To appear in the *International Journal of Production Research*
Vol. 00, No. 00, 00 Month 20XX, 1–28

A Reliable Decision Support System for Fresh-Food Supply Chain Management

G. Dellino[a,b*] , T. Laudadio[a], R. Mari[a], N. Mastronardi[a] and C. Meloni[a,b]

[a]*Istituto per le Applicazioni del Calcolo 'M. Picone', Consiglio Nazionale delle Ricerche, Bari, Italy*;
[b]*Dipartimento di Ingegneria Elettrica e dell'Informazione, Politecnico di Bari, Bari, Italy*

(*April 13, 2018*)

The paper proposes a Decision Support System (DSS) for the supply chain of packaged fresh and highly perishable products. The DSS combines in a unique tool sales forecasting with order planning which includes an individual model selection system equipped with ARIMA, ARIMAX and transfer function forecasting model families, the latter two accounting for the impact of prices. Forecasting models parameters are chosen via two alternative tuning algorithms: a two-step statistical analysis, and a Sequential Parameter Optimization framework for automatic parameter tuning. The DSS selects the model to apply according to user-defined performance criteria. Then it considers sales forecasting as a proxy of expected demand and uses it as input for a multi-objective optimization algorithm that defines a set of non-dominated order proposals with respect to outdating, shortage, freshness of products and residual stock. A set of real data and a benchmark—based on the methods already in use—are employed to evaluate the performance of the proposed DSS. The analysis of different configurations shows that the DSS is suitable for the problem under investigation; in particular, the DSS ensures acceptable forecasting errors and proper computational effort, providing order plans with associated satisfactory performances.

Keywords: fresh food supply chain; forecasting; order proposal; optimization; decision support systems

1. Introduction

The fresh food supply chain management has experienced great changes over the last years, and it has now become the major strategic issue for food firms (Bourlakis and Weightman 2004). In this context, one of the main goals for a company is to individuate a combination of purchasing, transportation, physical distribution and logistics to get a position to achieve economies of scale; the control of the supply chain tends to shift from producers to retailers, while the increased pressure for higher quality and cost efficiency affects all members of the food chain.
Consequently, the food chain is called to move toward a more vertically integrated structure that includes joint partnerships, strategic alliances, and more vertical co-ordination among different supply chain players; and retailers are asking for new value-added logistics services. In fact, traditional logistics firms cannot always fully meet the demands of the retailers, because they did not possess the adequate logistics and information technology solutions. Thus retailers started to co-operate also with logistics information technology suppliers as new actors in the field. The resulting value-added services increase the operational complexity of the food retail chain.
Today's food supply chain are required to be reliable and agile enough to respond to consumer demand and preferences and it is expected that e-logistics practices and e-commerce will have a growing relevance to support a customer-focused and more responsive organization. Sales forecasts and order planning play a central role in this customer and data centered process. They constitute a process that is dual to the physical process of delivering goods from the producer to the retailer. The whole process can be viewed as an integrated demand and supply chain management (Dotoli et al. 2005) which integrates both the pull action from

*Corresponding author. Email: gabriella.dellino@poliba.it

the customer's demand and the push effect from the producer and the retailer in the form of promotions and advertising actions.
The forecasting techniques to be employed on the store level are called to fulfill the requirements of the so called micro-forecasting (i.e., a specific forecast for every item in each store), while aggregation across items and/or across stores is considered not meaningful or even misleading. The order plan per store must be based on the individual sales forecast for the individual store. Since calendar events, advertising, and promotional events may have very different effects in the different stores depending on the respective locations, the forecasting techniques should not be based only on the observed time series of sales but should also be able to take this information about external factors into account also including the possibility of judgmental adjustments (Wang and Petropoulos 2016). To deal with these aspects, collaborative planning and forecasting approaches can be adopted involving different actors of the supply chain organization (Arminger 2004; Du et al. 2009; Yang et al. 2017).
Forecasts for sales and orders for stores are usually based on short-term forecasts, ranging from daily to between one and four weeks depending on the adopted delivery plans. At supply chain level and/or in a vendor/retailer collaboration (Borade and Sweeney 2015; Yang et al. 2017), to forecast the large number of items and store combinations, as required for demand planning, some automatic computer-assisted ordering systems are needed to process data and execute the recurrent computations (Tan et al. 2017).
Clearly, the efficiency and effectiveness of quantitative methods for optimizing supply chain operations strictly depend on the quality of available data. For this reason, sales forecasting represents the first crucial step in this integrated process and may influence the development of a realistic supply chain model. Sales forecasting accuracy typically has close interactions with inventory management performance. In particular, in a fresh and highly perishable food supply chain, integrating forecasting and optimization is even more important: because the shelf life of products can be very limited, reliable order proposals are fundamental to manage and reduce inefficiencies such as stock-out and outdating (Ma et al. 2013; Nahmias 2011). The fresh food supply chain represents a very challenging application, characterized by several interrelated variables and constraints, and possible sources of uncertainties: it requires an effective management of logistics operations, and perishability is a particularly critical issue (Alvarez and Johnson 2011; Nahmias 2011). This calls for integrated and flexible approaches able to properly account for the characteristics of customers, retailers, and producers in order to meet their needs.
This paper proposes methods and models devoted to fresh food supply chain operations management to be included in a unique software tool to form a Decision Support System (DSS), and presents a case study involving a set of retailers with both small and medium sized stores located in the Apulia region (Italy). Both forecasting and optimization are crucial issues in such a context and the main criticalities are related to the uncertainty on future sales. In the considered context, a detailed forecast, i.e. forecast by product or even by product-location, is necessary. In fact, the retailer needs sales forecast to determine how much of each product to deliver or to stock at the different retail locations. We focus on this situation that drives the planning of the considered supply chain. The considered forecasting process uses sales data, such as point-of-sale scanner data, and promotion and marketing plans. These information flows also show the collaborative nature of the forecasting activity and its integration in the decision-making process. The demand forecasts become a key input to the order planning function. The fresh food supply chain is usually characterized by a high number of products, where each item may account for only a small fraction of both the total volume and revenue. A relatively cheap and efficient way to repetitively forecast the sales of each of a large number of items is to adopt an automatic forecasting system. Otherwise, it could seem neither practical nor economical to spend a lot of effort in forecasting single items, except for the few best-selling products. These forecasting issues are addressed by an integrated, flexible and simple-to-use tool proposed to support the decision maker to determine, within a vendor/retailer collaboration, reliable operations plans with respect to waste reduction and other different criteria, such as shortage, freshness and residual stock of products.

1.1 *Literature review*

The literature clearly shows how the integration of sales forecasting and planning activities plays a crucial role in the market of fresh and highly perishable food: in fact, the shelf life of products is very limited, therefore reliable forecasts are essential to reduce inefficiencies (Nahmias 2011; Stevenson 2015; Wagner 2002; Wagner and Meyr 2002). In particular, we refer to the analyses proposed by Fildes and Petropoulos (2015); Leung and Bhaskaran (2011) and Wagner (2002) on the role and functions of forecasting tools in order to give adequate support to decision makers. These analyses point out that the selection of a forecasting model and the estimation of its parameters are fundamental issues in the configuration of a demand planning system or during the update of forecasting parameters. These activities should be made regularly providing some kind of automatic software tool to search all available statistical forecasting procedures and parameter combinations and select the one which produces the best forecast accuracy in the period of interest. More specifically, selecting the appropriate individual (i.e., for each item or even each item-location) forecasting method for a large number of time series is considered as a crucial problem for many demand forecaster. In addition, the forecast performance of such a system may deteriorate over time if they are not adequately diagnosed and tuned. These studies also observe that although dedicated software packages usually provide some parameter optimization capability, forecasting modules of many integrated supply chain management systems do not provide such a feature or it is very limited. The survey of McCarthy et al. (2006) offers a guide to expand the knowledge of understanding the importance of performance indicators when a decision maker is called to choose, managing or design a forecasting system.

This paper focuses on the problem of sales forecasting and order planning integration through a DSS, and mainly refers to base forecast generated from historical data (i.e., time series) captured in a continuous planning and monitoring system. It is important to observe that final adjustments could be made to improve the base forecasts considering several factors that have not been part of historical data (Eksoz et al. 2014; Goodwin 2002; Wang and Petropoulos 2016). For instance, recent changes in the market, weather conditions, competitors' plan or last minute changes at retail or supply side are among the factors that should be considered in the order planning process through an experts' adjustment of base forecasts (Lawrence et al. 2006; Syntetos et al. 2015). In a fresh food supply chain, this aspect could much more relevant due to demand variability, risk of expiry, freshness and sales miss. How to implement these judgmental adjustments (which cannot be underestimated in the fresh food industry) and the analysis of their impact are outside the scope of this paper and could be addressed in further research works.

Different mathematical modeling approaches have also been proposed to offer quantitative methods to decision makers. A review of the state-of-the-art in the area of planning models for the different components of agri-food supply chains is offered in Ahumada and Villalobos (2009). The links between sustainability in food supply chains and quantitative methods to support the decision makers are analyzed in detail by Beske et al. (2014) and Soysal et al. (2012). Indeed, fresh food production and distribution potentially generate considerable waste through poor planning of operations for both packaged and unpackaged items (Brandenburg et al. 2014). This paper deals with the first case in which each item has its own nominal shelf life which can be considered as a constant, while in the case of unpackaged products the planner has to face with a random lifetime (Kouki et al. 2014).

The problem of determining the optimal economic operating policy when a number of non-instantaneous deteriorating items are jointly replenished has been recently addressed by Ai et al. (2017). While analytical methods to reduce the overproduction wastes in the convenience food production are proposed in Darlington and Rahimifard (2006) and Darlington et al. (2009). The optimization issues in the supply chain composed of retailers and potential recipients that practice food recovery are addressed in Aiello et al. (2014); Muriana (2015). Wang et al. (2009) develop approaches to integrate traceability initiatives with operations management objectives for perishable food products. While Van Der Vorst et al. (1998) investigates the effects of supply chain management on logistical performances in food supply chains, showing the crucial role of the reduction (or elimination) of uncertainties to improve the overall behavior of the chain. In such a context, a robust supply chain operations management can only be obtained by taking uncertainties of future demand into account; for this reason a good and reliable forecasting plays a crucial role (Dellino et al. 2010, 2012; Fleischmann et al. 2002; Simangunsong et al. 2012). At this aim, an extended version of the classical newsvendor problem has been proposed by Huang (2013), to account for specific issues related

to random demand and item deterioration over time. van Donsellar et al. (2006) investigate inventory control policies for perishable items in supermarkets, providing directions for improving the automated store ordering system currently in use in two Dutch supermarket chains. Moreira and Tjahjono (2016) study the beverage industry decision-making process at an operational level, with the aim to increase supply chain flexibility to satisfy fluctuating customer demand quantitatively considering the impact on the supply chain. The close interactions between demand forecasting and inventory management performance are addressed by Babai et al. (2013) focussing on their impact on both inventory costs and service levels. Galasso et al. (2009) propose a mixed integer linear planning model embedded in a framework simulating a rolling horizon planning process in order to assist the decision makers for coping with an uncertain or flexible demand, while Kanet et al. (2010) show how the implementation of dynamic planned safety stocks can reduce unnecessary safety stocks and improve service in supply networks. The study conducted by Rijpkema et al. (2014) shows that, in perishable product supply chain design, a trade-off should be defined between transportation costs, shortage costs, inventory costs, product waste, and expected shelf life, suggesting to adopt a multi-criteria approach. However, Yakovleva et al. (2012) point out that experts give considerably different relative weights to various elements of sustainability in the supply chain, while Kaipia et al. (2013) observe that the sustainability (in terms of waste reduction) of the perishable food chain needs more efficient information sharing.
The approach presented in this paper refers to the operational level of a supply network in which the supply configuration (i.e., actors, products and supply policies) is fixed at a prior decision phase and the supply management is conducted in co-operation by suppliers and retailers. Thus the considered models rely only on four KPIs accounting for waste, freshness, stock-out and residual stock. Indeed, in the fresh food business these are among the ones that are used in practice. However, the previous discussion and the relevant literature suggest that for further research, in a more general context or for more integrated approaches, it is worthwhile to enrich the set of KPIs to consider customer service and supply chain costs, risks, uncertainties, and sustainability issues (Abbou et al. 2017; Diabat et al. 2012; Simangunsong et al. 2012; Yakovleva et al. 2012).

1.2 *Contribution*

In this paper, we address the problem of sales forecasting and order planning integration in a unique software tool, which is framed within the general problem of supply chain management, for a set of fresh and highly perishable packaged food products. This goal is reached by designing a modular and reliable DSS, whose main building blocks are the following: the first one addresses sales forecasting adopting an automatic individual model selection approach; the information about sales forecasts is then used as input for a multi-objective optimization algorithm to define the best order policy. Data, coming from a set of small and medium sized retailers operating in Apulia region, Italy, collected for model estimation are pre-processed for a two-fold reason: identify seasonality and remove noise. Following an individual model selection approach, three different forecasting model families are considered; besides, the DSS has a modular architecture, so it is open to include other model families or consider possible judgmental adjustments on forecasts. The first family is one of the best known classes of mathematical models for time series forecasting represented by the Autoregressive Integrated Moving Average (ARIMA) models (Box et al. 2008). ARIMA models are widely used in statistics, econometrics and engineering for several reasons: (i) they are considered as one of the best performing models in terms of forecasting, (ii) they are used as benchmark for more sophisticated models, (iii) they are easily implementable and have high flexibility due to their multiplicative structure. Nevertheless, they do not take the effect of exogenous variables into account. In the case study under investigation, sales of fresh goods are influenced by prices and the impact of the latter on the forecasting process should be considered. Several alternative approaches can be found in the literature that make forecasting more robust and reliable by including the effect of exogenous variables. The easiest approach is to adapt ARIMA models to account for the aforementioned variables, obtaining the so-called ARIMAX models (Box et al. 2008). Another common technique is based on the identification of a transfer function (TF) relating the time series of the variable of interest with the one of the exogenous variable (Makridakis et al. 2008). We investigated the performance of these model families on our data set. Specifi-

cally, given a time series, for each family the best model is identified and estimated by varying parameters in predefined intervals. The best parameter setting for each family can be selected according to two alternative approaches. The first is a grid search based tuning framework that adopts a set of statistical indicators. The second is the Sequential Parameter Optimization (SPO) framework proposed by Bartz-Beielstein et al. (2010): it makes the parameter tuning automated according to modern statistical techniques based on meta-models' construction and design of experiments. The SPO framework is very general and it is considered an efficient and effective tuning procedure in a variety of applications (Bartz-Beielstein et al. 2010).
The DSS includes a module devoted to automatically operate an individual (i.e., item-location) forecasting model selection—on the basis of some general user-defined criteria—the tuned forecasting model, among those preliminarily chosen for each family, which provides the input data (i.e., the forecasted demand) for the multi-objective optimization method. The latter is designed to be user-interactive and to provide a Pareto front of optimal order proposals according to some crucial Key Performance Indicators (KPIs) for packaged fresh and perishable products such as outdatings, stock-out and freshness of goods in the context of a pre-configured supply network in which the planning activities are conducted in co-operation by suppliers and retailers. The modular structure of the proposed DSS enables to easily consider different KPIs of interest for different research works and applications. We propose to compute the order proposal through a meta-heuristic approach, based on a genetic algorithm, that considers the forecasting sales as a proxy of demand. Then, the user (i.e., the manager) specifies the criteria for selecting an order plan among the pool of non-dominated solutions identified by the optimization module. Therefore, the DSS provides the order proposal to be implemented on the basis of the manager preferences. Demand uncertainty is also taken into account through a simulation analysis to assess the impact of demand perturbations on the optimal order quantity. A scheme describing the proposed DSS is displayed in Figure 1. As described above, each

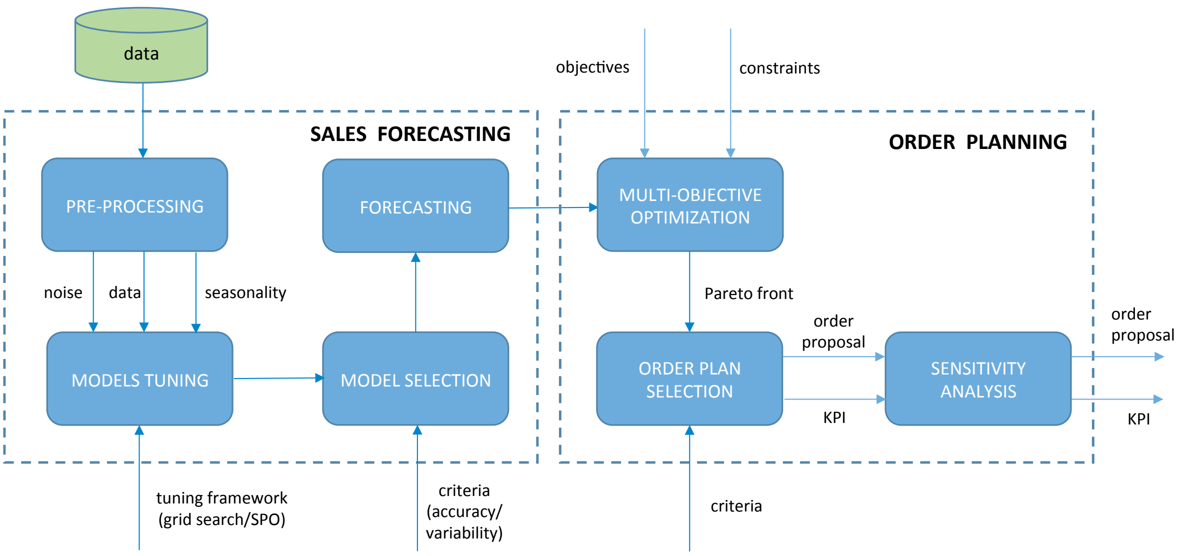


Figure 1.: DSS.

building block implements specific approaches and algorithms and their integration into a user interactive system is an innovative contribution in order to provide a well defined DSS that combines the order planning problem with sales forecasting, making it tailored for real world applications.
The paper is organized as follows. Section 2 details the building blocks of the proposed DSS, which combines pre-processing, alternative tuning procedures for the different forecasting model families, statistical indicators for model selection, and order planning. Next, computational results are discussed in Section 3: specifically, we investigate the DSS performance under different alternative configurations based on possible user preferences. Intermediate forecasting results are analyzed and also compared to some models belonging to the exponential smoothing family (Makridakis et al. 2008). The latter represents a robust approach commonly used in practice in this area, often implemented in the forecasting and supply chain

planning systems available commercially (Fildes and Petropoulos 2015; Leung and Bhaskaran 2011). They are previously adopted by the retailers participating in the case study for sales forecasting, and taken here as benchmark. Finally, conclusions are drawn in Section 4.

2. The DSS structure

In the following we discuss the methods adopted in the proposed DSS for order management, as illustrated in Figure 1. Specifically, in Section 2.1 we first describe the data pre-processing step of the forecasting module. Then, we use the pre-processed data as input for the model tuning block, which requires the user to specify the tuning approach (either grid search or SPO based). This block is composed of an automatic tuning procedure and a statistical analyzer to select the best model. Three forecasting techniques (ARIMA, ARIMAX and TF) are described in Section 2.2, while tuning procedures and indicators are reported in Section 2.3. The module devoted to the dynamic selection of the forecasting model to use in the current planning activity is described in Section 2.4. The order planning module, consisting of a multi-objective optimization block and an order plan selection block, is finally discussed in Section 2.5.

2.1 *Pre-processing*

Data, coming from a set of small and medium sized retailers operating in Apulia region, Italy, are collected for model estimation and need to be pre-processed for a two-fold reason: identify seasonality and remove noise.

Sales are normalized as follows. Let y_t be the quantity of product P sold in store V at time t. Then, the corresponding normalized quantity z_t is

$$z_t = \frac{y_t - \mu}{\sigma}, \tag{1}$$

where μ and σ are the sample mean and the standard deviation, respectively, of time series y_t over the data set under investigation.

Seasonality typically emerging from historical sales can be detected by applying an Independent Component Analysis (ICA) (Hyvärinen and Oja 2001), a well known Blind Source Separation (BSS) technique widely used in signal and image processing to extract the most significant information from the given data. In our study, we assume that the sales time series are linear combinations of several mutually independent components. ICA finds these components by minimizing the mutual information between the unmixed time series. Before applying ICA, a pre-processing step is required in order to get the major advantage. This step is called pre-whitening and consists in uncorrelating the time series and reducing the data dimensionality; i.e., pre-whitening performs a Principal Component Analysis (Jolliffe 2002) and discards the principal components with lowest variance, which are associated to the noise affecting the sales data. Based on the impact of noise on historical sales, the pre-processing module might filter this noise in case it represents a significant component of the original time series.

2.2 *Sales forecasting*

Concerning the model selection procedure each time-series is decomposed in three distinct time-segments: a) training set, used for training the considered models; b) validation set, used to compare the methods and select the best performing one; and c) test set, which provides the forecasting horizon where the selected models are tested to get forecasts after refitting over all samples from both training and validation sets. As mentioned in Section 1, we adopt three time series forecasting model families: ARIMA, ARIMAX and TF. The latter two methods account for prices' effect on sales.

2.2.1 ARIMA models

Let z_t be the realization of a stochastic process at time t, that is an observation of time series at time t, and let a_t be a random variable with normal distribution, having zero mean and variance equal to σ_a^2. Thus, the random variable a_t represents the realization at time t of a white noise process. An ARIMA model with *seasonality* is defined as follows:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla^D z_t = \theta_q(B)\Theta_Q(B^s)a_t \,, \tag{2}$$

where B is the backward shift operator which is defined by $Bz_t = z_{t-1}$ and

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 \ldots - \phi_p B^p \,, \tag{3}$$

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} \ldots - \Phi_P B^{Ps} \,, \tag{4}$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 \ldots - \theta_q B^q \,, \tag{5}$$

$$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} \ldots - \Theta_Q B^{Qs} \,, \tag{6}$$

$$\nabla^d = (1-B)^d \,, \tag{7}$$

$$\nabla^D = (1-B^s)^D \,. \tag{8}$$

The parameter p defines the order of the autoregressive non-seasonal component, q defines the order of the moving average non-seasonal component, and the parameter d represents the order of non-seasonal integration necessary to obtain a stationary time series. The parameters p, q and d are commonly used for referring to non-seasonal models in a concise way. For more complex models, in which a similar pattern at regular time intervals is observed, it is more realistic to take seasonality under consideration. A data set comprised of food sales on large-scale stores is typically affected by a weekly seasonality, which reflects the customers' habit to buy foods especially on the weekend. The seasonal component is defined by the parameters P, D and Q, where P defines the order of the seasonal autoregressive component, Q defines the order of the seasonal moving average component, and D is the order of seasonal differences. Finally, s defines the time series' seasonality. A seasonal ARIMA model is synthetically described as ARIMA$(p,d,q) \times (P,D,Q)_s$. Model tuning and parameters' identification is performed by the module discussed in Section 2.3. A disadvantage of classical ARIMA models is that the effect of exogenous variables on data is not taken into account. In the following sections we present the alternative forecasting model families included in our DSS.

2.2.2 ARIMAX models

ARIMA model with exogenous variables, also referred to as ARIMAX, can be defined as follows:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla^D z_t = \theta_q(B)\Theta_Q(B^s)a_t + \beta x_t \,, \tag{9}$$

where x_t is the vector of exogenous variables and β is the vector of regression coefficients. The latter has to be estimated and its initial value is set equal to the canonical correlation between z_t (series of sales) and x_t (series of prices). According to Makridakis et al. (2008), defining β as a regression coefficient is not appropriate, so they propose an alternative formulation of the ARIMAX model. As the difference between these two models is not remarkable, we implement the ARIMAX forecasting model (9) as it is more similar to the classical ARIMA model; i.e., the two models preserve a similar theoretical structure. The best ARIMAX model for a given time series will be selected by the procedures discussed in Section 2.3.

2.2.3 TF Models

TF models are based on the assumption that the relation between time series and exogenous variables can be modeled by a TF (to be estimated) plus an error vector described by an ARIMA model. More formally,

$$z_t = \frac{\omega(B)B^b}{\delta(B)} x_t + n_t \,, \tag{10}$$

where the TF is defined by v zeros, r poles and a delay b, with

$$\omega(B) = \omega_0 - \omega_1 B - \omega_2 B^2 \ldots - \omega_v B^v \,, \tag{11}$$

$$\delta(B) = 1 - \delta_1 B - \delta_2 B^2 \ldots - \delta_r B^r \,, \tag{12}$$

and the error vector n_t is described by the following ARIMA model

$$\phi_p(B)\Phi_P(B^s)\nabla^d \nabla^D n_t = \theta_q(B)\Theta_Q(B^s) a_t \,. \tag{13}$$

Unlike the previous two models, in this third one additional parameters have to be estimated, namely parameters v, r and b. Therefore, a TF model will be described by a tuple $(p,d,q) \times (P,D,Q)_s \times (v,r,b)$, whose values will be identified as discussed in Section 2.3.

2.3 *Model tuning and selection*

For each forecasting model family, given a data set (i.e., an item-store sales time series), the best forecasting model to use has to be identified. In the literature, diverse approaches are proposed for accomplish this task (e.g., see Box et al. (2008)). Our DSS implements two procedures, namely a grid search (GS) and the SPO approach, which are discussed in the following sections 2.3.2 and 2.3.3, respectively. The user is required to specify in the system configuration the identification approach to apply for model tuning (the default choice is SPO), together with the forecasting families to consider for the pre-processed input data (the default choice is ARIMA). Both GS and SPO procedures use a set of statistical indicators which will be introduced in the following section.

2.3.1 Statistical Indicators

In this section we analyze the tuning and selection phase in more detail and provide a complete description of the statistical indicators adopted to identify the best model for each family in the forecasting module of our DSS. As the tuning and selection procedure refers to the training set and the validation set (previously introduced in Section 2.2), these indicators can be divided into two groups, in-sample (referring to the training set) and out-of-sample indicators (referring to the validation set). We perform the so-called *in-sample analysis*, working on the training set, to identify a set of non-dominated models with respect to the first group of indicators. Then, in order to finalize the search for the most accurate model, we perform the so-called *out-of-sample analysis* on the validation set, which is used to compare forecasts to observed data, and to assess the effectiveness of the forecasting model in order to individuate the best representative for each considered forecasting model family.

In-sample indicators. This subset includes indicators that are computed on the training set. These are mostly used as lack of fit measures, based on the information entropy and parsimony of models. Thus, the objective of in-sample analysis is to measure the matching between real data and simulated data obtained by the mathematical model under analysis.
We compute two different indicators: the Ljiung-Box test and the Hannan-Quinn Information Criterion (HQC) (Box et al. 2008; Burnham and Anderson 2002). The Ljiung-Box test is a statistical test in which the null hypothesis is that the autocorrelations of the residuals are assumed to follow a white noise process.

HQC is a well known criterion used to quantify the entropy of the information and the information lost in the fitting process. It can be used under the assumption that the residuals are independent and identically distributed, and it tends to penalize lack of parsimony. The interested reader is referred to Burnham and Anderson (2002) for a detailed description of these (and others) in-sample indicators.
For the sake of implementation of our forecasting models, the proposed indicators are used as follows: a non-domination analysis is performed with respect to variance, residuals and HQC for all the models obtained by changing the tuple of parameters, e.g., $(p,d,q) \times (P,D,Q)_s$ for ARIMA models. The dominated models, as well as models not satisfying the Ljung-Box test, are excluded by the next analysis, while the remaining models undergo the out-of-sample analysis.
Note that the joint use of both the Ljiung-Box test and the HQC indicator is relevant from a theoretical standpoint. In fact, the uncertainty associated to the forecast provided by the model strictly depends on two assumptions: (i) the time series model describing data is correctly identified and (ii) the residuals follow a normal distribution; see Chatfield (2001). In this respect, the HQC indicator supports the selection of the best model (see e.g. Hyndman and Khandakar (2008)), while Ljiung-Box test guarantees the normal distribution of residuals.

Out-of-sample indicators. The out-of-sample indicators are well known statistical indicators for quality and accuracy of forecasting (Makridakis et al. 2008; Makridakis and Hibon 2000; Armstrong 2001). Specifically, we use them to compare forecast data and real data within the validation set and to make a quantitative comparison among different models in terms of quality of forecast, working on the models passing the previous in-sample analysis.
The first group of indicators includes absolute measures: they are scale dependent and for this reason can only be used when computed on the same data set, but they cannot be used to compare the behaviour of a forecasting model on different data sets. Let us define f_t as the forecast of quantity of product P (indicated as z_t) sold in store V at time $t \in \{1,\dots,n\}$, where n denotes the size of the validation set. We compute the following indicators:

- Root Mean Squared Error (RMSE): $\sqrt{\frac{1}{n}\sum_{t=1}^{n}(z_t - f_t)^2}$,
- Mean Absolute Error (MAE): $\frac{1}{n}\sum_{t=1}^{n}|z_t - f_t|$,
- Maximum Absolute Error (MaxAE): $\max_{t=1,\dots,n} |z_t - f_t|$.

The second set of indicators is comprised of relative measures that are scale independent. We computed the following indicators:

- Mean Absolute Percentage Error (MAPE): $100 \cdot \frac{1}{n}\sum_{t=1}^{n}\left|\frac{z_t - f_t}{z_t}\right|$,
- Maximum Absolute Percentage Error (MaxAPE): $\max_{t=1,\dots,n}\left|\frac{z_t - f_t}{z_t}\right|$,
- Coefficient of determination R^2: $1 - \frac{\sum_{t=1}^{n}(z_t - f_t)^2}{\sum_{t=1}^{n}(z_t - \mu)^2}$,

where μ is the average value of z_t over the validation set. On the one hand, they allow to compare the performance of the same model on data sets with different scales; on the other hand, MAPE and MaxAPE are not defined for time t in which the denominator is equal to zero. Thus, these indicators should be used with caution in case of either missing data or data too close to zero. To this aim, we replace MAPE by the following indicator:

- Mean Absolute Scaled Error (MASE): $\frac{T-1}{n} \cdot \frac{\sum_{t=1}^{n}|z_t - f_t|}{\sum_{j=2}^{J}|\zeta_j - \zeta_{j-1}|}$

proposed by Hyndman and Koehler (2006) to overcome the main issues related to the other measurements (Hyndman 2006; Franses 2016). In particular, this indicator scales the absolute error (on the validation set) of the forecasting model under study by the error on the training set (indexed with $j \in \{1, \ldots, J\}$) associated to the naïve forecast, which takes the actual value of one period (ζ_j) as the forecast for the next period. For seasonal time series, this indicator will be adapted using seasonal naïve forecasts and, in this case, indicated as Seasonal MASE (SMASE).

Finally, in order to assess the precision of forecast within the validation set, prediction intervals are computed. They can be defined as an upper and a lower bound between which the future forecasted value is expected to lie with a defined probability (Chatfield 2001). Thus, the $100(1-\alpha)\%$ prediction intervals are computed as:

$$f_t(h) \pm g_{\alpha/2} \cdot \sqrt{Var[z_{t+h} - f_t(h)]}\,, \tag{14}$$

where $g_{\alpha/2}$ is the α-percentile point of the normal distribution of forecasting errors (modeled as white noise), $f_t(h)$ represents the h steps ahead forecast made at time t and $Var[z_{t+h} - f_t(h)]$ is the variance of the h steps ahead forecast error made at time t; α is usually set at 5%. According to Chatfield (2001), for time series forecasts based on ARIMA the variance of the forecast error is the Mean Squared Error (MSE), under the assumption that the model is properly identified and that the residuals are normally distributed. As stated above, the latter conditions are guaranteed by the in-sample indicators. Formally, the h steps ahead MSE is defined as the expected square loss, $E[(z_{t+h} - f_t(h)|z_t, x_t)^2]$, conditional on time series z_t up to time t and to exogenous components x_t, if they are included in the forecasting model. MSE provides information on the forecast precision. Indeed, a forecast may be very accurate in terms of out-of-sample indicators, meaning that there is a good matching between real and forecast data, nevertheless a high MSE reveals that the forecast may vary within a wider prediction interval. That is, MSE is an indicator of the forecast dispersion and this feature has been taken into account by including MSE as an additional out-of-sample indicator.

2.3.2 Grid Search approach

We propose a grid search approach in which each model parameter ranges in predefined intervals; see, for instance, Höglund and Östermark (1991). The main advantage of this approach is that an extremely large set of combinations, i.e., forecasting models, can be compared and the resulting model may be very accurate. On the contrary, the larger are the parameters' intervals the higher is the required computational effort, since a set of statistical tests and analyses have to be performed for every considered forecasting model.

In the grid search approach for ARIMA and ARIMAX, we consider all possible combinations of parameters p, d, q, P, D and Q ranging in predefined intervals, while seasonality s is set based on results from the preprocessing phase. For each tuple $(p,d,q) \times (P,D,Q)_s$ the maximum likelihood principle is adopted for model parameters' estimation. Then the analysis of the forecasting model is conducted by means of two kinds of statistical indicators, in-sample and out-of-sample, that are used to determine the best model. Such indicators have been described in Section 2.3.1. For TF, the grid search is implemented as follows: parameters v and r range in predefined intervals, while parameter b is estimated according to the maximum likelihood principle. In order to choose the best value of unknown parameters, for each pair (v,r) the simulated time series $\hat{z}_t$ is defined as

$$\hat{z}_t = \frac{\omega(B)B^b}{\delta(B)} x_t\,. \tag{15}$$

The original time series z_t is compared to $\hat{z}_t$ and the following goodness of fit measure F is computed:

$$F = 100 \cdot \left(1 - \frac{\|z - \hat{z}\|}{\|z - \mu\|}\right)\,, \tag{16}$$

where μ is the average value of z_t over the training set. The pair (v,r) providing the best fitting F is selected and the error vector is computed as $n_t = z_t - \hat{z}_t$. This component is described by an ARIMA model whose parameters are estimated through the grid search approach proposed above for ARIMA models.

2.3.3 SPO approach

The main drawback of the GS approach is that, regardless the forecasting model to be tuned, computational cost for determining the best parameter setting may be very high, thus it is necessary to let parameters range within intervals of limited size. In order to pursue efficiency in tuning the forecasting models, we embed an automatic tuning procedure in our DSS, based on the SPO framework proposed by Bartz-Beielstein et al. (2010). SPO is a heuristic that combines classical and modern statistical techniques; it can be used in scenarios where performing an exhaustive parameter tuning is too complex and time consuming, because of the heavy number of parameter combinations. The general idea of the SPO framework can be described as follows:

(1) define a budget (e.g., number of forecasting model estimations), an initial population (i.e., a set of parameter settings) and an objective function (i.e., an out-of-sample indicator);
(2) starting from the initial population, explore the search space and infer information about the evolution of the objective function by building a metamodel;
(3) based on predictions by the metamodel, define new design points and increase the search space;
(4) refine the metamodel until budget is available.

The SPO framework tries to determine a functional relationship between a parameter setting and the objective function by building a metamodel. Many different metamodels, such as standard regression techniques, tree-based regression or Kriging models (Bartz-Beielstein et al. 2010), are supported by the SPO approach. In particular, we adopt Kriging models in our implementation since they are among the most frequently used. The metamodel is used to predict MAE for a new and wider set of parameter settings. The setting with the best expected improvement is selected (Bartz-Beielstein et al. 2010), the population is increased and it is used to refine the metamodel. The whole scheme is repeated until the budget is consumed. The main feature of the framework is that the model to be tuned is treated like a black box. Thus we can use ARIMA, ARIMAX or TF models without any distinctions. We define a budget equal to a maximum number of parameter estimations of the forecasting model. The initial population, i.e., the alternative parameter settings to test the model forecasting performance, is designed according to the Latin Hypercube Sampling (LHS), introduced by McKay et al. (1979) for computer experiments and allowing to define space-filling designs. For each parameter setting the forecasting model is estimated and the out-of-sample indicator MAE (see Section 2.3.1) is used as objective function to measure the forecasting quality. Note that, for the TF model, it is necessary to make two parameter estimations, thus the SPO approach is performed twice: parameters (v, r) are estimated according to the best goodness of fit measure F defined in (16), then parameters (p, d, q) and (P, D, Q) are estimated according to the best MAE.

2.4 *Dynamic Forecasting Model Selection*

Automating tuning operations in the design of a forecasting system is relevant when a high number of item/store combinations needs to be analyzed (i.e., a micro-forecasting approach is adopted). Besides, we investigated different models (i.e., ARIMA, ARIMAX and TF) and—in general—no method always dominate the others (Petropoulos et al. 2014). Therefore, an autonomic character of the forecasting module is of high practical interest also to choose the forecasting method to use (in an actual instance) without the necessary intervention of the decision maker (Wagner 2002). In fact, the user or the decision maker at the store level should not be exposed to the complexity of the forecasting system. In general, he/she often has neither the expertise nor enough time to directly configure these technical details, and should only pay attention to items with the highest volumes (or values) or to exceptions. To this aim, the proposed DSS is equipped with a Model Selection Module (as reported in Figure 1) devoted to analyze the set of candidate forecasting methods considering different criteria (e.g., RMSE, MAE, MaxAE, MASE, MaxAPE, R^2, and MSE, already used for model validation purposes).
Our procedures derive, for each forecasting model family, a group of non-dominated models on the basis

of in-sample indicators; then, they further perform a non-domination analysis by means of the selected out-of-sample indicators. Among the latter non-dominated models, the selection procedure concludes by choosing, for each family, the model with the best performance in terms of MAE. The reason for adopting MAE as the criterion for the final choice was agreed with the management, as retailers are mostly interested in minimizing the absolute deviation from actual sales, rather than other measures.
The selection between forecasts produced by different models is conducted adopting an individual (for each single item/store pair) approach, and the final choice among the available models can follow either a specific (yet pre-configured) or user-defined rule. This approach has the advantage of taking into account specific time series characteristics shown in each series individually in the period of interest. Moreover, the performance of individual selection generally outperforms that of aggregate selection at the expense of an additional complexity and computational cost (Armstrong 2001; Fildes and Petropoulos 2015). Nevertheless, tuning the forecasting model, which represents the most expensive component in the module, does not need to be performed at each run of the DSS; investigating the frequency of such update goes beyond the purpose of this study and deserves further research. Considering that inventory and order performances are, in general, mainly affected by the forecast errors and their variance (Fleischmann et al. 2002; Nahmias 2011), in this paper we limit our analysis considering two main possible criteria: accuracy and variability, while further investigations form a subject for future research. The considered criteria are associated to the MAE and MSE model validation measures, respectively, and the analysis of the performance of the selected models will be conducted—after a refitting using all the samples from both training and validation sets—on the test set representing, in our study, the forecasting/planning horizon.

2.5 *Order Proposal based on Sales Forecasting*

Once one of the forecasting models is selected as described in Section 2.2, in order to apply that model in the forecasting/planning horizon represented by the test set, we refit it using all samples available in both the training and the validation sets. Then sales forecasts are obtained and they can be considered as a proxy of expected demand and are provided as input to a multi-objective optimization algorithm. The aim is to identify an optimal order planning policy according to multiple (and often conflicting) objectives—reproducing management's needs—namely minimizing stock-outs and waste, as well as maximizing the quality of service perceived by customers, in terms of product freshness, while keeping residual stock levels under control. Deriving an order proposal which guarantees the best trade-off among these objectives requires a careful inventory management system. Specifically, it relies on the following assumptions:

- Items are sold according to a First-In First-Out (FIFO) rule (Nahmias 2011); i.e., they are sold on an oldest first basis.
- The age distribution of the on hand inventory has to be tracked daily, in order to identify outdates resulting in waste. As a consequence, the age of sold items needs to be tracked as well.
- When a new order is delivered to the store, it consists of fresh units whose residual shelf life is exactly equal to the nominal shelf life of the product; i.e., items arrive at the store with age equal to 1 day.
- Items leave the system due to outdating when their age passes the shelf life, meaning they have just expired and become waste.

Then, inventory is reviewed at the end of each day throughout the forecasting horizon, according to the procedure illustrated in Figure 2. The notation adopted is summarized in Table 1. Based on this notation, the following relation holds:

$$I_t = \sum_{l=1}^{SL} I_t^l \,.$$

From here, we compute the daily waste w_t as the remaining number of units with age SL on hand (if any) after satisfying the demand D_t. If D_t exceeds I_t^{SL}, no waste is observed, as resulting from line 6 in Figure 2, where the notation $(\cdot)^+$ is introduced to represent $\max(0,\cdot)$. Similarly, daily stockout s_t occurs whenever demand D_t exceeds total on hand inventory I, including the order quantity delivered in day t, q_t; otherwise,

$s_t = 0$, as resulting from line 7. Total on hand inventory at the end of day t is then obtained by difference of daily waste w_t and demand D_t from the total on hand inventory at the beginning of day t, I_t, including the order quantity delivered in day t, q_t, if no stockouts occur (so that $q_t + I_t - D_t \geq 0$), otherwise the total on hand inventory at the end of day t is equal to 0. This gives the total on hand inventory at the beginning of the next day, thus denoted as I_{t+1} in line 8.

Table 1.: Notation adopted by the order planning module.

Symbol	Description
T	forecasting horizon
t	day in the forecasting horizon
SL	shelf life
l	age of the item
I_0^l	number of units with age l in the starting inventory
I_t^l	number of units on hand with age l at the beginning of day t
I_t	total on hand inventory at the beginning of day t
D_t	demand in day t
q_t	number of units ordered (and delivered) at the beginning of day t
w_t	waste at the end of day t
s_t	stock-out at the end of day t
y_t^l	number of units with age l sold at the end of day t
$\mathcal{Y}^l$	number of units with age l sold in the forecasting horizon

Inventory Review

1: **Inputs:** T, SL, D_t, q_t;
2: $I_0^1, \ldots, I_0^{SL}$;
3: $I_1^l \leftarrow I_0^l,\ l = 1, \ldots, SL$
4: $I_1 \leftarrow \sum_{l=1}^{SL} I_0^l$
5: **for** $t = 1, \ldots, T$
6: $w_t \leftarrow \left(I_t^{SL} - D_t\right)^+$
7: $s_t \leftarrow [D_t - (q_t + I_t)]^+$
8: $I_{t+1} \leftarrow \left[(q_t + I_t - D_t)^+ - w_t\right]^+$
9: **for** $l = 1, \ldots, SL$
10: $\mathcal{Y}^l \leftarrow \mathcal{Y}^l + y_t^l$
11: $I_{t+1}^l \leftarrow I_t^{l-1} - y_t^{l-1}$
12: **end**
13: **end**

14: **Outputs:**
15: Waste $f_1 = \sum_{t=1}^{T} w_t$;
16: Freshness $f_2 = \dfrac{\sum_{l=1}^{SL} \left(\mathcal{Y}^l \sum_{t=1}^{T} y_t^l\right)}{\sum_{t=1}^{T} \sum_{l=1}^{SL} y_t^l}$;
17: Stock-outs $f_3 = \sum_{t=1}^{T} s_t$;
18: Residual stock $f_4 = I_{T+1}$;

Figure 2.: Inventory review and KPIs computation for order planning.

As shown in Figure 2, the KPIs of interest are computed as follows: *waste* f_1 (line 15) provides the overall number of items that must be discarded along the forecasting horizon due to outdating; *freshness* f_2 (line 16) is computed by tracking the age of the product when sold to the customer, and then averaging over the weekly sales; *stock-outs* f_3 (line 17) express the cumulative unmet demand at the end of the forecasting horizon; *residual stock* f_4 (line 18) corresponds to the items remaining in stock at the end of the forecasting horizon.

We formulate a multi-objective optimization problem based on the aforementioned KPIs, in order to derive a plan covering the whole forecasting horizon (typically, a week). The proposed formulation accounts for the following issues:

- Lot size constraint; i.e., orders are allowed only in multiples of a minimum order quantity.
- Fixed delivery date; i.e., orders can be delivered only in given days. The number and frequency of weekly deliveries are established by the supplier and taken as input by our DSS.
- Lead time, which determines when an order has to be placed in order to meet the delivery requirements.

The problem is solved through a meta-heuristic approach, based on a genetic algorithm implemented in Matlab (MATLAB 2014), which adopts a variant of NSGA-II (Deb 2001). As a result of the optimization process, we obtain a set of Pareto optimal solutions, which are non-dominated with respect to the four identified objectives.

The following step of the order planning module is the selection of a single order plan among all the non-dominated solutions computed by the meta-heuristic approach. To this aim, the order plan selection module is called by the DSS. This module takes two inputs: the Pareto front of optimal solutions and an indication of the criteria defined by the user. Alternative criteria may be specified in order to take management's preferences into account. Among the most widely used selection rules we cite the following two: (i) an aggregated objective function such as a weighted sum of the four KPIs, such that the non-dominated order plan minimizing the aggregated objective function is selected; (ii) ranking the KPIs by relevance in lexicographic order, thus selecting the best ranked non-dominated order plan.

It is clear that almost any kind of criterion may be introduced in the DSS. Moreover, the management's priority may change over time or it may depend on the specific kind of item and store, thus our DSS is flexible enough to satisfy different user's requirements. The order plan selection module (see Figure 1) is introduced to this objective: without altering the overall structure and behaviour of the DSS, the user may define the most suitable selection criteria fitting his/her needs.

The optimal order proposal is based on the forecasted demand provided by the forecasting module of the DSS. As demand might deviate from its forecasted value, it may be helpful for the management to evaluate the impact of demand uncertainty on the computed order proposal, in terms of KPIs variation. Therefore, we perform an ex-post analysis in which, given the optimal order plan, the DSS simulates demand variability and evaluates KPIs deviation. In particular, we simulate system performances for different realizations (say, N) of daily demand. To this aim, we estimate the distribution of daily demand based on predictions provided by each forecasting model. In particular, we assume daily demand to be normally distributed (Stevenson 2015) with mean equal to the forecasted value and standard deviation derived from the square root of the MSE associated to the forecast. Then, we sample N observations from the estimated distribution and compute the KPIs associated to the optimal order proposal when demand equals each sampled realization. Comparing the value of the KPIs for the base scenario to the N values of the same KPIs for the alternative scenarios, we are able to estimate how sensitive the order plan selected by our DSS is to demand variability. In particular, a small variability in the KPIs would be preferable to denote a stable order plan with a limited impact of demand variability on system's performances.

3. Experimental analyses

In this section we investigate the performance of the proposed DSS on a set of real sales' data for a number of sample items, chosen among the 'best sellers'. Details on the available data set and the selected sample products are reported in Section 3.1. After the pre-processing analysis, we study all the proposed forecasting models using classical statistical indicators; to this aim, we identify a training set of 90 days, a validation and a test set of 7 days each (corresponding to the usual length of the real planning horizon). Then, we run the whole DSS to derive an order proposal by optimizing the KPIs introduced in Section 2.5, based on forecasted sales provided by the model identified by the selection module described in Section 2.4. Finally, we perform a sensitivity assessment of the KPIs due to forecasts' uncertainty; this investigation helps the management to evaluate risks in terms of potential performance losses caused by sales deviating from their forecasts.

3.1 *Data Description*

The data set used for designing and setting the DSS comes from a real fresh food supply chain. The available data is made of three year sales, from 2011 to 2013, for a set of 19 small and medium sized supermarkets operating in Apulia region, Italy. We have been provided with a set of 156 fresh products identified as best sellers, for which the available sales were characterized by large and reliable values. Restricting our analyses to the freshest items—i.e., those having a shelf life lower than two months—the remaining 113 products belong to nine food categories, represented in Figure 3 together with their percentage composition. Tests have been performed on specific pairs of items and stores, according to the following criteria:

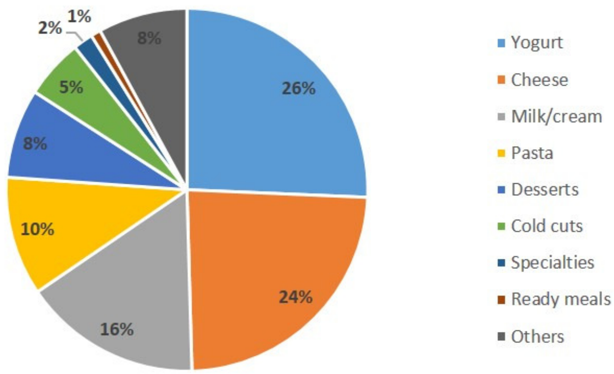


Figure 3.: Fresh food classification for our data set.

- We selected products belonging to the three biggest categories (namely, cheese, yogurt and fresh milk/cream), as they appear to be the most representative, covering 66% of the whole data set, as emerging from Figure 3.
- For each of these food categories, we selected the products showing the highest sales volume.
- For each selected product, we compared sales on different stores and computed the correlation among them; finally, we selected the store with the highest correlation.

Regarding store selection for a given product, we considered correlation as a measure of similarity among sales on different stores. Indeed, given a store A, a high correlation between the corresponding time series and the sales series from the other stores means that sales on store A precisely mimics the overall behaviour of sales on all stores.
We propose four examples to illustrate the performance of the proposed DSS. The first example refers to 1 liter of milk, selected for being a common fresh item, with a very short shelf life (4 days). In the second example we analyze another very common fresh product, 250 grams of mozzarella cheese, which has a medium shelf life (18 days) and can be supplied as single items. Another example is based on 125 grams of yogurt, having a medium-long shelf life (30 days). Finally, the last example considers 200 grams of salmon, selected for having the highest sales' volumes in our data set and a relatively longer shelf life (51 days). Key information on our sample products are summarized in Table 2. Further, Table 3 reports details on historical sales, including minimum and maximum value observed on a daily basis, as well as sample mean, standard deviation, and the corresponding coefficient of variation (CV), computed as the ratio of the stardard deviation to the mean. We also specify the percentage of zero sales occurring throughout the time horizon under study—apart from closing days, when sales are obviously zero. From here, it is clear that zero sales have a negligible impact on the following DSS data processing, ranging between 1% to 5.6% across the four sample products.

Table 2.: Data on sample products.

Product	Food Category	Shelf Life (days)	Lot size (# items)	Delivery days
Milk	Fresh milk/cream	4	12	MON,WED,THU,SAT
Mozzarella cheese	Cheese	18	1	MON,THU
Yogurt	Yogurt	30	20	MON,THU
Salmon	Specialties	51	10	TUE,SAT

Table 3.: Statistics on historical sales for sample products.

Product	Min	Max	Mean	Std. dev.	CV	% zero sales
Milk	10	47	20.89	11.43	0.55	1.0%
Mozzarella cheese	3	25	9.31	5.81	0.62	1.1%
Yogurt	13	63	25.83	16.81	0.65	5.6%
Salmon	2	26	8.36	5.87	0.70	2.2%

3.2 *Data pre-processing*

To identify seasonality in sales, we run ICA in its JADE implementation (Cardoso 1999). The pre-whitening procedure mentioned in Section 2.1 reveals that only one principal component is characterized by high variance, thereby reducing the problem dimensionality to one component. The remaining principal components are associated to the noise affecting historical sales data.
For each product, a seasonality of 7 days is observed. This is typical of food sold on large scale distribution in which customers tend to have well defined cyclical behaviour.
Concerning the impact of noise on the historical sales, we observe that it is less than 1% and, therefore, we process the original sales time series without filtering it.

3.3 *Forecasting models*

All forecasting models are implemented in Matlab and, even if different and faster implementation are possible, the analysis of the computation times among the considered models is fair. In our implementation, the user is required to specify store and item, retrieved by a database, the forecasting/planning horizon, and what kind of exogenous variables has to be accounted for (in our study, the price). When tuning the forecasting models, the following intervals have been set for their parameters:

- for grid search based tuning, we let p, d, q, v and r vary in the set $\{0,1,2\}$, while P, D, and Q may take values 0 or 1; we also set $b = 0$, as no delay is expected;
- for SPO-based tuning, we set the interval $[0,3]$ for all parameters (p, d, q, P, D, Q, v, and r), while keeping $b = 0$.

In fact, because the SPO approach explores the parameters' search space more efficiently, we used this framework to investigate whether increasing the intervals' width might provide better forecasting results while keeping computational time under control.
Table 4 provides validation results of the six forecasting models (i.e., one for each family and for tuning method) in terms of out-of-sample indicators, the bold values denoting the best value for each indicator. Notice that, whenever historical sales are zero, we do not include them in the computation of MaxAPE. In fact, dealing with best-selling products, no sales are mostly due to store closings while being seldom observed during the week, as confirmed by sparsity results from Table 3. When this is the case, it appears reasonable to exclude the corresponding prediction error from the computation of the out-of-sample indicator. Zero-demand samples still impact on all the other error measures, which are well-defined also for these values. The computational times (in seconds) required by the different methods (including tuning/selection and forecasting activities) are also reported in Table 4. The out-of-sample analysis shows that ARIMA and ARIMAX have a quite similar performance, while TF performs worse, especially in the basic version,

Table 4.: Validation (out-of-sample) analysis for milk.

	ARIMA	ARIMAX	TF	ARIMA-SPO	ARIMAX-SPO	TF-SPO
(p,d,q)	$(1,0,0)$	$(0,0,1)$	$(0,0,0)$	$(3,0,1)$	$(0,0,0)$	$(3,0,2)$
$(P,D,Q)_s$	$(1,1,1)_7$	$(1,1,1)_7$	$(1,1,1)_7$	$(2,0,0)_7$	$(0,2,0)_7$	$(0,0,0)_7$
(v,r,b)	-	-	$(1,1,0)$	-	-	$(3,0,0)$
RMSE	2.79	2.87	3.72	**2.14**	2.85	2.72
MAE	2.49	2.38	3.08	**1.78**	2.26	2.12
MaxAE	4.57	5.03	6.98	**3.70**	6.43	4.56
SMASE	0.51	0.49	0.63	**0.36**	0.46	0.44
MaxAPE	30.60%	34.78%	58.16%	**10.27%**	20.22%	22.78%
R^2	92.36%	91.90%	86.38%	**95.51%**	91.99%	92.72%
MSE	**31.32**	**31.32**	64	46.98	50.98	128.02
time(sec)	233	318	**224**	475	533	724

showing, in this case, a not relevant impact of prices. When integrated with SPO the differences among the three models are smoothed over: the integration of every model with SPO substantially increases the computational time (more than 50%, on average) but it also improves the performance indicators in almost all cases. ARIMA-SPO seems to be the best performing model according to the adopted statistical indicators, while TF is the least time-consuming.
Concerning mozzarella cheese, Table 5 shows results for the out-of-sample statistical indicators. In this

Table 5.: Validation (out-of-sample) analysis for mozzarella cheese.

	ARIMA	ARIMAX	TF	ARIMA-SPO	ARIMAX-SPO	TF-SPO
(p,d,q)	$(1,1,2)$	$(0,0,0)$	$(2,0,0)$	$(2,0,1)$	$(1,0,3)$	$(2,0,2)$
$(P,D,Q)_s$	$(1,1,1)_7$	$(1,0,0)_7$	$(1,1,0)_7$	$(0,2,3)_7$	$(0,0,0)_7$	$(2,0,0)_7$
(v,r,b)	-	-	$(2,1,0)$	-	-	$(3,1,0)$
RMSE	4.69	4.05	3.13	3.63	2.04	**1.78**
MAE	3.94	3.58	2.59	2.79	1.75	**1.50**
MaxAE	7.78	6.45	5.65	7.97	3.81	**3.43**
SMASE	1.24	1.12	0.81	0.88	0.55	**0.47**
MaxAPE	66.79%	65.26%	54.90%	**46.81%**	63.54%	48.93%
R^2	8.82%	32.17%	59.52%	45.25%	82.72%	**86.92%**
MSE	**9.12**	13.48	18.24	13.19	17.9	19.92
time(sec)	245	307	**231**	506	493	715

case, TF is the best performing among the basic models in terms of computational time and all the statistical indicators except for MSE. Considering the integration with SPO, TF-SPO is the best one except for MaxAPE, MSE and computational time. Hence, in this case, prices have a relevant role as exogenous data. Again we notice that the SPO approach allows to improve the out-of-sample indicators at overall computational time's expense. However, the tuning phase is not always required at each use of the system allowing significant savings. Results for yogurt are summarized in Table 6. It is evident that there is no model clearly outperforming the others over all the indicators. We note that negative values of R^2 observed for almost all models do not reflect a low forecasting accuracy, as supported by the other indicators; rather, it depends on the limited variability of historical sales around their average values: based on the definition of the R^2 coefficient, this makes the denominator very small thus implying a negative R^2. The validation analysis for salmon is reported in Table 7. ARIMA-SPO seems to be the best model in terms of RMSE, MaxAE, MaxAPE and R^2 while ARIMAX-SPO performs slightly better as far as MAE and SMASE are concerned. ARIMA is the least time consuming model; note that the forecasting quality of the three models with SPO is almost comparable, but the average MSE is higher than the models using GS, that is, the provided forecast is the least accurate one. Conversely, ARIMAX is the model with the smallest average MSE. Overall, TF and ARIMAX seem to outperform ARIMA in terms of all the out-of-sample indicators.

Table 6.: Validation (out-of-sample) analysis for yogurt.

	ARIMA	ARIMAX	TF	ARIMA-SPO	ARIMAX-SPO	TF-SPO
(p,d,q)	$(0,1,1)$	$(1,1,2)$	$(1,1,1)$	$(1,0,2)$	$(2,1,1)$	$(2,0,2)$
$(P,D,Q)_s$	$(1,0,1)_7$	$(1,1,0)_7$	$(1,1,0)_7$	$(2,1,2)_7$	$(2,1,3)_7$	$(1,0,1)_7$
(v,r,b)	-	-	$(2,1,0)$	-	-	$(1,3,0)$
RMSE	2.65	**2.35**	2.95	2.70	2.52	2.96
MAE	2.20	**1.35**	2.11	2.12	2.05	1.99
MaxAE	5.38	5.92	6.29	**4.72**	4.87	7.08
SMASE	0.55	**0.34**	0.53	0.53	0.51	0.50
MaxAPE	89.69%	98.71%	104.81%	77.77%	97.30%	**33.52%**
R^2	-17.46%	**7.71%**	-46.37%	-22.29%	-6.49%	-47.43%
MSE	**15.2**	17.9	28.65	24.48	25.18	23.42
time(sec)	**225**	331	232	383	440	613

Table 7.: Validation (out-of-sample) analysis for salmon.

	ARIMA	ARIMAX	TF	ARIMA-SPO	ARIMAX-SPO	TF-SPO
(p,d,q)	$(1,0,0)$	$(1,0,0)$	$(2,0,2)$	$(1,1,1)$	$(0,1,3)$	$(1,1,1)$
$(P,D,Q)_s$	$(0,1,1)_7$	$(1,0,1)_7$	$(1,0,1)_7$	$(3,1,2)_7$	$(1,2,2)_7$	$(3,1,2)_7$
(v,r,b)	-	-	$(2,1,0)$	-	-	$(2,1,0)$
RMSE	6.94	5.34	5.21	**4.12**	4.22	4.28
MAE	4.96	4.21	4.08	3.59	**2.98**	3.33
MaxAE	13.74	11.51	10.15	**6.90**	9.85	8.30
SMASE	0.58	0.49	0.48	0.42	**0.35**	0.39
MaxAPE	35.23%	20.56%	26.70%	**18.68%**	25.93%	21.54%
R^2	80.24%	88.31%	88.87%	**93.05%**	92.71%	92.48%
MSE	93.3	**70.6**	115.82	129.94	474.71	251.48
time(sec)	**219**	299	241	465	512	722

Summing up, if we consider all the out-of-sample indicators and compare all the selected models, ARIMA-SPO is the non-dominated model for milk, ARIMA-SPO and TF-SPO are the two non-dominated models for mozzarella cheese, ARIMAX, ARIMA-SPO and TF-SPO are the non-dominated models for yogurt, while ARIMA-SPO and ARIMAX-SPO are the two non-dominated models for salmon. Taking the computational time into account leads to different conclusions, as the integration of the forecasting models with SPO almost doubles the total solve elapsed time. ARIMA seems to be preferable for milk, while ARIMAX performs better for yogurt and TF seems to provide the best compromise between forecasting quality and computational burden for mozzarella cheese and salmon. This behaviour may be affected by the impact of prices: for primary goods and highly fresh products, such as milk, having a shelf life of a few days, price seems to be irrelevant for the forecasting process, also because their high perishability prevents from buying too many pieces when the price is lower. On the contrary, for secondary goods, such as mozzarella cheese, salmon and yogurt, price plays a more relevant role.

The family of Exponential Smoothing (ES) models represents a simple robust approach commonly used in practice and also previously adopted in the application context under study (e.g., see Fildes and Petropoulos (2015); Gardner and McKenzie (2011); McKenzie and Gardner (2010)). Hence, we consider three different ES models as benchmarks, namely the simple ES (indicated as ETS), the seasonal ES (indicated as ETSs) and the linear damped trend ES with seasonality (ETSDs). Table 8 reports a comparison of the selected models with some Exponential Smoothing (ES) models in terms of both validation and test performance. The results are summarized reporting the averages on the considered cases for the SMASE indicator obtained in the validation phase and the MAE measured in the test phase. Notice that several models belonging to the ES family can be expressed as an ARIMA and are among those investigated by the sales forecasting module of the proposed DSS; therefore, they would be selected in case they perform better than other forecasting models. Comparing the forecasting performance of the ES models against the models tested

Table 8.: Performance comparison of ARIMA (A), ARIMAX (AX), and TF models with Exponential Smoothing methods.

	A	AX	TF	A-SPO	AX-SPO	TF-SPO	ETS	ETSs	ETSDs
SMASE (Validation)	0.72	0.61	0.61	0.55	0.47	**0.45**	1.31	0.74	0.89
MAE (Test)	0.49	0.41	0.42	0.37	**0.32**	**0.32**	0.96	0.49	0.59

by the DSS in terms of the out-of-sample indicators, it emerges that there is always at least one model per group—either the one using a grid search or SPO—among those returned by the DSS which performs better than ESs. These observations are confirmed by the results (in terms of MAE) obtained over the forecasting horizon. When forecasts are poor, it is possible that the forecasting model is structurally inadequate, or that the parameters of the model are not set properly. In general it is difficult for the user to distinguish between the two cases. At this aim, the proposed DSS offers a simple-to-use tool able to give high quality forecasts automatically adapting the structure of the forecasting model to the underlying data, if necessary.

3.4 *Order Proposal*

As emerging from the results discussed in the previous subsection, there is no forecasting model clearly outperforming the others over all the analyzed products. Moreover, the forecast performance of a selected model tends to deteriorate over time without adequate diagnosis and tuning actions. Therefore, it is useful for the user to have a DSS that dynamically adapts to the available data, each time selecting the most appropriate forecasting model according to some user-specified criteria. In fact, this avoids to choose a forecasting technique a priori, regardless of the specific characteristics of the underlying data. Specifically, the proposed DSS implements two alternative final model selection criteria among those individuated for each family: the *accuracy criterion* selects the model having the smallest MAE, corresponding to the smallest deviation in terms of units of product; the *variability criterion* chooses the model with the smallest MSE, which implies a more stable forecast. Combining these two criteria with the alternative tuning techniques described in Section 2.3, we analyze four configurations. Summing up, we analyze the following four DSS configurations:

- Config #1 $\rightarrow$ grid search based tuning, accuracy criterion;
- Config #2 $\rightarrow$ grid search based tuning, variability criterion;
- Config #3 $\rightarrow$ SPO tuning, accuracy criterion;
- Config #4 $\rightarrow$ SPO tuning, variability criterion.

For each of these four configurations, once sales forecasts are available from the forecasting module, our DSS calls the order planning module aiming to identify an optimal order proposal with respect to the KPIs (f_i with $i = 1, \ldots, 4$) discussed in Section 2.5, assuming empty inventory at the beginning of the planning horizon. So, we derive the optimal order plan associated to the daily forecasted sales, covering the whole planning horizon; we refer to this scenario as the *base* scenario. The computational time to solve the optimization problem usually remains below 30 seconds, thus impacting by less than 9% on the overall computational cost required by each DSS configuration considered when tuning and selection of the forecasting model are also performed.

After the multi-objective optimization algorithm computes the Pareto frontier, it is necessary to define a criterion for the order plan selection module of the DSS. In our experimental analyses, we test the weighted sum of KPI as selection criterion: specifically, the order proposal is derived by equally weighting the four KPIs and selecting the one minimizing their weighted sum. Note that different combinations of weights can be used to reflect the corresponding management's policies.

Next, we analyse the variability of the four selected KPIs resulting from demand uncertainty. To this aim, we sample $N = 100$ different realizations of daily demand over the planning horizon, using the estimated probability distribution. This results in as many scenarios, including the base scenario. Then, we simulate system's performance in terms of its identified KPIs when different demands occur, as expressed by the

N alternative scenarios, while the order proposal remains fixed at the optimal level associated to the base scenario. We note that this analysis is usually very fast (less than 1 second, on average), so its contribution to the overall computational cost is negligible.
For each product we compare the four DSS configurations over the following results:

- daily sales forecasting along the planning horizon, denoted as *forecasted sales*;
- daily *sales variability* as square root of the forecasting MSE;
- sales coefficient of variation (CV), based on the daily average forecasted sales and variability;
- *order proposal* for the base scenario;
- expected *inventory* at the end of each day associated to the optimal order proposal in the base scenario;
- *KPIs* (f_1-f_4) associated to the optimal order proposal in the base scenario;
- sample mean (*avg_inventory*) and standard deviation (*stddev_inventory*) of daily inventory resulting from alternative scenarios generated during the sensitivity analysis;
- sample mean (*avg_KPI*), standard deviation (*stddev_KPI*), and coefficient of variation (*CV_KPIs*) of the KPIs associated to the alternative scenarios generated during the sensitivity analysis.

Results are summarized in Tables 9-11, specifying also the forecasting model selected by the model selection module in each configuration. We also include box plots showing variability of the KPIs due to demand uncertainty along the forecasting horizon. Possible outliers have been identified as follows: data points are considered outliers if they fall outside the interval $[q_1 - w(q_3 - q_1), q_3 + w(q_3 - q_1)]$, where q_1 and q_3 are the 25^{th} and 75^{th} percentiles, respectively, and we set $w = 1.5$ to guarantee 99.3% coverage, approximately corresponding to $\pm 2.7\sigma$. Table 9 reports results for milk, which can be delivered every other day (see Table 2 and zero values in the order proposal section of Table 9). We notice that overall weekly forecasted sales are different across configurations, ranging from 135 to 146, while order proposals always sum up to 144, although with different assortments along the planning horizon. This may result in stock-outs and/or residual stock at the end of the week (as in configs #1-2). Nevertheless, based on the choice of the weights for the KPIs and due to lot size constraints, it turns out to be preferable to order less, thus accepting to face stock-outs, instead of ordering one lot more to prevent stock-outs, rather risking to have waste. Figure 4 depicts the box plots associated to the KPIs for milk. Blue dots superimposed to the box plots correspond to the KPI values for the base scenario, when demand equals forecasted sales. We also included squared markers for the KPIs measured at the end of the planning horizon, resulting from the selected order proposal having faced the historical demand, as it occurred in the planning horizon. Triangular markers reproduce the KPIs we would obtain by optimizing the order proposal in the ideal case of forecasting sales perfectly matching the historical demand observed at the end of the planning horizon.
We notice that zero waste is reached by all DSS configurations across the N scenarios, so we do not include the corresponding boxplot. Freshness remains around 1.5 days for the base scenario: this is a satisfactory result, since milk has a shelf life of 4 days, and in the base scenario it will be always sold within the next half a day following the delivery day. Config #1 ensures the lowest freshness, both in terms of median and in terms of variability across the N scenarios. The other configurations show similar behaviour with higher median values — still remaining below 2 days — and increased variability, particularly for configs #3-4. All configurations show similar stock-outs performance, being config #1 slightly better in terms of lower third quartile and maximum values. All the other configurations show similar median and third quartile values. The lowest residual stock is provided by configs #3 and 4, in terms of both quartile and maximum values. Higher interquartile ranges are observed for config #2; it has high third quartile and maximum values.

Mozzarella cheese is delivered (at most) twice a week. Based on the delivery days specified in Table 2, we notice from Table 10 that the DSS usually suggests to order as many units to fully cover the overall forecasted sales up to the next delivery: in fact, mozzarella cheese is sold as single item, so the best KPIs are obtained when ordering as many units as expected to sale. In the base scenario this ensures zero stock-outs and residual stock at the end of the planning horizon. However, the identified order proposal becomes subject to higher variability when tested in the alternative scenarios: in fact, both average stock-out and residual stock values and their standard deviations increase. This is clearly shown in Figure 5, representing the box plots of the KPIs. Again all DSS configurations ensure zero waste across the N scenarios, and are

Table 9.: DSS configuration results for milk.

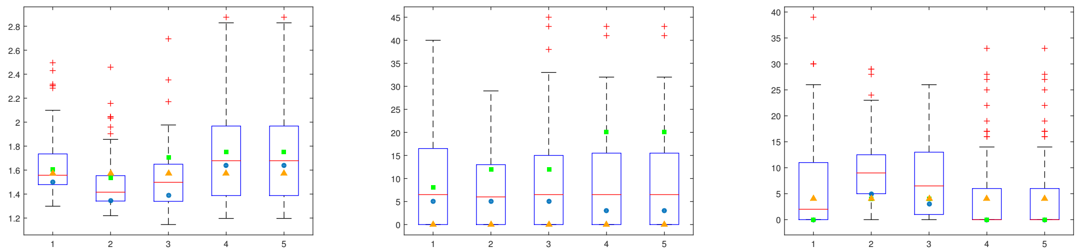
# config	**1**	**2**	**3**	**4**
Forecasting model	ARIMAX	ARIMA	ARIMA-SPO	ARIMA-SPO
Forecasted sales	16	16	12	12
	20	22	20	20
	41	41	40	40
	0	0	0	0
	22	22	21	21
	24	24	22	22
	21	21	20	20
Sales variability	5.37	5.32	6.07	6.07
	5.55	5.59	6.17	6.17
	5.55	5.62	6.17	6.17
	5.55	5.62	6.19	6.19
	5.55	5.62	6.19	6.19
	5.55	5.62	6.19	6.19
	5.55	5.62	6.19	6.19
Sales CV	0.23	0.23	0.27	0.27
Order proposal	36	36	36	36
	0	0	0	0
	36	48	48	48
	0	0	0	0
	48	36	36	36
	0	0	0	0
	24	24	12	12
Inventory	20	20	24	24
	0	0	4	4
	0	7	12	12
	0	7	12	12
	26	21	27	27
	2	0	5	5
	5	3	0	0
KPIs	0	0	0	0
	1.35	1.39	1.64	1.64
	5	5	3	3
	5	3	0	0
Avg_inventory	20.11	20.65	22.88	22.88
	3.87	2.77	4.65	4.65
	2.61	10.05	13.46	13.46
	2.61	10.05	13.46	13.46
	28.86	23.69	28.55	28.55
	5.8	3.69	8.01	8.01
	9.47	7.78	4.54	4.54
Stddev_inventory	4.98	5.69	5.92	5.92
	5.47	4.29	5.69	5.69
	4.73	7.23	9.23	9.23
	4.73	7.23	9.23	9.23
	6.91	8.66	11.02	11.02
	5.90	5.42	8.62	8.62
	7.01	6.77	7.25	7.25
Avg_KPIs	0	0	0	0
	1.48	1.53	1.71	1.71
	7.55	9.28	9.26	9.26
	9.47	7.78	4.54	4.54
Stddev_KPIs	0	0	0	0
	0.21	0.26	0.36	0.36
	7.62	10.21	10.62	10.62
	7.01	6.77	7.25	7.25
CV_KPIs	–	–	–	–
	0.14	0.17	0.21	0.21
	1.01	1.10	1.15	1.15
	0.74	0.87	1.60	1.60

Table 10.: DSS configuration results for mozzarella cheese.

# config	**1**	**2**	**3**	**4**
Forecasting model	TF	ARIMA	TF-SPO	ARIMA-SPO
Forecasted sales	0	0	0	0
	6	10	6	5
	7	9	9	9
	10	9	12	11
	0	8	0	8
	11	12	10	10
	18	18	14	15
Sales variability	4.24	3.01	4.41	3.54
	4.24	3.01	4.41	3.62
	4.29	3.01	4.48	3.65
	4.29	3.01	4.48	3.66
	4.29	3.01	4.48	3.66
	4.29	3.01	4.48	3.66
	4.29	3.01	4.48	3.66
Sales CV	0.49	0.27	0.53	0.38
Order proposal	0	0	0	0
	23	29	27	26
	0	0	0	0
	0	0	0	0
	29	37	24	32
	0	0	0	0
	0	0	0	0
Inventory	0	0	0	0
	17	19	21	21
	10	10	12	12
	0	1	0	1
	29	30	24	25
	18	18	14	15
	0	0	0	0
KPIs	0	0	0	0
	2.42	2.18	2.39	2.28
	0	0	0	0
	0	0	0	0
Avg_inventory	0	0	0	0
	16.19	19.14	20.29	20.68
	8.27	10.21	10.98	11.48
	1.83	3.04	2.59	2.36
	30.83	32.29	26.59	25.84
	19.30	20.20	15.93	15.48
	3.47	3.84	4.34	3.24
Stddev_inventory	0	0	0	0
	3.53	3.10	4.34	3.18
	4.96	4.20	5.86	4.62
	3.02	3.78	4.03	3.68
	3.02	5.29	4.03	5.21
	5.05	6.55	5.74	6.05
	4.40	5.05	5.40	4.91
Avg_KPIs	0	0	0	0
	2.44	2.26	2.48	2.28
	5.99	3.49	5.49	4.77
	3.47	3.84	4.34	3.24
Stddev_KPIs	0	0	0	0
	0.29	0.29	0.45	0.30
	6.59	5.26	6.08	5.44
	4.40	5.05	5.40	4.91
CV_KPIs	–	–	–	–
	0.12	0.13	0.18	0.13
	1.10	1.51	1.11	1.14
	1.27	1.31	1.24	1.51

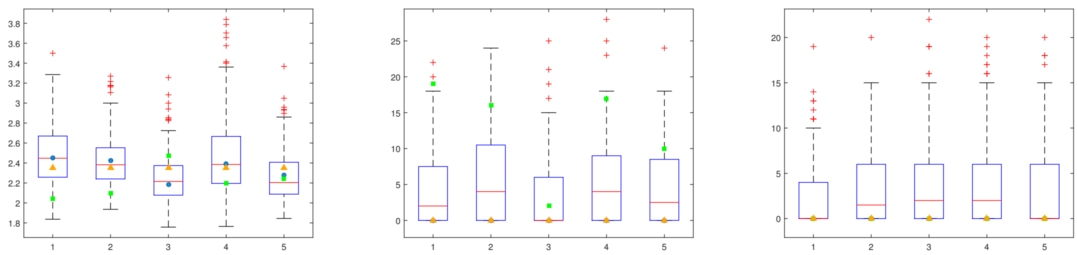
not reported in the figure. Freshness varies between 1.8 and 3.4 days, remaining below 2.5 days on average, given a shelf life of 18 days for mozzarella cheese. Stock-outs show low median and limited variability in config #2; values increase in config #4 and further worsen in configs #3 and 1, which has the highest third quartile and maximum values. As for residual stock, the lowest quartile values and variability across the N scenarios are provided by config #4, while the other configurations have higher median and third quartile values.

Assuming to have no inventory at the beginning of the planning horizon and having fixed delivery days, all yogurt forecasted sales for the first day result in stock-outs, as shown in Table 11. Figure 6 illustrates the box plots related to the KPIs for yogurt, which can be delivered twice a week. All DSS configurations guarantee zero waste (not reported) and high freshness, which always remains below 4.5 days, i.e., around



(a) Freshness (b) Stock-out (c) Residual stock

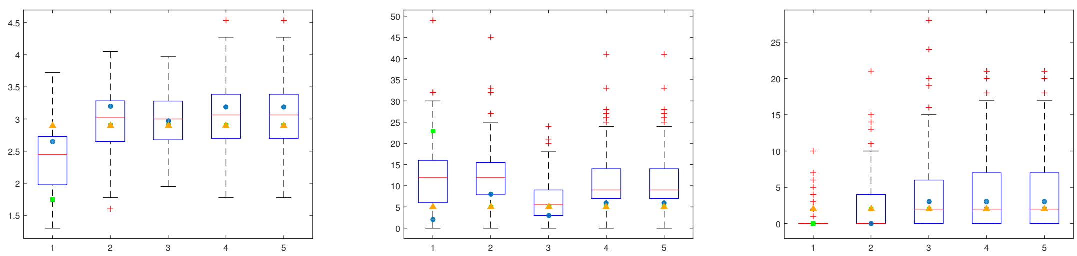
Figure 4.: Box plots of the KPIs for milk.



(a) Freshness (b) Stock-out (c) Residual stock

Figure 5.: Box plots of the KPIs for mozzarella cheese.

15% of its shelf life; the shortest interquartile range is provided by config #2, while configs #1, 3 and 4 show similar performance, with higher variability. Stock-outs show the smallest variability in config #2, which also has the lowest quartile values. The best residual stock is provided by config #1. While Configs #2, 3 and 4 have higher values for the first quartile and median values, and similar maximum values. Figure 7



(a) Freshness (b) Stock-out (c) Residual stock

Figure 6.: Box plots of the KPIs for yogurt.

shows box plots of the KPIs for salmon, whose delivery days are the first and fourth day of the week. Again zero waste (not reported) and (almost) no stock-outs are guaranteed by all DSS configurations in the base scenario. Config #4 provides the lowest stock-out values across scenarios, while the worst is config #3, having the highest median and third quartile values and the widest interquartile range. Given a shelf life of 51 days, average freshness is also good, showing limited variability around 2.2 days, while remaining always under 3.5 days; config #2 has the best quartile values and the lowest variability across scenarios. Residual stock is zero or close to zero in the base scenario; similar performance is observed across scenarios by configs #1-2: the latter has the lowest third quartile and maximum values, which slightly increases in configs #1, and increases even more in configs #3 and 4.

The proposed tool combining simulation-based analyses and box plots appears to be adequate to evaluate the KPIs' variability induced not only by the forecasts themselves and the planning model but also by the criteria adopted by the decision maker to select an order plan from the set of Pareto-optimal solutions. Such

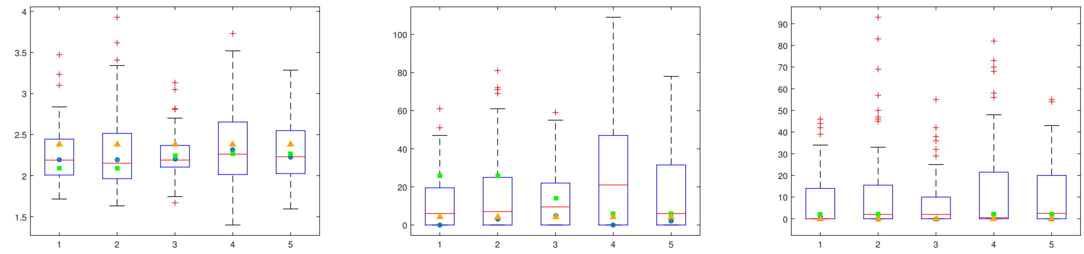
Table 11.: DSS configuration results for yogurt.

# config	1	2	3	4
Forecasting model	ARIMAX	ARIMA	TF-SPO	TF-SPO
Forecasted sales	7	3	6	6
	9	9	10	10
	6	7	6	6
	12	11	8	8
	0	0	0	0
	7	6	6	6
	7	4	7	7
Sales variability	4.27	3.91	4.72	4.72
	4.27	3.91	4.72	4.72
	4.27	3.91	4.87	4.87
	4.27	3.91	4.87	4.87
	4.27	3.91	4.93	4.93
	4.27	3.91	4.93	4.93
	4.27	3.91	4.93	4.93
Sales CV	0.53	0.59	0.68	0.68
Order proposal	0	0	0	0
	40	40	40	40
	0	0	0	0
	0	0	0	0
	0	0	0	0
	0	0	0	0
	0	0	0	0
Inventory	0	0	0	0
	31	31	30	30
	25	24	24	24
	13	13	16	16
	13	13	16	16
	6	7	10	10
	0	3	3	3
KPIs	0	0	0	0
	3.20	2.97	3.19	3.19
	8	3	6	6
	0	3	3	3
Avg_inventory	0	0	0	0
	30.17	31.36	30.03	30.03
	23.49	24.62	23.30	23.30
	12.6	13.69	15.29	15.29
	12.60	13.69	15.29	15.29
	6.42	7.50	8.99	8.99
	2.58	4.07	4.20	4.20
Stddev_inventory	0	0	0	0
	4.45	3.70	4.47	4.47
	5.71	4.95	6.03	6.03
	6.54	6.47	7.51	7.51
	6.54	6.47	7.51	7.51
	6.17	6.19	7.09	7.09
	4.08	5.44	5.58	5.58
Avg_KPIs	0	0	0	0
	2.95	2.98	3.03	3.03
	12.82	6.27	11.50	11.50
	2.58	4.07	4.20	4.20
Stddev_KPIs	0	0	0	0
	0.53	0.42	0.57	0.57
	7.53	4.86	7.39	7.39
	4.08	5.44	5.58	5.58
CV_KPIs	–	–	–	–
	0.18	0.14	0.19	0.19
	0.59	0.77	0.64	0.64
	1.58	1.34	1.33	1.33

Table 12.: DSS configuration results for smoked salmon.

# config	1	2	3	4
Forecasting model	TF	ARIMAX	ARIMAX-SPO	ARIMA-SPO
Forecasted sales	25	35	31	36
	28	34	28	31
	26	29	28	23
	32	35	36	37
	53	44	58	54
	0	0	0	0
	29	28	27	31
Sales variability	10.79	8.08	11.26	10.29
	10.79	8.35	15.92	11.37
	10.80	8.36	19.50	11.59
	10.80	8.36	22.01	11.64
	10.81	8.36	24.26	11.65
	10.81	8.36	26.31	11.65
	10.81	8.36	28.22	11.65
Sales CV	0.34	0.24	0.58	0.32
Order proposal	110	130	130	130
	0	0	0	0
	0	0	0	0
	0	0	0	0
	80	70	80	80
	0	0	0	0
	0	0	0	0
Inventory	85	95	99	94
	57	61	71	63
	31	32	43	40
	0	0	7	3
	27	26	29	29
	27	26	29	29
	0	0	2	0
KPIs	0	0	0	0
	2.19	2.20	2.32	2.23
	3	5	0	2
	0	0	2	0
Avg_inventory	84.88	94.49	99.42	93.53
	55.80	59.74	69.58	60.89
	30.48	30.32	41.55	37.85
	7.49	4.87	11.93	9.79
	34.03	31.42	33.00	36.31
	34.03	31.42	33.00	36.31
	10.77	7.18	13.07	10.64
Stddev_inventory	9.71	7.68	12.10	9.72
	16.41	11.89	18.50	15.40
	19.45	13.80	24.51	18.24
	12.41	8.01	17.50	12.63
	18.77	12.83	24.41	17.78
	18.77	13	24.41	17.78
	17.95	10.79	19.48	14.29
Avg_KPIs	0	0	0	0
	2.29	2.24	2.35	2.29
	15.72	13.91	27.26	15.99
	10.77	7.18	13.07	10.64
Stddev_KPIs	0	0	0	0
	0.45	0.25	0.47	0.36
	19.37	14.77	29.58	19.48
	17.95	10.79	19.48	14.29
CV_KPIs	–	–	–	–
	0.20	0.11	0.20	0.16
	1.23	1.06	1.09	1.22
	1.67	1.50	1.49	1.34

a tool provided support to compare the different DSS configurations investigated in this study. Overall, we notice that all configurations have a significantly smaller dispersion than the historical one. Configurations adopting SPO offer in general good performances not only in terms of quality of sales forecasting but also in terms of KPIs' dispersion. Configurations adopting the *variability* criterion (i.e., configs #2 and 4) to select the forecasting model give, on average, lower dispersion of the predictions; however, this result tends to be absorbed throughout the planning process due to the combined effects of solution integrality, delivery days restrictions and lot size constraints. On the basis of these observations, it seems useful—from



(a) Freshness (b) Stock-out (c) Residual stock

Figure 7.: Box plots of the KPIs for salmon.

an overall DSS performance and reliability point of view—to prefer configurations oriented towards more accurate predictions, as those based on the *accuracy* selection criterion and SPO tuning.

4. Conclusions

In this paper we propose a modular and reliable DSS for sales forecasting and order planning in the supply chain management of packaged fresh food products. The proposed DSS combines, in a unique, flexible and easy-to-use software tool, a forecasting module to derive sales forecasts from historical data and exogenous variables, supported by a model selection and tuning module for the automatic choice and configuration of the forecasting method, and a multi-objective optimization module equipped with an order plan selection module, to derive the best order proposal based on a set of KPIs accounting for cost and quality of service. Three different forecasting model families were considered and tested on a set of sample products in a real supply network.

Our results clearly show the benefits of deriving an optimal order proposal based on sales forecasting, explicitly accounting for demand variability and the possible impact of exogenous variables. The proposed analyses highlight the capability of the DSS to absorb relevant differences in terms of forecasting behaviour, thus limiting their impact on the order planning phase. Another advantage offered by this DSS relies on its flexibility, as it is designed to be easy-to-use and to automatically run alternative approaches in terms of forecasting and model tuning techniques, depending on the characteristics of the data set. In this respect, we notice that results point out that there is no dominant forecasting model and there is not convenience to use a single model for all the cases (i.e., pairs item/store), and also the performance of a model selected for a specific case might deteriorates over time. Hence, instead of a one-size-fits-all approach, an individual selection including the identification of the best method for each series is considered, though it is more computationally intensive. More specifically, configurations using SPO tend to provide more accurate forecasts, although the computational time, when tuning is required, is higher than the configurations adopting the grid search.

For a given configuration of the DSS, the forecasting model selected may change depending on the product or the store under investigation, as well as on the specific period (i.e., sales time series) under study. In fact, it is not required to identify the 'best' forecasting technique for a given product. Results show that almost all forecasting techniques are selected at least once across the DSS configurations, which confirms there is no technique always 'worse' than the others. Overall, this supports the usefulness of a DSS in identifying the most appropriate forecasting model for the specific situation, accounting for both product's characteristics, its historical sales and user's preferences. Specifically, it is preferable to adopt a DSS configuration pursuing the *accuracy* criterion, as it provides more accurate sales forecasting, while its dispersion—whenever relevant—tends to be absorbed throughout the planning process. In fact, as the sales forecasting module can be computationally more expensive than the order planning module, particularly because of the micro-forecasting environment we are working on, investing more computational resources on the forecasting module might pay off in terms of reliability of the whole DSS.

Further research may focus on the forecasting side extending the considered families of forecasting meth-

ods (e.g., including Exponential Smoothing and Neural Networks), and developing more sophisticated tools for parameter tuning of models, and on performing long-run simulations (including the implementation and test of both monitoring and re-tuning policies), e.g. over one year, to test alternative settings for the tuning module (McCarthy et al. 2006). Moreover, further research efforts could be oriented to evaluate the overall behaviour of the proposed DSS against the current supply chain; and design suitable robust planning approaches to cope with the uncertainty associated to the obtained forecasts (Fleischmann et al. 2002; Simangunsong et al. 2012). Another interesting theme for future researches could be devoted to adapt and extend the proposed approach to cases including products characterized by a random shelf life (Kouki et al. 2014) as often occurs for unpackaged fresh foods.
Many elements, such as market and weather conditions, competitor's plans, last minute changes at retail or supply side, promotions or festivities, as well as intermittent demand behaviour, may have an impact on sales and considering their contribution in the forecasting process may represent promising future research directions. The latter could be addressed considering managers judgment to adjust the statistical baseline forecasts (Goodwin and Fildes 1999; Önkal et al. 2013; Syntetos et al. 2009).
The modular and flexible structure of the proposed DSS enables for an easy implementation of different possible extensions for more general contexts or for more integrated approaches. At this aim, it is worthwhile to enrich the set of KPIs to consider customer service and supply chain costs, risks, uncertainties, and sustainability issues.

Acknowledgements

Authors would like to thank the editor and the anonymous reviewers for their insightful and constructive comments. This work was supported by the E-CEDI project, funded by Apulia Region under the POR FESR 2007-2013 grant, Asse I, Linea 1.2 - Azione 1.2.4. The work of the second author was partly supported by PRIN 2012 n. 2012MTE38N.

References

Abbou, R., Loiseau, J. J., Moussaoui, C., (2017), Robust inventory control of production systems subject to uncertainties on demand and lead times, *International Journal of Production Research*, 55:8, 2177–2196,

Ahumada, O., J.R. Villalobos, (2009) Application of planning models in the agri-food supply chain: A review. *European Journal of Operational Research*, vol. 195, 1–20.

Ai, X-Y., Zhang, J-L., Wang, L., (2017), Optimal joint replenishment policy for multiple non-instantaneous deteriorating items, *International Journal of Production Research*, 55:16, 4625–4642,

Aiello, G., M. Enea, C. Muriana (2014). Economic benefits from food recovery at the retail stage: An application to Italian food chains. *Waste Management*, Vol. 34, 1306–1316.

Alvarez, J., and Johnson, R. (2011), Doug Rauch: Solving the American Food Paradox. *Harvard Business Review*, 9, 12–22, November.

Arminger, G., (2004), Sales and order forecasts in CPFR. *ECR Journal* 4 (1), 55-67

Armstrong, J. (2001), *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Springer.

Babai, M., Ali, M., Boylan, J., and Syntetos, A. (2013), Forecasting and inventory performance in a two-stage supply chain with ARIMA(0,1,1) demand: Theory and empirical analysis. *International Journal of Production Economics*, 143(2):463–471.

Bartz-Beielstein, T., Chiarandini, M., Paquete, L., and Preuss, M. (2010), *Experimental Methods for the Analysis of Optimization Algorithms*. Springer.

Beske, P., A. Land, S. Seuring, (2014), Sustainable supply chain management practices and dynamic capabilities in the food industry: A critical analysis of the literature. *International Journal of Production Economics*, vol. 152, 131–143.

Borade, A. and Sweeney, E. (2015), Decision support system for vendor managed inventory supply chain: a case study. *International Journal of Production Research*, 53(16):4789–4818.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2008), *Time Series Analysis: Forecasting and Control*. Wiley, 4th edition.

Brandenburg, M. K. Govindan, J. Sarkis, and S. Seuring, (2014), *Quantitative models for sustainable supply chain management: Developments and directions*. European Journal of Operational Research, Vol. 233, 299–312.

Bourlakis, M. A. and Weightman, P. W. H. (2004), *Food supply chain management*. Oxford, UK, Blackwell.

Burnham, K. P. and Anderson, D. R. (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer.

Cardoso, J. (1999), High-order contrasts for independent component analysis. *Neural Computation*, 11(1), 157–192.

Chatfield, C. (2001), Prediction intervals for time-series forecasting. In Armstrong, J. (ed.), *Principles of Forecasting, International Series in Operations Research & Management Science*, vol. 30, 475–494. Springer US.

Darlington, R., S. Rahimifard (2006), A responsive demand management framework for the minimization of waste in convenience food manufacture, *International Journal of Computer Integrated Manufacturing*, Vol. 19, 751–761.

Darlington, R., T. Staikos, S. Rahimifard, (2009), Analytical methods for waste minimisation in the convenience food industry. *Waste Management*, Vol. 29, 1274–1281.

Deb, K. (2001), *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley & Sons.

Dellino, G., J.P.C. Kleijnen, C. Meloni (2010), Robust optimization in simulation: Taguchi and response surface methodology, *International Journal of Production Economics*, 125, 1, 52–59.

Dellino, G., J.P.C. Kleijnen, C. Meloni (2012), Robust optimization in simulation: Taguchi and Krige combined. *INFORMS Journal on Computing*, 23, 4, 471–484.

Diabat, A., Govindan, K., Panicker, V. V., (2012), Supply chain risk management and its mitigation in a food industry, *International Journal of Production Research*, 50:11, 3039–3050

Dotoli, M., Fanti, M., Meloni, C., Zhou, M., (2005), A multi-level approach for network design of integrated supply chains. *International Journal of Production Research*, 43 (20), 4267-4287.

Du, X. F., Leung, S. C. H., Zhang, J. L., and Lai, K. K. (2009), Procurement of agricultural products using the CPFR approach. *Supply Chain Management: An International Journal*, 14 (4), 253–258.

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., and Petropoulos, F., (2015), Simple versus complex selection rules for forecasting many time series. *Journal of Business Research*, 88, 1692–1701.

Fleischmann, B., H. Meyr, M. Wagner (2002), Advanced planning. In: Stadtler, H., Kilger, C. (Eds.), Supply Chain Management and Advanced Planning, 2nd Edition. Springer, Berlin, 71–96.

Franses, P.H. (2016), A note on the Mean Absolute Scaled Error. *International Journal of Forecasting*, 32, 1, 20–22.

Galasso, F., C. Mercé, B. Grabot (2009), Decision support framework for supply chain planning with flexible demand. *International Journal of Production Research*, 47, 2, 455–478.

Gardner, E.S.Jr., McKenzie, E. (2011), Why the damped trend works. *Journal of the Operational Research Society*, 62(6), 1177-1180.

Goodwin, P., and Fildes, R., (1999), Judgmental Forecasts of Time Series Affected by Special Events: Does Providing a Statistical Forecast Improve Accuracy?, *Journal of Behavioral Decision Making*, 12(1), 37–53.

Goodwin, P., (2002), Integrating management judgment and statistical methods to improve short-term forecasts, *Omega*, 30(2), 127–135.

Höglund, R. and Östermark, R. (1991), Automatic ARIMA modelling by the cartesian search algorithm. *Journal of Forecasting*, 10(5):465–476.

Huang, M.-G. (2013), Economic ordering model for deteriorating items with random demand and deterioration. *International Journal of Production Research*, 51(18):5612–5624.

Hyndman, R.J. (2006), Another look at forecast-accuracy metrics for intermittent demand. *Journal of Statistical Software*, 27(3):1–22.

Hyndman, R.J. and Khandakar, Y. (2008), Automatic time series forecasting: The forecast package for R. *Foresight*, 8(4):43–46.

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

Hyvärinen, A. and Oja, E. (2001), *Independent Component Analysis*. Wiley.

Jolliffe, I. (2002), *Principal Component Analysis*. Series in Statistics. Springer-Verlag, New York, 2nd edition.

Kaipia, R., I. Dukovska-Popovska, and L. Loikkanen (2013), Creating sustainable fresh food supply chains through waste reduction. *International Journal of Physical Distribution and Logistics Management*, 43, 3, 262–276.

Kanet, J.J., M.F. Gorman, and M. Stößlein (2010), Dynamic planned safety stocks in supply networks. *International Journal of Production Research*, 48, 22, 6859–6880.

Kouki, C., Jemai, Z., Sahin, E., Dallery, Y., (2014), Analysis of a periodic review inventory control system with perishables having random lifetime, *International Journal of Production Research* 52, 1, 283–298,

Lawrence, M., Goodwin, P., O'Connor, M., Önkal, D., (2006), Judgmental forecasting: A review of progress over the last 25 years, *International Journal of Forecasting*, 22(3), 493–518.

Ma, Y., Wang, N., Che, A., Huang, Y., and Xu, J. (2013), The bullwhip effect on product orders and inventory: a perspective of demand forecasting techniques. *International Journal of Production Research*, 51(1):281–302.

Leung, Y.T., and Bhaskaran, K., (2011), A Practical Approach to Diagnosing and Tuning a Statistical Forecasting System. In: K.G. Kempf et al. (eds.), Planning Production and Inventories in the Extended Enterprise, International Series in Operations Research & Management Science, 152, Springer Science+Business Media, New York, 289–311.

Makridakis, S., and Hibon, M. (2000), The M3-competition: results, conclusions and implications. *International Journal of Forecasting*, 16(4):451–476.

Makridakis, S., Wheelwright, S. C., and Hyndman, R. J. (2008), *Forecasting Methods and Applications*. Wiley India Pvt. Limited, 3rd edition.

MATLAB (2014), *version 8.4.0 (R2014b)*. The MathWorks Inc., Natick, Massachusetts.

McKenzie, E., Gardner E.S.Jr., (2010), Damped trend exponential smoothing : A modelling viewpoint. *International Journal of Forecasting*, 26(4), 661–665.

McCarthy, T. M., Davis, D. F., Golicic, S. L., Mentzer, J. T., (2006), The evolution of sales forecasting management: a 20-year longitudinal study of forecasting practices, *Journal of Forecasting*, 25(5), 303–324.

McKay, M., Beckman, R., and Conover, W. (1979), A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245.

Moreira, M., B. Tjahjono (2016), Applying performance measures to support decision-making in supply chain operations: a case of beverage industry. *International Journal of Production Research*, 54, 8, 2345–2365.

Muriana, C. (2015), Effectiveness of the food recovery at the retailing stage under shelf life uncertainty: An application to Italian food chains. *Waste Management*, 41, 159–168.

Nahmias, S. (2011), *Perishable Inventory Systems*. Springer.

Önkal, D., Sayim, K. Z., and Gönül, M. S., (2013), Scenarios as channels of forecast advice, *Technological Forecasting and Social Change*, 80(4), 772–788.

Petropoulos, F., S. Makridakis, V. Assimakopoulos, K. Nikolopoulos (2014), Horses for Courses in Demand Forecasting. *European Journal of Operational Research* 237, 152–163.

Rijpkema, W.A., R. Rossi, J.G.A.J. Van Der Vorst (2014), Effective sourcing strategies for perishable product supply chains. *International Journal of Physical Distribution and Logistics Management*, 44, 6, 494–510.

Simangunsong, E., L.C. Hendry, M. Stevenson (2012), Supply-chain uncertainty: a review and theoretical foundation for future research, *International Journal of Production Research*, 50:16, 4493–4523.

Soysal, M., J.M. Bloemhof-Ruwaard, Meuwissen, M.P.M., J.G.A.J. Van Der Vorst (2012), A Review on Quantitative Models for Sustainable Food Logistics Management. *International Journal on Food System Dynamics*, 3, 2, 136–155.

Stevenson, W. (2015), *Operations Management*. Series in Operations and Decision Sciences. McGraw-Hill Education, 12th edition.

Syntetos, A. A., Nikolopoulos, K., Boylan, J. E., Fildes, R., Goodwin, P., (2009), The effects of integrating management judgement into intermittent demand forecasts, *International Journal of Production Economics*, 118(1), 72–81.

Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., Nikolopoulos, K., (2015), Supply chain forecasting: Theory, practice, their gap and the future, *European Journal of Operational Research*, 1(26), 1–26.

Tan, K. H., Ji, G., Lim, C. P., Tseng, M. L. (2017), Using big data to make better decisions in the digital economy. *International Journal of Production Research*, 55, 17, 4998–5000

Yakovleva, N., J. Sarkis, T. Sloan (2012), Sustainable benchmarking of supply chains: the case of the food industry. *International Journal of Production Research*, 50:5, 1297–1317.

Yang, Y., Pan, S., Ballot, E., (2017), Innovative vendor-managed inventory strategy exploiting interconnected logistics services in the Physical Internet. *International Journal of Production Research*, 55:9, 2685–2702,

Wagner, M., (2002), Demand planning. In: Stadtler, H., Kilger, C. (Eds.), Supply Chain Management and Advanced Planning, 2nd Edition. Springer, Berlin, 123–141.

Wagner, M., Meyr, H., (2002), Food and beverages. In: Stadtler, H., Kilger, C. (Eds.), Supply Chain Management and Advanced Planning, 2nd Edition. Springer, Berlin, 371–388.

Wang, X., Petropoulos, F., (2016), To select or to combine? The inventory performance of model and expert forecasts. *International Journal of Production Research*, 54, 17, 5271–5282.

Wang, X., D. Li, C. O'Brien (2009), Optimisation of traceability and operations planning: an integrated model for perishable food production. *International Journal of Production Research*, 47, 11, 2865–2886.

Van Der Vorst, J.G.A.J., A.J.M. Beulens, W. De Wit, P. Van Beek (1998), Supply Chain Management in Food Chains: Improving Performance by Reducing Uncertainty. *International Transactions in Operational Research*, 5, 6, 487–

499.

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