanced with a data-driven decision support system (DSS) (Power, 2008). In particular, a DSS that supports the retailer at different organizational level during the planning process by providing demand forecasts is required (Holsapple & Sena, 2005; Holsapple & Whinston, 1996) in order to standardize and optimize the process. It needs to access a satisfying amount of historical as well most recent point-of-sale (POS) data from the operational database in order to apply techniques and methods like pattern recognition, statistical analysis, regression analysis or predictive modeling. Thereby, POS data needs be aggregated in real-time to the required organizational level. Hence, all prediction models are able to access the same operational data that contains near real-time information of the sales. Those requirements are met by state-of-art technology.

In this paper, we present the core of a DSS that predicts the demand for articles and article groups at different organizational levels. Such a DSS belongs to the first phase of the evolution of Business Intelligence and Analytics (BI&A 1.0) (Chen, Chiang, & Storey, 2012). Our work is motivated by the requirements of an industrialized bakery with respect to demand forecasting. The bakery runs a large number of stores that are daily delivered with baked goods which are produced in a centralized production facility. In this scenario, the total demand for each article as well as the demand at every store needs to be forecast on a daily basis. Based on the forecasts, production and distribution of goods takes place. In the following, we summarize the characteristics and challenges of the domain and justify why it is important to provide demand forecasts to support day-to-day operations.

1.1. Perishable fast moving consumer goods

Fast moving consumer goods (FMCG) comprise articles that are sold at a high frequency as they are mostly required to fulfill the daily needs (e.g. food, drinks, toiletries) (Kaiser, 2011). This group includes goods that have a short shelf life due to the high demand and/or because of the perishable character of the product which makes regular ordering necessary. Van Donselaar et al. (2006) classify items as perishable goods if they have a high rate of deterioration at ambient storage conditions (e.g. vegetables) or an obsolescence date that makes reordering impractical (e.g. newspapers). They report that perishable items have a 50% higher number of average sales per week and a 40% smaller median case pack size compared to non-perishable items. Thus, they conclude that the time between two orders is 2.5 times smaller for perishable goods which indicates that they are rather fast moving goods. Baked goods are classified as daily fresh items which are not only daily ordered but have also a high number of sales. The major cost factor for this type of articles are excessive stock levels that lead to marked down or thrown away items. For this product category, it is possible to rely on the customers' willingness to substitute and to keep the assortment limited (Van Woensel, Van Donselaar, Broekmeulen, & Fransoo, 2007). Studies indicate that the willingness to substitute is higher for perishable items (84%) than for other product categories (50% (Gruen, Corsten, & Bharadwaj, 2002)) which is caused by an immediacy effect, i.e., the item is needed on the day it is bought (Van Woensel et al., 2007). A high service level for all items leads to plenty of leftovers as the demand prediction is uncertain. Thus, it is beneficial to ensure the availability only for a subset of items of a category and to avoid waste of costly items.

1.2. Goods-in-stock: waste vs. out-of-stock/shelf

Inaccurate forecasts lead to overestimation or underestimation of the demand. The effects of overestimation can be quantified as the unsold items of perishable goods are waste and cannot be sold after the shelf life expires. From a financial point of view, the retailer loses the costs related to the production and delivery of the unsold items and has to pay waste collection fees or donate them to charity. For articles having a small profit margin, it is important to limit avoidable costs. On the other hand, demand underestimation leads to OOS which is much harder to quantify as the customer reaction is uncertain. Ehrenthal and Stölzle (2013) consider that an article is OOS if it cannot be bought by a customer at a given point in time. Studies suggest that the global average of OOS is 8.3% (Corsten & Gruen, 2003). OOS leads to an immediate revenue loss of 4% (Gruen et al., 2002) but also affects customer loyalty and jeopardizes future sales (Zinn & Liu, 2008). Hence, inaccurate demand forecasts have a negative financial impact. In the following, we summarize possible effects as well as causes of OOS.

The effects of OOS have been widely investigated (Campo, Gijsbrechts, & Nisol, 2000; Gruen & Corsten, 2007; Gruen et al., 2002; Helm, Hegenbart, & Endres, 2013). Campo et al. (2000) state that customers switch stores, substitute items, postpone the purchase or do not buy anything if the required item is not available. However, the actual response depends on factors like a pre-shopping agenda, urgency of the purchase, brand loyalty and store prices (Zinn & Liu, 2001). So, OOS leads to lost sales, dissatisfied shoppers and diminishes store loyalty. It also obstructs sales planning as the historic sales data is distorted and does not reflect the actual demand. This effects the forecast accuracy because of demand underestimation of items that were occasionally sold out in the past as well as to demand overestimation due to substitution effects. These effects are not limited to the directly affected article category. Ehrenthal and Stölzle (2013) report that OOS of fresh goods leads to the highest turnover loss compared to other categories. Hence, decreasing OOS is a possibility to reduce costs and to increase sales, especially for fresh items like baked goods.

The described effects of OOS underline that a retailer gains a competitive advantage by avoiding OOS. Thus, understanding the causes of OOS is required as it points to issues that need to be improved in order to achieve a better service level. Ehrenthal and Stölzle (2013) report that the causes for OOS in the retail industry are specific to retailer, store, category and item. However, many researcher identified inefficient store operations (Ehrenthal & Stölzle, 2013; Gruen & Corsten, 2007; Gruen et al., 2002) and not issues in the upstream supply chain (e.g. shortage) as primary cause for OOS (Aastrup & Kotzab, 2010). They also observed that the article availability decreases on the downstream towards the retail shelves. However, collaboration and communication between supplier and retailer provoke less problems regarding the article availability. Ehrenthal and Stölzle (2013) optimize the flow of goods by simplifying and structuring the tasks for the store personnel and bundling store deliveries and shelf replenishment. After these operational changes, OOS was mainly caused by erroneous orders instead of fulfillment and replenishment problems.

1.3. Addressing the bullwhip effect

The bullwhip effect is an issue in many supply chains (Lee, Padmanabhan, & Whang, 1997b). It describes the effect of increasing demand oscillation along the upstream supply chain and leads to excessive inventory levels and stock-outs. It is predominantly caused by long lead times, order batching/rationing, shortage gaming and price fluctuations (Lee, Padmanabhan, & Whang, 1997a). In order to resolve the bullwhip effect, the general concept of supply chain cooperation (e.g. collaborative planning forecasting and replenishment (CPFR)) is suggested (Danese, 2011; Hollmann, Scavarda, & Thomé, 2015; Kaipia, Korhonen, & Hartiala, 2006; Thomé, Hollmann, & Scavarda do Carmo, 2014). Thereby, it is typically tried to increase the responsiveness of the supply chain and to make demand more visible by sharing information (Danese, 2007; Eksoz, Mansouri, & Bourlakis, 2014; Holweg, Disney, Holmström, & Småros, 2005; Kaipia et al., 2006; Lee & Whang, 2000; Ramanathan, 2014). Trust among supply chain partners as well as high quality of information sharing supported by adequate information technology are crucial for successful supply chain collaboration (Hollmann et al., 2015). Trust can be gained from long-term relationships (Barratt & Oliveira, 2001) but depends also on information security and confidentiality (Büyüközkan & Vardaloğlu, 2012). The benefits of successful supply chain collaboration include increased responsiveness, increased sales revenues, reduced costs, and improved forecast accuracy (Danese, 2011; Hollmann et al., 2015; Holweg et al., 2005). However, the results depend on the requirements and characteristics of the supply chain which need to be addressed with different types of collaborations (Danese, 2011; Hollmann et al., 2015). Another possibility to mitigate the bullwhip effect is the application of more advanced forecasting techniques that are capable of handling sudden demand changes (Datta et al., 2009; Jaipuria & Mahapatra, 2014). Nevertheless, it is suggested that data (e.g. point-of-sales data) which reflects the true customer demand should be used for time series forecasting (Barratt & Oliveira, 2001; Lee et al., 1997a).

In this study, we investigate how real-time access to point-of-sales data can be exploited to support day-to-day operations in a very responsive bakery supply chain. We focus on the demand forecasting aspect as accurate predictions are crucial to increase the availability and to limit the amount of discarded goods. The company controls all parts of the supply chain from production to delivery and shelf replenishment in the stores. This reduces the barriers of collaboration between different parts of the supply chain. For instance, trust is not an issue and unfiltered access to real-time demand information is given to all parties of the supply chain using state of the art information technology. However, other supply chains for perishable goods involve multiple organizations that need to collaborate (Eksoz et al., 2014). The high responsiveness of the supply that is caused by the characteristics of the products makes it unlikely that the bullwhip effect appears. The perishability of the goods does not allow to keep inventory (Holweg et al., 2005), i.e., the goods are only sold on the day on which they are produced. The lead time is also only one day which allows to provide demand forecasts based on most recent demand observations.

1.4. Demand forecasting

The performance of a supply chain also depends on the accuracy of the demand forecasts (Adebanjo, 2009; Adebanjo & Mann, 2000). Hence, supply chain forecasting is an active field of research (Fildes, Nikolopoulos, Crone, & Syntetos, 2008; Syntetos, Babai, Boylan, Kolassa, & Nikolopoulos, 2016). A study in the fast moving consumer goods sector reveals increased product availability, lower inventory levels along the supply chain and more effective use of current capital assets as major benefits of effective forecasting (Adebanjo & Mann, 2000).

Retailers use (automated) ordering systems for most items. However, it is often the case that orders for perishable items are based on the experience and judgment of the store manager (Fildes, Goodwin, Lawrence, & Nikolopoulos, 2009; Syntetos, Nikolopoulos, & Boylan, 2010) as the systems are not adapted to perishable items (Van Donselaar et al., 2006). This is unsatisfactory as store clerks in different stores apply different rules for different article categories. Additionally, it is non-transparent and the predictions may not be as accurate as desired. In order to estimate future demand, the clerks consider comparable historic data (e.g. sales of previous weeks) as well as various factors like price, promotions, product quality (e.g. based on the origin country) or weather data. This imposes another problem as the time for an order decision is limited and an individual cannot apply her/his logic to several hundred items. Hence, automated decision support systems that implement the respective logic and provide accurate forecasts and help to reduce the workload of the personnel are required. Van Donselaar et al. (2006, 2010) argue that automated ordering systems should be customized for specific product groups in order to provide more reliable results. Based on their observations, they suggest that an inventory control system for perishable items should monitor the total orders for a group of substitutable items as this is an indicator for expected waste.

A variety of time series forecasting methods exists. The most popular traditional approaches (De Gooijer & Hyndman, 2006) are exponential smoothing models (Gardner, 1985; 2006) and auto regressive integrated moving average (ARIMA) models (Box & Jenkins, 1976). Besides those traditional methods, data-driven approaches like artificial neural networks (ANNs) are also considered for time series forecasting (Zhang, 2012; Zhang, Patuwo, & Hu, 1998). We will use ARIMA models as they fit the requirements of our application and are successful applied in comparable contexts (see Section 2). We do not provide a comparison with other forecasting models as this paper focuses on the identification of article groups and the exploitation of the resulting hierarchy.

Hierarchical forecasting (Gross & Sohl, 1990) can be applied to exploit the structure of time series data. The main approaches are top-down and bottom-up forecasting. The top-down approach requires forecasting at an aggregate level and allocating the forecasts to the lower level time series (derived forecasts). The bottom-up approach requires forecasting at a lower level and summing the forecasts to obtain the aggregate level forecasts (cumulative forecasts). The applied approach depends on the objective of the forecast. According to Kahn (1998), top-down forecasting is preferred for strategical planning (e.g. budgets) while bottom-up forecasting is preferred for tactical forecasting where detailed forecasts are required (e.g. production and distribution).

This paper addresses the aforementioned issues and contributes to the literature in various ways. We present a novel approach to detect groups of articles based on intra-day sales pattern (1). We create feature vectors based on point-of-sales data and apply cluster analysis. The obtained clusters are used to build a hierarchy and apply hierarchical forecasting (2). Moreover, we empirically evaluate our approach with real data of an industrialized bakery (3). The remainder of this paper is structured as follows: We give an overview on related work in Section 2. In Section 3, we present our approach to provide demand forecasts by exploiting the domain-specific hierarchy. We evaluate our approach and report the results in Section 4. A conclusion as well as an outline of future work is given in Section 5.

2. Related work

The related work part of this paper comprises work in which demand forecasting and hierarchical forecasting are applied in a comparable context.

Gilbert (2005) and Liu, Bhattacharyya, Sclove, Chen, and Lattyak (2001) use ARIMA models in the context of supply chain forecasting. Liu et al. (2001) predicts the daily demand of ingredients of a fast-food restaurant franchise in order to support the store managers (e.g. inventory management).They require an automated approach as they deal with a large number of time series and also consider outlier detection and adjustment of events that are known or unknown apriori (Chang, Tiao, & Chen, 1988). Shukla and Jharkharia (2013) apply ARIMA models to daily sales data of a wholesale vegetable market and achieve a mean absolute percentage error (MAPE) around 30%. Those approaches do not rely on any product clusters or hierarchies to perform the forecasts.

Kahn (1998) argues that direct forecast are preferable over hierarchical forecasts if the data characteristics (e.g. seasonality, trend) differ. An aggregated time series does not necessarily reflect the characteristics of the lower levels time series which leads to inaccurate forecasts. While the bottom-up approach leads to better forecasts at the lower level, the errors might aggregate and lead to poor forecasts at intermediate and top-levels. Thus, hierarchical forecasting works best if the low level time series share the same pattern. With respect to top-down forecasting, he proposes to proportion the forecast based on seasonal indices. Gross and Sohl (1990) propose a set of disaggregation methods for top-down forecasting that are based on past observations. In order to select the appropriate method, they suggest to compare the disaggregated forecast with the individual forecast. Widiarta, Viswanathan, and Piplani (2009) perform a simulation study to examine the relative effectiveness of top-down and bottom-up forecasting for substitutable products. At product level, the top-down approach is preferred if the degree of substitutability is high. However, if the variability of the demand proportions is high, direct forecasts are required as an accurate the demand allocation is not possible. At product group level, the direct forecast outperforms the bottom-up approach when the demand variability at product level is high and the degree of substitutability between products decreases. Williams and Waller (2011) conduct a study based on weekly point-of-sales (POS) data of cereals which are a fast moving consumer good. They conclude that the bottom-up approach is preferred for forecasting each stock-keeping unit at store level and region level if POS data is available. While those approaches are concerned with hierarchical forecasting, they provide no insights on the hierarchy construction or rely on the given category assignments.

Kalchschmidt, Verganti, and Zotteri (2006) cluster customers (e.g. stores) according to various criteria (e.g. weekly sales pattern, penetration rate) in order obtain homogenous groups. The demand per group becomes less uncertain and variable which leads to more accurate predictions. Their focus is on predicting the total demand of certain products at company level rather than at store level. Thereby, they do also not consider product groups. Zotteri, Kalchschmidt, and Caniato (2005) analyze the impact of the aggregation level (chain, store and cluster level) on the forecasting performance. The objective is to forecast the demand at chain and store level for each item. They propose to cluster time series based on their characteristics (e.g. demand pattern) rather than more intuitive but misleading features like the allocation to a distribution center (e.g. geographical proximity). They report that the bottom-up approach is more accurate than the direct forecasts at chain level. However, clustering demand leads to better predictions for fast-moving items while the bottom-up approach is better for slow moving items. The top-down approach has the advantage of minimizing the total number of forecasts which leads to reduced computational costs. They do only exploit the hierarchical structure of the retailer and do not rely on product categories.

We identify the following gaps in existing literature: We are not aware of any work that exploits the organizational structure and the product hierarchy. To the best of our knowledge, no empirical study focusing on hierarchical forecasting for substitutable perishable goods exists. We are also not aware of a data-driven approach to identify product groups based on intra-day sales pattern.

3. Approach

We propose to forecast the demand of substitutable perishable goods by constructing and exploiting the domain-specific hierarchy (see Fig. 1). The main objective of the underlying decision support system is to optimize day-to-day operations by providing daily demand forecasts for each offered product at different organizational levels. In particular, daily demand estimation is important for production planning at company level and at store level to coordinate distribution for each article. Accurate demand predictions are a key for an increased availability and a limited amount of discarded goods. For this purpose it is sufficient to provide a short-term forecast (e.g. 3 days). In order to construct the hierarchy (see Section 3.1), we cluster articles having comparable sales pattern (see Section 3.2). ARIMA(X) models are used to forecast time series that are derived from point-of-sales data that is aggregated to the desired level (see Section 3.3).

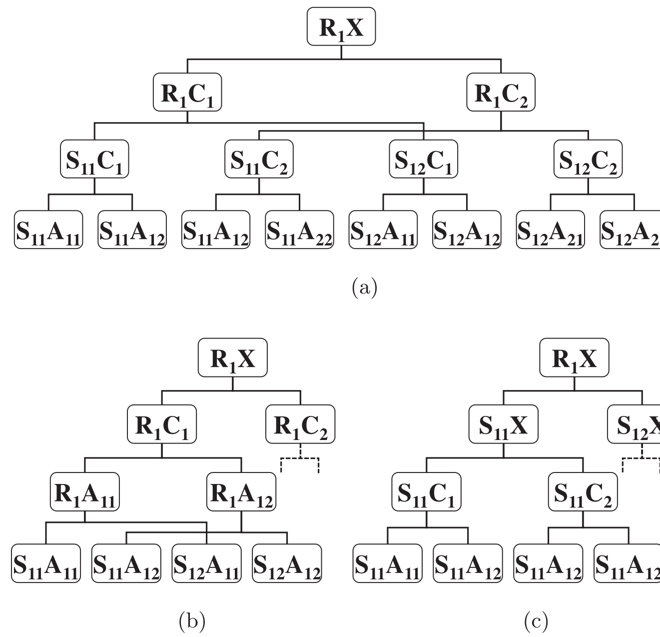

Fig. 1. The figures illustrate hierarchies that can be used for hierarchical forecasting. Each node refers to a time series of a certain level that is identified by two letters. The first letter specifies a region (R) or a store (S). The second letter specifies an article (A), an article group (C), or the group of all articles (X). (a) RX-RC-SC-SA, (b) RX-RC-RA-SA, (c) RX-SX-SC-SA.

3.1. Hierarchy construction

The organizational structure of retailers can often be presented as a hierarchy (see Fig. 1). The stores are grouped into regions having their own distribution centers. Moreover, the articles that are offered by the retailer build a hierarchy itself, i.e., several articles can be grouped into an article category. Hence, various hierarchies can be build and exploited for demand forecasting. In the context of this work, we consider the following six levels of the hierarchy that can be mapped to the time series:

- Region (RX): The total quantity sold within the region.
- Region Category (RC): The total quantity sold of articles of a specific category (cluster) within the region.
- Region Article (SA): The total quantity sold of a specific article within the region.
- Store (SX): The total quantity sold at a specific store.
- Store Category (SC): The total quantity sold of articles of a specific category (cluster) at a specific store.
- Store Article (SA): The total quantity sold of a specific article at a specific store.

While forecasts are only required at article level (RA + SA), the category levels (RC + SC) are also useful as they allow to monitor the demand within a group of items and to validate the predictions at article level (RA + SA). Moreover, forecasts for a cluster of substitutable items are valuable if the assortment is changing or some articles are temporarily not available due to delivery problems or item damage. For instance, the demand forecast of a cluster can be used to estimate the demand for a new article if a seasonal article gets replaced. Moreover, the demand for an article group is less volatile than for single articles. This helps to reduce the risk of excessive stock levels and stock-outs for the whole article cluster. Another advantage is that differently aggregated time series allow tracking the dynamics at different scales. We also introduce the total aggregation of sales (SX + RX). The purpose of these levels is to analyze the benefit of the introduced clustering approach with respect to the forecasting accuracy.

The introduced hierarchy can be exploited by applying a hierarchical forecasting approach. Thus, we forecast the demand at category level and allocate it to the article level (top-down). Therefore, we compute an allocation scheme based on historic proportions as suggested by Gross and Sohl (1990). For each weekday w, the demand proportion PR of an article a at layer l_1 depends on the total demand at the upper level l_2 based on past sales observations:

$$PR_{l_1,l_2,a,w} = \frac{\sum_t s_{l_1,a,w,t}}{\sum_t s_{l_2,w,t}} \tag{1}$$

3.2. Article clustering

In order to gain the most benefit from a hierarchy, it is important to rely on meaningful article clusters. We propose to automatically detect article groups by clustering articles according to their intra-day sales pattern. We do this as studies indicate that perishable articles have a high substitution rate in case of OOS, e.g., customers buy another article of the same category in the same store (Van Woensel et al., 2007). Hence, it is important to maintain a high service level for at least one product of a cluster of substitutable items if not for every article. By risking that some articles are OOS, the total amount of waste can be limited while the expected revenue loss might be acceptable due to the substitution effects. Moreover, predicting the demand of the complete cluster might lead to more accurate forecasts as the time series of the single articles can be distorted due to various effects (e.g. promotions, OOS). By using a data-driven approach for determining article groups, we do not need to rely on a possible erroneous or not fine-grained article category assignment offered by a retailer.

In order to perform a cluster analysis, we transform the point-of-sales data into feature vectors $P_{a,\,q,\,w}$ representing the intra-day sales pattern for each article a in a specific quarter q on each weekday w. Hence, each article is represented by 24 vectors if the stores open from Monday to Saturday.

$$P_{a,q,w} = (p_{a,q,w,1}, p_{a,q,w,2}, \ \ldots, \ p_{a,q,w,T}) \tag{2}$$

We introduce a vector for each weekday and quarter in order to reveal possible differences in the demand pattern and to cover seasonal aspects. For instance, the demand pattern of working days and weekends could be distinguishable. Moreover, different environmental factors (e.g. weather conditions) might cause different demand pattern in the summer compared to the winter. The length of $P_{a,\,q,\,w}$ depends on the maximal number of hours T the stores are open. Each element $p_{a,\,q,\,w,\,t}$ represents the average relative proportion of the total daily sales that is sold in the respective hour t.

$$p_{a,q,w,t} = \frac{s_{a,q,w,t}}{\sum_t s_{a,q,w,t}} \tag{3}$$

The variable $s_{a,\,q,\,w,\,t}$ represents the total sales for article a in quarter q on weekday w and hour t. We cluster the generated features with the k-means algorithm. The algorithm ensures that each vector is assigned to exactly one cluster (strict partitioning). Moreover, the center of a cluster can be interpreted as general demand pattern of the allocated articles. In order to apply the algorithm, one has to set the number of cluster k. It is noteworthy that the number of clusters should be aligned to the characteristics of the demand pattern. Therefore, we suggest to apply agglomerative hierarchical clustering in a preceding step as this helps revealing a hierarchical structure and determining a suitable number of clusters. A suitable linkage criterion for our use case is Ward's method (Ward Jr, 1963) which merges clusters so that the within cluster variance is minimal.

After the feature vectors are allocated to clusters using the k-means algorithm, we determine the final article groups by majority vote. This is necessary as each article is represented by several feature vectors and it is not guaranteed that all are part of the same cluster. Thus, we assign an article to the cluster to which most of its feature vectors belong. The obtained article clusters are complemented with the organizational structure to build the hierarchy. Moreover, the clusters can be used for defining forecasting models that are optimized for different article groups.

3.3. Forecasting

In order to forecast time series data, we rely on the auto regressive integrated moving average (ARIMA) model as well as on an extension of ARIMA, called ARIMAX, which also allows incorporating external factors/explanatory variables. These models are well established in time series forecasting and have been shown to perform well in other settings (see Section 2). The auto regressive (AR) part of ARIMA represents a linear combination of past values while the moving average part (MA) of ARIMA is a linear combination of past forecast errors. The time series has to be stationary which can be achieved by differencing (I). The standard model can be written as ARIMA(p, d, q) while the seasonal model can be written as ARIMA(p, d, q, P, D, Q)s. Here, p (P) represent the order of the (seasonal) auto-regressive part, d (D) are the orders of non-seasonal (seasonal) differencing and q (Q) are the order of the non-seasonal (seasonal) moving average part. Moreover, s states periodicity of the time series. The seasonal ARIMA model for a time series Y_t (t= 1, ..., n) is defined as follows:

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D Y_t = \mu + \theta(B)\Theta(B^s)e_t \tag{4}$$

The operator B is the backshift operator, i.e., $BY_t = Y_{t-1}$, μ is a constant and e_t are error terms. The auto-regressive ($\phi(B)$, $\Phi(B^s)$) and moving average ($\theta(B)$, $\Theta(B^s)$) parts are expressed as polynomials.

$$\phi(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - \ldots - \phi_p B^p \tag{5}$$

$$\Phi(B^s) = 1 - \Phi_1 B^{1s} - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps} \tag{6}$$

$$\theta(B) = 1 - \theta_1 B^1 - \theta_2 B^2 - \ldots - \theta_q B^q \tag{7}$$

$$\Theta(B^s) = 1 - \Theta_1 B^{1s} - \Theta_2 B^{2s} - \ldots - \Theta_Q B^{Qs} \tag{8}$$

The main challenge of using ARIMA(X) is to set the parameters (p, d, q, P, D, Q ,s) properly which requires statistical knowledge and depends on the nature of the data as well as on assumptions about the nature of customer behavior. Hyndman and Khandakar (2008) and Rojas et al. (2008) propose methods that automatically determine the parameters. We rely on an implementation of a method developed by Hyndman and Khandakar (2008) that automatically selects the ARIMA(X) model having the lowest Akaike Information Criterion (AIC) (Akaike, 1974) for seasonal and non-seasonal data. AIC is a measure for describing the relative quality of a statistical model for a data set by estimating the information loss and considering the complexity of the model. In order to select the parameters, they use a step-wise approach and traverse the space of possible models in an efficient way until the optimal model is found.

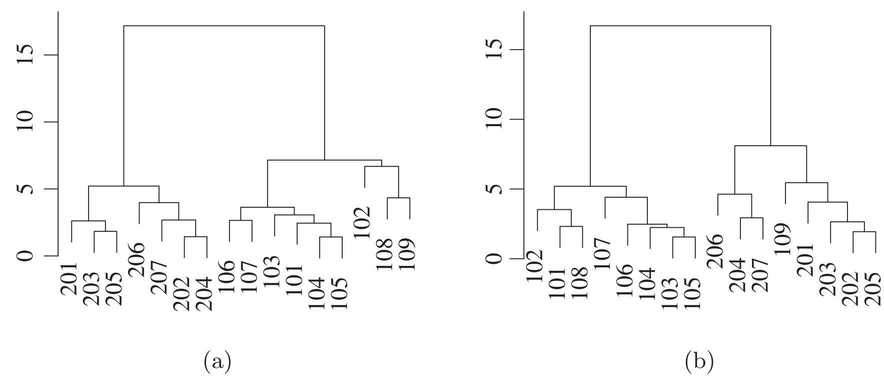

Fig. 2. Hierarchical cluster analysis based on intra-day sales pattern of articles on working days and weekend. The dataset yields two main clusters which nearly perfectly match the two article categories. (a) Working Days (Mon – Fri), (b) Weekend (Sat).

For daily forecasts, we have various options to construct the time series based on point-of-sales data. The most straight forward approach is using a time series describing the daily sales for all weekdays on which the stores are open (e.g. Monday to Saturday). In order to forecast the demand for the next days, we can directly set the forecast horizon to the desired length. Another possibility is to decompose the time series into one series for each weekday. This seems to be reasonable as a weekly sales pattern is prevalent (Adebanjo & Mann, 2000; Van Donselaar et al., 2010). This scenario requires to learn a model for each weekday that falls in the forecast horizon. Hence, for a three day forecast, we need three models that predict the next point of the respective time series. This reduces the length of the time series by 83.33% to which a prediction model needs to be fitted. In total, only 50% of the existing data is used to predict the sales for the next three days which also causes a reduced calculation time. In this work, we use a rather limited set of explanatory variables as we focus on the exploitation of the hierarchy. This includes seasonal dummy variables as well as three variables that model the effects of public holidays. Those variables state on how many of the next days the store will be closed, how many days passed since the store was closed, and how many days are left till the store closes again.

4. Results and discussion

We evaluate the proposed approach in the context of demand forecasting for an industrialized bakery. The dataset comprises point-of-sales data over 18 months of 16 articles that are sold in 6 stores. The stores belong to the same region while the articles can be grouped into the two categories buns (9 articles (id: 101–109)) and breads (7 articles (id: 201–207)). The articles are produced in a centralized production facility and get daily delivered to the stores. Items that are not sold by the end of the day have to be discarded as the articles are perishable goods that quickly deteriorate. In this scenario, it is important to provide demand forecast at regional level in order to plan the production as well as on store level for deliveries. We present and discuss the results of the clustering module in Section 4.1 and the results of the forecasting module in Section 4.2. Moreover, we present the practical and managerial implications of our work in Section 4.3.

4.1. Article clustering

We apply the proposed clustering approach (see Section 3.2) to the 16 articles of the bakery dataset. The cluster analysis is based on 384 feature vectors as the stores are only closed on Sundays. The hierarchical cluster analysis reveals that the demand pattern of working days are distinguishable from weekends. Moreover, we observe two groups of articles that nearly perfectly match the article groups buns and bread (see Fig. 2). Based on these observations, we decide to split the feature vectors into one set that contains all vectors of working days (320 vectors) and one set the contains all vectors of weekend sales pattern (64 vectors). For each set of vectors, we apply k-means with $k = 2$.

The results of the cluster analysis are depicted in Fig. 3. Overall, the resulting clusters are quite pure and accurate compared to the given article category assignment. In this case, we use the original category assignment as gold standard as the two categories already contain substitutable goods and thus are reasonable clusters. However, this has not to be the case in other scenarios. The cluster analysis shows that the demand pattern for buns (see Fig. 3a and c) are distinguishable from breads (see Fig. 3b and d). Based on these results, we assign the articles 101–109 to cluster C1 (buns) and articles 201–207 to cluster C2 (breads). It is also mentionable that the pattern for different buns (respectively breads) are similar which underlines the assumption that they have comparable characteristics with respect to the customer demand. The clusters show that buns are mostly sold in the mornings while the demand for breads is higher in the afternoon. This suggests that buns are the preferred product in the morning while bread sales are rather equally distributed over the day. Moreover, we observe peaks during lunchtime and in the afternoon which seem to be related to the working times of employees in Germany. We also observe that the demand pattern of working days is different from the weekend. On Saturday, the demand for buns is very high in the mornings and drops continuously during the day. For breads, we do not observe the second peak in the afternoon that we see on working days. For all clusters, we observe that the sales drastically decrease during the last opening hours due to less demand. Hence, running OOS during the last hour of the opening hours may not have a big impact on the revenues and might be acceptable if it decreases the amount of discarded goods.

Clustering articles based on their intra-day sales pattern is a possibility for detecting comparable articles (e.g. breads). This is required as the substitution effects for baked goods are higher than for other article categories and thus maintaining a high service level for only one article per group might be reasonable in order to limit the amount of discarded goods without sacrificing the revenues. This can be exploited by hierarchical forecasting. Intra-day sales pattern give also implications for store operations, e.g., shelf replenishment needs to be aligned with periods of high demand. Moreover, having clusters of comparable articles allows defining cluster-specific prediction models. Specifying optimized models for each article cluster might result in more accurate forecasts than one global model. This is reasonable as different categories depend on distinguishable explanatory variables.

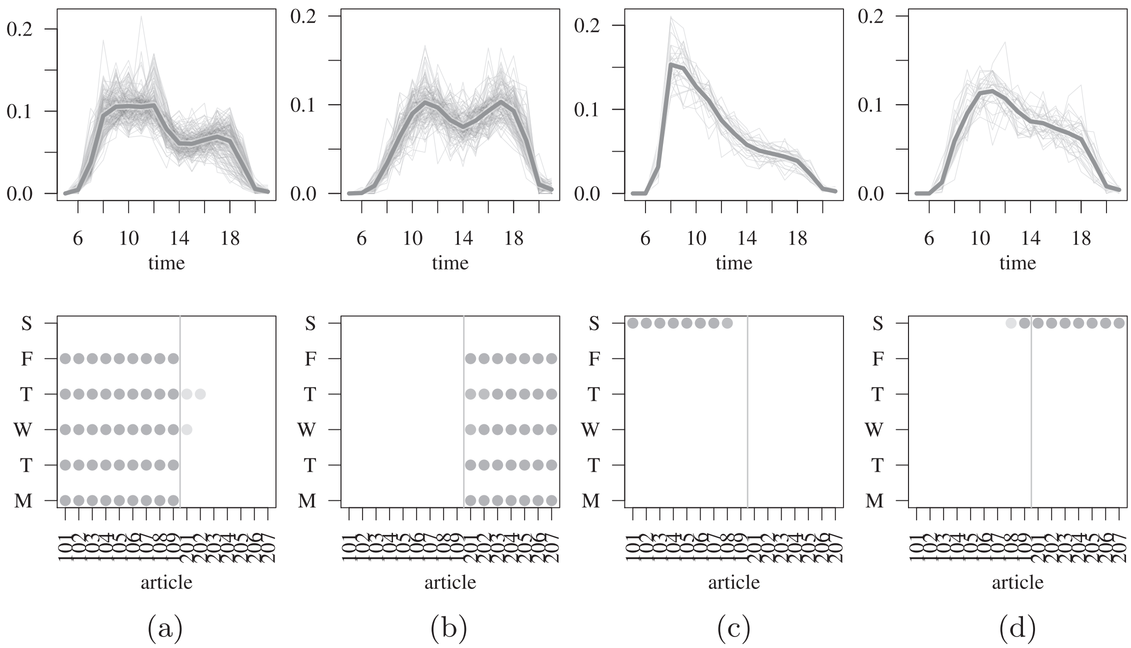

Fig. 3. The intra-day sales pattern are clustered using *k*-Means. The graphs in the first row show the sales percentage of the daily sales in each hour. The graphs in the second row indicate which sales pattern with respect to article (x-axis) and weekday (y-axis: Monday (M) - Saturday (S)) are represented in each cluster. Cluster WD (1+2) are obtained using the sales pattern of working days while Cluster WE (1+2) are based on sales pattern of weekends. The Rand index (accuracy) is 0.9814 for Cluster WD and 0.8537 for Cluster WE. The purity is 0.9906 for Cluster WD and 0.9219 for Cluster WE. The overall purity is 0.9791. (a) Cluster WD 1, (b) Cluster WD 2, (c) Cluster WE 1, (d) Cluster WE 2.

4.2. Forecasting

We report the results of the evaluation that is mainly concerned with determining how well the different levels of the hierarchy (see Fig. 1) can be forecast. However, we focus on the prediction accuracy for the daily demand at store article level (SA) and region article level (RA) as these are the most relevant levels in the present use case. This is caused by the fact that decisions about production and delivery quantities are made on article level. We evaluate the predictions with the percentage based error measure *mean absolute percentage error* (MAPE) and the scale-dependent measure *root mean square error* (RMSE) (Hyndman & Koehler, 2006):

$$MAPE = 100 \cdot \frac{1}{n}\sum_t \left|\frac{Y_t - F_t}{Y_t}\right| \tag{9}$$

$$RMSE = \sqrt{\frac{\sum_t (Y_t - F_t)^2}{n}} \tag{10}$$

We use the MAPE to assess the accuracy of the predictions as it is widely used and its disadvantages do not apply in our application scenario. This is possible as baked articles are fast moving goods that are sold in high quantities every day. Hence, the MAPE is always defined and meaningful. Additionally, we compute the RMSE as it penalizes larger errors more than small errors which is desired in this context. However, it has the drawback that it is scale-dependent and does not allow comparisons between different categories or levels of the hierarchy. In this work, we want to measure the accuracy of the predicted demand which does not reflect the optimal order quantity. Therefore, we use measures that do not take the costs of demand underestimation or demand overestimation into account.

However, the forecast represents the optimal order quantity if costs of underestimation and overestimation are equal. If this is not the case, the optimal order quantity can be obtained by applying a newsvendor model (Silver, Pyke, & Peterson, 1998). In order to assess the supply chain performance of our approach, we assume symmetric costs and measure the loss rate (*LR*), fill rate (*FR*), and service level (*SL*):

$$LR = \frac{1}{n}\sum_t \frac{(F_t - Y_t)^+}{Y_t} \tag{11}$$

$$FR = 1 - \frac{1}{n}\sum_t \frac{(Y_t - F_t)^+}{Y_t} \tag{12}$$

$$SL = \frac{1}{n}\sum_t \mathbb{I}(Y_t \le F_t) \tag{13}$$

The loss rate indicates the percentage of goods that need to be discarded in relation to the actual sales, the fill rate represents the percentage of demand that can be fulfilled while the service level specifies the probability that the demand can be completely fulfilled on a given day. The expected service level is 50% for unbiased forecasts. Please note that we refer to fill rate when we use the term “service level” outside of this section.

The data set comprises time series data over 18 months for 16 articles that are sold in 6 stores. This results in 96 time series at the SA level. We use the first 15 months as training set and the last 3 months as test set. Within the test set we randomly pick 25 dates that are used as starting dates for daily forecasts having a horizon of three days. In the following, we investigate the direct forecasts at each level (see Section 4.2.1) and based on these results the hierarchical forecasts (see Section 4.2.2).

4.2.1. Direct forecasts

We compute the forecasts at different levels using multivariate ARIMAX model as well as a univariate ARIMA model (ARIMA (1-WD)) that uses only data of the weekday it has to predict. Hence, for a 3-day forecast it extracts three time series from the original daily time series and forecasts for each of those the next value. The process of the time series disaggregation is depicted in Fig. 4. This approach seems reasonable as the demand for each weekday is relatively stable. The parameters of the ARIMA(X) models

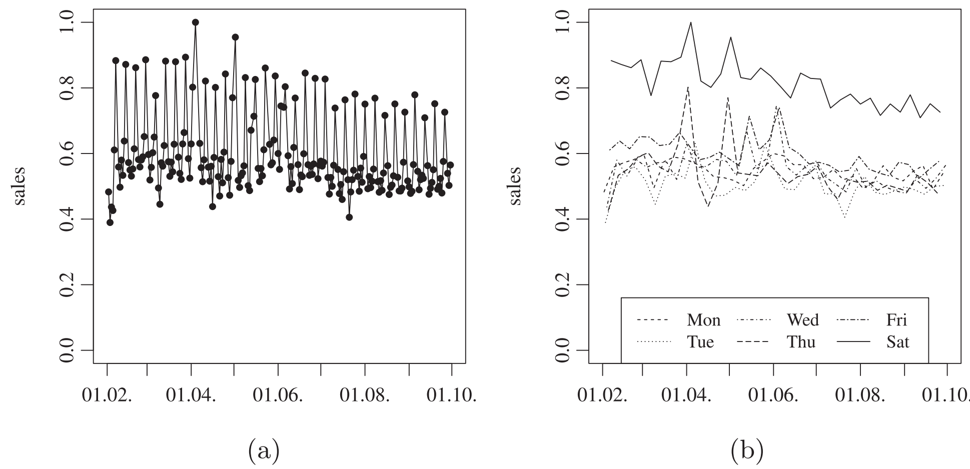

Fig. 4. An excerpt of time series RC 1 (buns at regional level) in standard form (left) and disaggregated by weekday (right). (a) standard, (b) per weekday.

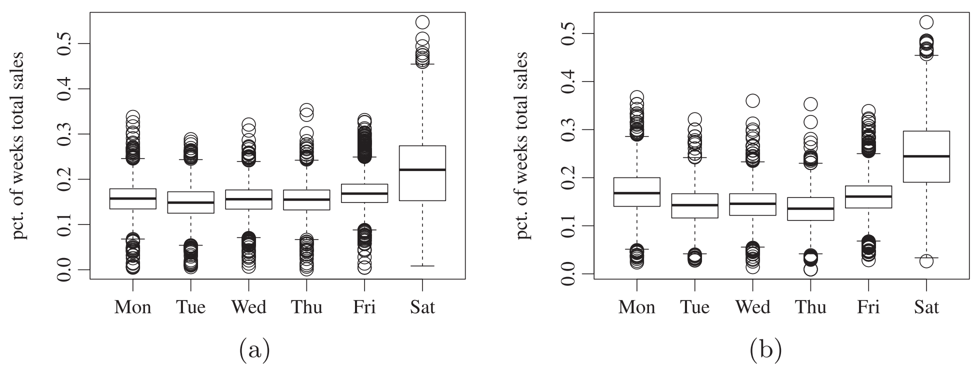

Fig. 5. Weekly sales pattern. (a) Buns, (b) Breads.

Table 1
Overview of the used ARIMAX models for the different levels.

Level	p	d	q	P	D	Q	s
RX	5	1	0	2	0	0	6
RA 1	5	1	0	1	0	0	6
RA 2	5	1	0	1	0	0	6
RC 1	5	1	0	2	0	0	6
RC 2	1	1	0	2	0	0	6
SX	5	1	0	2	0	0	6
SA 1	5	1	0	2	0	0	6
SA 2	5	1	0	2	0	0	6
SC 1	5	1	0	2	0	0	6
SC 2	3	1	0	2	0	0	6

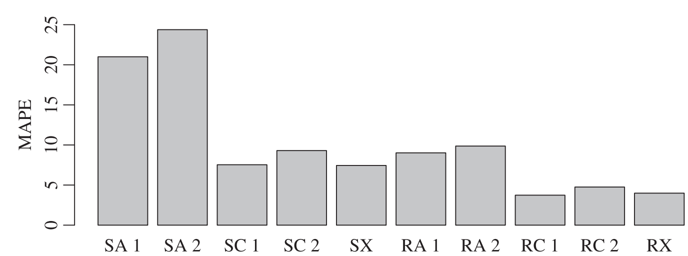

Fig. 6. Comparison of the forecast accuracy at different levels using the combined model.

are determined within the training set for each level and cluster. The selected models are given in Table 1. We rely on seasonal ARIMAX models as a weekly seasonality is prevalent for both groups of articles (see Fig. 5). The demand of working days (i.e. Monday–Friday) is on a comparable level while the demand on the weekend (Saturday) is higher. This observation is consistent with the cluster analysis of the intra-day demand pattern that also reveals differences between working days and the weekend. The demand process seems to be purely auto-regressive as no moving-average term is obtained in the models. Based on these two models, we also compute a combinatorial forecast that weights the ARIMAX models with 0.75 as it incorporates more information and the ARIMA (1-WD) model with 0.25. Moreover, we use the seasonal naïve method as baseline, i.e., the forecast is the last observation of the same weekday, as it comes close to the actual behavior of store managers that are not supported by a DSS and have only limited information on the sales history.

The results of the evaluation of the forecast models at the different levels are shown in Table 2. The ARIMAX model outperforms the ARIMA (1-WD) model at store level with respect to MAPE and RMSE. The results at regional level are rather comparable. This leads to the conclusion that weekly seasonality is even more prevalent as the noise in the time series decreases. Thus, the benefit of considering sales information of all weekdays is reduced. However, the combined model leads to slightly better results than both single ARIMA models. The MAPE for cluster C1 is smaller than for cluster C2 which might be caused by the fact that the sales volumes for buns are higher and more reliable than for breads. Overall, aggregating the demand leads to a higher prediction accuracy (see Fig. 6). At store level the MAPE drops from 20.61 (SA1) to 7.53 (SC1) for buns and from 24.38 (SA2) to 9.30 (SC2) for breads using the combined model. A comparable relative decrease can be measured at regional level (RA vs. RC). This underlines the importance of detecting groups of substitutable articles. Such forecasts are at least a possibility to validate single forecasts, e.g., the sum of the

Table 2
Forecast accuracy of the considered models at different levels.

Level	Baseline		ARIMAX		ARIMA (1-WD)		Combination	
	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
RX	4.72	277.15	4.24	246.79	**3.88**	237.59	3.99	**232.69**
RC 1	4.98	259.59	3.97	207.46	3.81	206.34	**3.74**	**196.90**
RC 2	5.56	41.53	**4.56**	**32.17**	6.56	43.75	4.75	32.39
RA 1	9.92	40.44	9.34	33.87	9.31	33.57	**8.92**	**32.10**
RA 2	11.79	10.64	**9.44**	8.50	11.29	9.93	9.47	**8.44**
SX	8.31	75.71	7.63	67.67	8.34	74.00	**7.44**	**65.65**
SC 1	8.54	69.80	7.69	60.84	8.47	66.72	**7.53**	**59.20**
SC 2	10.83	12.40	9.63	10.42	10.79	11.65	**9.30**	**10.05**
SA 1	23.40	13.75	20.81	11.16	22.07	11.83	**20.61**	**10.93**
SA 2	27.03	3.40	24.46	2.90	26.73	3.11	**24.38**	**2.85**

Table 3
Supply chain performance of the baseline and the combined model at different levels.

Level	Baseline			ARIMA + ARIMAX (1-WD)		
	Fill rate	Loss rate	Service level	Fill rate	Loss rate	Service level
RX	0.977	0.024	0.320	0.982	0.022	0.493
RC 1	0.977	0.027	0.360	0.981	0.019	0.467
RC 2	0.973	0.028	0.514	0.983	0.031	0.583
RA 1	0.956	0.056	0.390	0.962	0.052	0.485
RA 2	0.951	0.069	0.421	0.968	0.067	0.581
SX	0.963	0.046	0.444	0.970	0.044	0.473
SC 1	0.961	0.046	0.413	0.968	0.044	0.484
SC 2	0.956	0.064	0.472	0.968	0.061	0.563
SA 1	0.916	0.150	0.425	0.932	0.142	0.480
SA 2	0.911	0.181	0.368	0.935	0.179	0.544

article forecasts has to be close to the category forecast, and can also be used if the assortment is changing to ensure that the demand for a cluster is matched. The forecasts for all articles (SX, RX) lead also to a high prediction accuracy. However, this is not useful for production planning as those groups contain articles that are not comparable with respect to required raw materials and demand characteristics.

The measures related to the performance of the supply chain are presented in Table 3. We focus on the performance of the combined ARIMA model in the following paragraph. The expected service level of 0.5 is only roughly matched which indicates that the models are slightly biased. The demand for cluster C1 is more often underestimated than overestimated while the opposite is true for cluster C2. However, this does not translate to higher loss rates for articles in cluster C1. The reason for this is that the prediction accuracy for cluster C1 is significantly higher than for cluster C2 (see Table 2). Thus, the loss rates of cluster C1 are also lower than for cluster C2. The achieved fill rates are comparable for both clusters. At SA level the fill rate is slightly above 93% and at SC level close to 97%. As those measures do not take possible substitution effects into account, it can be assumed that the actual fill rate at store level is between 93% and 97%. Following this argument, the actual loss rate at store level ranges between those measured at article level and cluster level, i.e., for cluster C1 (C2) between 14.2% (17.9%) and 4.4% (6.1%). As the measures improve at regional level compared to the store level, it is reasonable to assume that substitution takes place, i.e., the demand is not generally underestimated for all articles within a cluster at certain days. The substitution rates for baked goods are as high as 84% (Van Woensel et al., 2007) which means that the fill rates and loss rates measured at SC level are close to the actual key figures. The results underline the value of the detected clusters. It is possible to limit the amount of discarded goods by relying on substitution and to avoid costly turnover losses by maintaining a high fill rate.

The combined ARIMA model outperforms the baseline with respect to the prediction accuracy and supply chain performance (see Table 4). The relative improvements of MAPE and RMSE range between 10% and 25%. Hence, the predictions provided by the DSS are a noticeable improvement compared to the baseline. Overall, the more accurate predictions translate to a slightly increased fill rate and a noticeable reduction of loss. The only exception is the loss rate for cluster C2 at regional level which is worse than the baseline. However, the loss rate is still low while the fill rate could be improved. On SA level the fill rate could be increased by 1.8% (2.6%) while the loss could be reduced by 5.3% (1.3%) for cluster C1 (C2).

4.2.2. Hierarchical forecasts

We investigate if the hierarchy can be exploited in order to obtain more accurate results or reduced computational costs. Therefore, we compute hierarchical forecasts based on the combinatorial forecasts presented in the previous section (see Table 2). We focus on the demand predictions for articles on store level (SA) and regional level (RA) as those are relevant levels with respect to production planning and deliveries. The bottom-up forecasts are the sum of the forecasts at the lower level. For the top-down forecasts, we compute the allocation proportions based on the observations in the training set. As an example, the proportions that are used to allocate the cluster forecast to the articles at store level are depicted in Fig. 7 for cluster C2 and a specific store. The proportions are rather stable which makes it reasonable to compute them based on historic observations. The (dis-)aggregation ratio, e.g., the number of time series to which a single forecast needs to be distributed (top-down) respectively the number of forecasts that are aggregated (bottom-up), is given in Table 5. For instance, a prediction at the highest level (RX) needs to be allocated to 96 time series at the SA level. The ratios also indicate possible savings with respect to computational costs. For the most extreme scenario (RX → SA) only one model needs to be fitted compared to 96 models at the bottom level.

Overall, the hierarchical forecasts do not improve the accuracy of the direct forecasts (see Table 6). However, some hierarchical

Table 4
Comparison of the baseline and the combined prediction model. The absolute (abs.) as well as relative (rel.) changes are given for the measures MAPE, RMSE, fill rate and loss rate.

Level	MAPE		RMSE		Fill rate		Loss rate	
	abs.	rel.	abs.	rel.	abs.	rel.	abs.	rel.
RX	−0.73	−15.5%	−44.47	−16.0%	0.005	0.5%	−0.002	−8.9%
RC 1	−1.24	−24.9%	−62.69	−24.1%	0.004	0.4%	−0.009	−31.5%
RC 2	−0.80	−14.5%	−9.14	−22.0%	0.011	1.1%	0.003	8.9%
RA 1	−1.00	−10.1%	−8.34	−20.6%	0.006	0.6%	−0.003	−5.6%
RA 2	−2.32	−19.7%	−2.20	−20.7%	0.017	1.8%	−0.002	−2.9%
SX	−0.87	−10.5%	−10.05	−13.3%	0.007	0.7%	−0.002	−4.5%
SC 1	−1.01	−11.8%	−10.60	−15.2%	0.007	0.7%	−0.003	−6.2%
SC 2	−1.53	−14.1%	−2.35	−19.0%	0.012	1.2%	−0.003	−5.3%
SA 1	−2.79	−11.9%	−2.82	−20.5%	0.016	1.8%	−0.008	−5.3%
SA 2	−2.65	−9.8%	−0.55	−16.1%	0.024	2.6%	−0.002	−1.3%

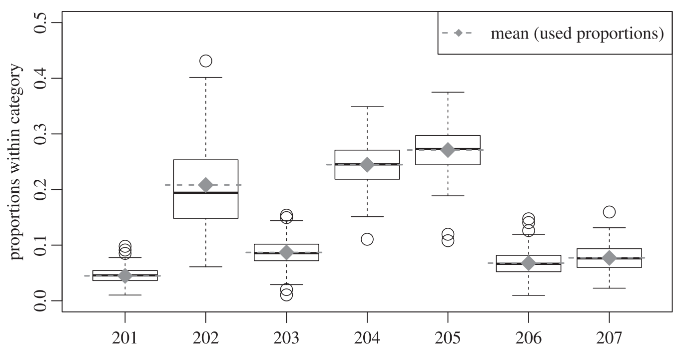

Fig. 7. Demand proportions at store article level in relation to the demand at store category level.

Table 5
(Dis-)aggregation ratio between the layers (top-down: ↓, bottom-up: ↑).

	SA	SC	SX	RA	RC	RX
SA	1	↓7 / 9	↓16	↓6	↓54 / 42	↓96
RA	↑6	–	–	1	↓7 / 9	↓16

forecasts achieve a comparable accuracy with the direct forecasts. We also perform a paired sample *t*-test on the mean difference of the accuracy measures of the direct forecast and each hierarchical forecast approach in order to determine if they differ significantly ($p < 0.05$). This is indeed the case for nearly all hierarchical forecasts. Only the bottom-up forecast for the region article level (RA) are not significantly different from the direct forecasts. However, the statistical significance does not necessarily indicate a practical significant difference as the sample size is quite large (> 1000). The introduced clusters (SC, RC) are suitable for top-down forecasting as they lead to better results than the aggregation of all articles (SX, RX) and are only slightly worse than the direct forecasts. At the SA level it is apparent that deriving the demand based on forecasts at regional level (RA, RC, RX) is worse than relying on the aggregated forecasts for the same organizational level (SC, SX). A reason for this might be that the number of the required allocation proportions is rather high (see Table 5). Moreover, the sales dynamics of the different stores may vary slightly and are not sufficiently (fine-grained) covered in the historic proportions. Hence, a tradeoff between forecast accuracy and computational costs is to forecast the demand at SC level. Based on these forecasts, the predictions at SA level can be calculated with a top-down forecasts. The derived predictions at SA level are then used to compute the RA level predictions bottom-up.

The qualitative results are compatible with those reported in the literature (see Section 2). Top-down predictions can be applied if the forecasts at the top-level and the allocation proportions are sufficiently accurate. In the present use case, the allocation pro-

Table 6
The accuracy measures MAPE and RMSE as well as their standard deviation (SD) of the hierarchical forecasts. Moreover, the p values of a paired sample *t*-test on the mean difference of the accuracy measures of the direct forecast and each hierarchical forecast approach are provided.

target	base	mode	MAPE	MAPE SD	p_{MAPE}	RMSE	RMSE SD	p_{RMSE}
RA 1	SA 1	bottom-up	**8.25**	5.00	0.134	32.54	43.84	0.449
	RA 1	–	8.92	5.88	–	**32.10**	39.47	–
	RC 1	top-down	9.99	5.89	0.004	35.99	42.22	0.0
	RX	top-down	10.13	5.86	0.002	36.89	42.71	0.0
RA 2	SA	bottom-up	9.71	5.36	0.181	**8.33**	5.51	0.118
	RA	–	**9.47**	4.86	–	8.44	5.54	0.032
	RC	top-down	10.77	4.94	0.0	9.67	5.95	0.032
	RX	top-down	12.20	6.16	0.0	10.68	7.07	0.003
SA 1	SA 1	–	**20.62**	40.54	–	**10.93**	16.11	–
	SC 1	top-down	21.95	43.40	0.006	11.22	16.31	0.019
	SX	top-down	21.58	42.20	0.011	11.38	16.51	0.008
	RA 1	top-down	22.98	48.78	0.0	13.53	23.41	0.0
	RC 1	top-down	22.84	45.56	0.0	13.56	23.15	0.0
	RX	top-down	22.77	45.13	0.0	13.56	23.21	0.0
SA 2	SA 2	–	**24.38**	21.48	–	**2.85**	2.03	–
	SC 2	top-down	25.92	21.85	0.0	3.07	2.14	0.0
	SX	top-down	28.01	23.96	0.0	3.23	2.27	0.0
	RA 2	top-down	26.39	21.77	0.0	3.13	2.27	0.0
	RC 2	top-down	26.93	22.38	0.0	3.22	2.33	0.0
	RX	top-down	28.68	24.28	0.0	3.34	2.42	0.0

portions between the SC level and the SA level are rather stable while the prediction accuracy at the top-level is higher than at the bottom-level (see Fig. 6). We also observe that it is important to group time series with similar pattern. For instance, the article clusters at store level (SC) seem to be slightly better suited to predict the demand at the SA level than top-down forecasts from the RA level even though the numbers of allocation proportions are comparable (see Table 5). In general, the forecasts at aggregated levels are quite accurate as the pattern of the lower level time series are similar. Hence, the information loss can be neglected while the noise is reduced. Nevertheless, direct forecasts are also possible in this use case as the items are fast-moving goods. This means that enough information is available at any level as the items are sold in high volume on a daily basis. Thus, bottom-up forecast are comparable to direct forecasts at an aggregated level. However, the advantage of top-down forecasts based on suitable clusters are reduced computational costs and an increased scalability of the system. The relative decrease of required runtime can be derived from Table 5.

4.3. Practical & managerial implications

The proposed approach is implemented at the bakery chain whose data is used for quantitative evaluation. The roll-out of the DSS had various effects on that company. In the past, the forecasts were provided by store managers based on their experience and limited information on the sales history. This is a quite common approach in the fresh food sector (Van Donselaar et al., 2006) even though it is very inefficient. It is time consuming and not reliable as the prediction skills of the store managers differ across the stores. The DSS addresses these issues by providing demand forecasts and limiting the human factor. This leads to reduced decision costs due to time savings and more reliable results. Moreover, the required skill set of the store managers changes as they are no longer required to estimate the future demand. Hence, the DSS provides already a benefit even if it would only match the store managers' performance with respect to the forecast accuracy. Actually, the presented forecasting approach led to an increased product availability as well as to a reduction of discarded goods. In this regard, the aggregated forecasts at cluster level are valuable as they are based on less noisy data and, thus, are more accurate then forecasts at article level. Moreover, the ownership of the forecasts is no longer with the store managers. Instead, dedicated positions at the headquarters were created that are in charge of monitoring the forecasts provided by the DSS. Thereby, they rely in particular on the aggregated forecasts at different levels of the hierarchy.

5. Conclusion and future work

We investigated the applicability of article clustering and hierarchical forecasting as part of a decision support system that enhances the ordering process in order to increase the service level and to limit the amount of discarded goods. Thereby, we focus on perishable fast moving consumer goods that are daily ordered. Traditionally, the replenishment process depends on the experience and skills of the respective store manager as judgmental forecasts are applied for perishable products. This approach is not optimal for retailers running several hundred stores as the prediction accuracy is unreliable which causes regular stock-outs and discarded goods. Both aspects negatively affect the revenue of the retailer. Thus, the proposed system addresses this issue by providing demand forecasts for all articles at store level and regional level. Therefore, we identify article clusters based on intra-day sales pattern. Predicting the demand of clusters is relevant to validate the article forecasts. Based on the article clusters and the organizational structure, we build a hierarchy that can be exploited for hierarchical forecasts. An evaluation based on point-of-sales data of an industrialized bakery chain indicates that the clustering approach works well in order to identify article clusters. The identified clusters contain substitutional articles which is convenient as the substitution rates in case of stock-outs are high for perishable goods. An analysis of the hierarchical forecasts reveals that the clusters are indeed useful to reduce the computational costs without sacrificing the forecasts accuracy.

We restricted the applied methods to ARIMA models as they are applied successful in comparable settings. However, the application and evaluation of other forecasting methods (e.g. neural networks) is part of the next steps as they have other desirable characteristics. With respect to the article cluster, it might also be beneficial to exploit the information about the substitutable articles and substitution rates. For instance, if an article of that group is promoted or discounted the demand for the other articles will be reduced. This distorts the time series of all articles which affects the forecast and needs to be considered by the model. The considered hierarchy ends at the regional level and implicitly assumes that the demand pattern are comparable at all stores. Hence, identifying comparable stores based on their reaction to environmental factors (e.g. weather) is necessary. In the fresh food sector, it also important to provide intra-day predictions as some articles can be produced during the day based on the customer demand. This adds another dimension (time) as well as an additional level to the hierarchy that needs to be evaluated. Moreover, it is important to incorporate the costs of underestimation and overestimation along with demand uncertainty in order to determine optimal order quantities.

Acknowledgments

The authors would like to thank two anonymous reviewers for their valuable comments which helped to improve the paper.

References

Aastrup, J., & Kotzab, H. (2010). Forty years of out-of-stock research–and shelves are still empty. *The International Review of Retail, Distribution and Consumer Research, 20*(1), 147–164.

Adebanjo, D. (2009). Understanding demand management challenges in intermediary food trading: A case study. *Supply Chain Management: An International Journal, 14*(3), 224–233.

Adebanjo, D., & Mann, R. (2000). Identifying problems in forecasting consumer demand in the fast moving consumer goods sector. *Benchmarking: An International Journal, 7*(3), 223–230.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723.

Barratt, M., & Oliveira, A. (2001). Exploring the experiences of collaborative planning initiatives. *International Journal of Physical Distribution & Logistics Management, 31*(4), 266–289.

Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control, revised ed*. Holden-Day.

Büyüközkan, G., & Vardaloğlu, Z. (2012). Analyzing of cpfr success factors using fuzzy cognitive maps in retail industry. *Expert Systems with Applications, 39*(12), 10438–10455.

Campo, K., Gijsbrechts, E., & Nisol, P. (2000). Towards understanding consumer response to stock-outs. *Journal of Retailing, 76*(2), 219–242.

Chang, I., Tiao, G. C., & Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics, 30*(2), 193–204.

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM, 54*(8), 88–98.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36*(4), 1165–1188.

Corsten, D., & Gruen, T. (2003). Desperately seeking shelf availability: an examination of the extent, the causes, and the efforts to address retail out-of-stocks. *International Journal of Retail & Distribution Management, 31*(12), 605–617.

Danese, P. (2007). Designing CPFR collaborations: Insights from seven case studies. *International Journal of Operations & Production Management, 27*(2), 181–204.

Danese, P. (2011). Towards a contingency theory of collaborative planning initiatives in supply networks. *International Journal of Production Research, 49*(4), 1081–1103.

Datta, S., Graham, D. P., Sagar, N., Doody, P., Slone, R., & Hilmola, O.-P. (2009). Forecasting and risk analysis in supply chain management: GARCH proof of concept. In T. Wu, & J. Blackhurst (Eds.), *Managing supply chain risk and vulnerability* (pp. 187–203). Springer London.

De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting, 22*(3), 443–473.

Ehrenthal, J. C., & Stölzle, W. (2013). An examination of the causes for retail stockouts. *International Journal of Physical Distribution & Logistics Management, 43*(1), 54–69.

Eksoz, C., Mansouri, S. A., & Bourlakis, M. (2014). Collaborative forecasting in the food supply chain: A conceptual framework. *International Journal of Production Economics, 158*, 120–135.

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting, 25*(1), 3–23.

Fildes, R., Nikolopoulos, K., Crone, F. S., & Syntetos, A. A. (2008). Forecasting and operational research: A review. *Journal of the Operational Research Society, 59*(9), 1150–1172.

Gardner, E. S. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting, 4*(1), 1–28.

Gardner, E. S. (2006). Exponential smoothing: The state of the artâ;;part ii. *International Journal of Forecasting, 22*(4), 637–666.

Gilbert, K. (2005). An arima supply chain model. *Management Science, 51*(2), 305–310.

Gross, C. W., & Sohl, J. E. (1990). Disaggregation methods to expedite product line forecasting. *Journal of Forecasting, 9*(3), 233–254.

Gruen, T. W., & Corsten, D. S. (2007). *A comprehensive guide to retail out-of-stock reduction in the fast-moving consumer goods industry*. Grocery Manufacturers of America.

Gruen, T. W., Corsten, D. S., & Bharadwaj, S. (2002). *Retail out-of-stocks: A worldwide examination of extent, causes and consumer responses*. Grocery Manufacturers of America.

Hackathorn, R. (2004). The bi watch real-time to real-value. *DM Review, 14*, 24–29.

Hahn, G., & Packowski, J. (2015). A perspective on applications of in-memory analytics in supply chain management. *Decision Support Systems, 76*, 45–52.

Helm, R., Hegenbart, T., & Endres, H. (2013). Explaining costumer reactions to real stockouts. *Review of Managerial Science, 7*(3), 223–246.

Hollmann, R. L., Scavarda, L. F., & Thomé, A. M. T. (2015). Collaborative planning, forecasting and replenishment: A literature review. *International Journal of Productivity and Performance Management, 64*(7), 971–993.

Holsapple, C. W., & Sena, M. P. (2005). Erp plans and decision-support benefits. *Decision Support Systems, 38*(4), 575–590.

Holsapple, C. W., & Whinston, A. B. (1996). *Decision support systems: A knowledge-based approach*. West Publishing Company.

Holweg, M., Disney, S., Holmström, J., & Småros, J. (2005). Supply chain collaboration:: Making sense of the strategy continuum. *European Management Journal, 23*(2), 170–181.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software, 26*(3), 1–22.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688.

Jaipuria, S., & Mahapatra, S. (2014). An improved demand forecasting method to reduce bullwhip effect in supply chains. *Expert Systems with Applications, 41*(5), 2395–2408.

Kahn, K. B. (1998). Revisiting top-down versus bottom-up forecasting. *The Journal of Business Forecasting, 17*(2), 14.

Kaipia, R., Korhonen, H., & Hartiala, H. (2006). Planning nervousness in a demand supply network: An empirical study. *The International Journal of Logistics Management, 17*(1), 95–113.

Kaiser, W. (2011). Fast moving consumer goods. In *Qualitative marktforschung in theorie und praxis* (pp. 605–615). Springer.

Kalchschmidt, M., Verganti, R., & Zotteri, G. (2006). Forecasting demand from heterogeneous customers. *International Journal of Operations & Production Management, 26*(6), 619–638.

Lee, H. L., Padmanabhan, V., & Whang, S. (1997a). The bullwhip effect in supply chains. *MIT Sloan Management Review, 38*(3), 93.

Lee, H. L., Padmanabhan, V., & Whang, S. (1997b). Information distortion in a supply chain: The bullwhip effect. *Management Science, 43*(4), 546–558.

Lee, H. L., & Whang, S. (2000). Information sharing in a supply chain. *International Journal of Manufacturing Technology and Management, 1*(1), 79–93.

Liu, L.-M., Bhattacharyya, S., Sclove, S. L., Chen, R., & Lattyak, W. J. (2001). Data mining on time series: An illustration using fast-food restaurant franchise data. *Computational Statistics & Data Analysis, 37*(4), 455–476.

Plattner, H. (2009). A common database approach for oltp and olap using an in--memory column database. In *Proceedings of the 2009 acm sigmod international conference on management of data* (pp. 1–2).

Power, D. J. (2008). Decision support systems: A historical overview. In *Handbook on decision support systems 1* (pp. 121–140). Springer.

Ramanathan, U. (2014). Performance of supply chain collaboration–a simulation study. *Expert Systems with Applications, 41*(1), 210–220.

Rojas, I., Valenzuela, O., Rojas, F., Guillén, A., Herrera, L. J., Pomares, H., et al. (2008). Soft-computing techniques and arma model for time series prediction. *Neurocomputing, 71*(4), 519–537.

Sahay, B., & Ranjan, J. (2008). Real time business intelligence in supply chain analytics. *Information Management & Computer Security, 16*(1), 28–48.

Shukla, M., & Jharkharia, S. (2013). Applicability of arima models in wholesale vegetable market: An investigation. *International Journal of Information Systems and Supply Chain Management, 6*(3), 105–119.

Sikka, V., Färber, F., Lehner, W., Cha, S. K., Peh, T., & Bornhövd, C. (2012). Efficient transaction processing in sap hana database: The end of a column store myth. In *Proceedings of the 2012 acm sigmod international conference on management of data* (pp. 731–742).

Silver, E. A., Pyke, D. F., & Peterson, R. (1998). Style Goods and Perishable Items. In *Inventory management and production planning and scheduling chapter 10* (pp. 382–420). New York: John Wiley and Sons.

Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research, 252*(1), 1–26.

Syntetos, A. A., Nikolopoulos, K., & Boylan, J. E. (2010). Judging the judges through accuracy-implication metrics: The case of inventory forecasting. *International Journal of Forecasting, 26*(1), 134–143.

Thomé, A. M. T., Hollmann, R. L., & Scavarda do Carmo, L. F. R. R. (2014). Research synthesis in collaborative planning forecast and replenishment. *Industrial Management & Data Systems, 114*(6), 949–965.

Van Donselaar, K., van Woensel, T., Broekmeulen, R., & Fransoo, J. (2006). Inventory control of perishables in supermarkets. *International Journal of Production Economics, 104*(2), 462–472.

Van Donselaar, K. H., Gaur, V., Van Woensel, T., Broekmeulen, R. A., & Fransoo, J. C. (2010). Ordering behavior in retail stores and implications for automated replenishment. *Management Science, 56*(5), 766–784.

Van Woensel, T., Van Donselaar, K., Broekmeulen, R., & Fransoo, J. (2007). Consumer responses to shelf out-of-stocks of perishable products. *International Journal of Physical Distribution & Logistics Management, 37*(9), 704–718.

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236–244.

Watson, H. J. (2009). Tutorial: business intelligence-past, present, and future. *Communications of the Association for Information Systems, 25*(1), 39.

Widiarta, H., Viswanathan, S., & Piplani, R. (2009). Forecasting aggregate demand: An analytical evaluation of top-down versus bottom-up forecasting in a production planning framework. *International Journal of Production Economics, 118*(1), 87–94.

Williams, B. D., & Waller, M. A. (2011). Top-down versus bottom-up demand forecasts: The value of shared point-of-sale data in the retail supply chain. *Journal of Business Logistics, 32*(1), 17–26.

Zhang, G. (2012). Neural networks for time-series forecasting. In *Handbook of natural computing* (pp. 461–477). Springer Berlin Heidelberg.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14*(1), 35–62.

Zinn, W., & Liu, P. C. (2001). Consumer response to retail stockouts. *Journal of Business Logistics, 22*(1), 49–71.

Zinn, W., & Liu, P. C. (2008). A comparison of actual and intended consumer behavior in response to retail stockouts. *Journal of Business Logistics, 29*(2), 141–159.

Zotteri, G., Kalchschmidt, M., & Caniato, F. (2005). The impact of aggregation level on forecasting performance. *International Journal of Production Economics, 93-94*, 479–491. Proceedings of the Twelfth International Symposium on Inventories