Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 149 (2019) 3–10

Procedia Computer Science

www.elsevier.com/locate/procedia

ICTE in Transportation and Logistics 2018 (ICTE 2018)

Demand forecasting in pharmaceutical supply chains: A case study

Galina Merkuryeva[a,]*, Aija Valberga[b], Alexander Smirnov[c]

[a]Riga Technical University, 1 Kalku Str., Riga, LV-1050, Latvia
[b]Frankopharm Group SIA, 18B Uriekstes Str., Rīga, LV-1005, Latvia
[c]St. Petersburg Institute for Informatics and Automation of the Russian Academy of Science, 39, 14 Linia,St. Petersburg, 199178, Russia

Abstract

Demand forecasting plays a critical role in logistics and supply chain management. In the paper, state-of-art methods and key challenges in demand forecasting for the pharmaceutical industry are discussed. An integrated procedure for in-market product demand forecasting and purchase order generation in the pharmaceutical supply chain is described. A case study for supply of pharmaceutical products from a wholesaler to a distribution company located in an emerging market is presented. Alternative forecasting scenarios for thebaseline demand calculations using the SMA model, multiple linear regressions and symbolic regression with genetic programming are experimentally investigated, and their practical implicationsare discussed.

Keywords: Demand forecasting; Phamathetical supply chain; Logistics; Multiple linear regression; Symbolic regression

1. Introduction

Demand forecasting is an integral part of business process management. Despite complexity and execution of forecasting processes across different businesses, the intended purpose stays the same: obtaining a fairly accurate estimation of future demand for a product or service given historical data and the current state of the environment (e.g., political, social, economic) to plan and organize businesses accordingly. Forecasting accuracy is still a big challenge in the pharmaceutical industry [1].

* Corresponding author. Tel.: +371-26428694.
E-mail address: galina.merkurjeva@rtu.lv

10.1016/j.procs.2019.01.100

Demand forecasts form the basis of all managerial decisions in logistics and supply chain management. Regardless of a push or pull type of a supply chain system, demand forecasting is the starting point for all planning activities and execution processes. Consider push processes that are performed in anticipation of customer needs - sourcing, production, transportation, operating activities and actions – all require demand forecasts as data input; the same is true for the pull processes – to plan necessary levels of activity and inventory the customer demand data should be a starting point.

Pharmaceutical industry is known as one of the strongest in R&D industries that paid comparatively less attention to development of supply chain technology. High margins gained from sales of original products allowed the industry to have high supply chain costs [2]. Expiration of patents [2, 3] and, as a result, a considerable increase in the number of generic production companies, who focus on the development of efficient, effective and less costly supply chains, require the pharmaceutical industry to turn its attention to challenges in supply future demand forecasting and inventory management, thus confirming the importance of the effectiveness of a supply chain for the industry further development.

Moreover, entering into emerging markets requires the expansion of supply chains to be more cost effective compared to ones operating in developed economies as the amount of money spending on medicines in emerging markets is relatively low [3, 4]. Distances between distribution centers (mainly, in Europe, and USA) and emerging markets (Uzbekistan, Turkmenistan, Kazakhstan, etc.) put the efficiency of logistics and supply chain on the top spot. Mix of demand forecasts with high error rates and long delivery lead times causes oversupplies and overstocks. If the distance between the market and the point of sales is fixed, the demand forecasting error can be decreased by using more efficient and advanced demand forecasting methods. The actuality of the forecasting task has been evidently recognized and valued by top-level management [5].

2. State-of-the-art

The supply chain of pharmaceutical products is characterized by high complexity, and supply and delivery channels to customers are limited and highly regulated. The complexity is considered as one of the main barriers to performance and efficiency improvements of a pharmaceutical supply chain [4].

There is no room for error as it might directly affect health of population and nations in a negative way. This is the reason why the costs of a stock outs for pharmaceutical products cannot be expressed only in monetary terms. Avoiding out-of-stock and forecast accuracy of the customer demand lead to high inventory levels in comparison with the consumer-centric best practices [2].

The most commonly used demand forecast accuracy metrics to assess the quality of forecasts are variations of the following basic metrics: Forecast Bias, Mean Absolute Deviation (*MAD*), Mean Squared Error (*MSE*)*, and* Mean Absolute Percentage Error (*MAPE*).Recommendations on how to select an appropriate accuracy metricand how to use it and monitorcould be found in [6].There are a few more things that should be considered when deciding on how to measure forecast errors or accuracy at each level of a supply chain. In particular, interrelations between accuracy based measures and forecast aggregation levels over the products and forecasting timeline should be taken into account and need to be matched by selected accuracy metrics. In addition, it is well known that forecast errors are dependent on forecasters' location along the chain. This means that forecasts created closer to the demand point will be more accurate, while forecasts created further up the supply chain will have larger forecast errors. Let's note that for the upstream organizations collaborative forecasts based on sales to the end customer allow to reduce forecast error [7].

A variety of forecasting methods have been developed based on two well-known approaches to forecasting: qualitative and quantitative. Correspondingly, qualitative methods such as Executive opinions, Delphi technique, Sales force polling and Customer services generate forecasts based on judgements or opinions, while quantitative techniques may be grouped under historical data forecasts, e.g., Naive method, Trend Analysis, Time Series Analysis, Holt's and Winter's models, or under so called associative forecasts which identify causal relationships between variables using Simple, Multiple or Symbolic regression. In addition, mixed or combined models enable integration of both approaches. In the pharmaceutical industry, time-series models are used most often (52%) and causal models account for 24%, while judgmental – for 19% and remaining 5% represent mixed or combined models [8].

Demand forecasting for pharmaceutical products is also dependent on the product lifecycle; for new and already existing (in-market) products it varies significantly [9]. A reasonable estimate of a market size at the product development stage is a big challenge. At this stage, judgemental methods are more likely used for forecasting as there are no quantitative data available. For in-market products one should also estimate potential market share considering future market growth or decline while shifting to quantitative or mixed models.

Benchmarking studies show that forecasting is a relatively new task for the pharmaceutical industry. This might explain the dominant position (82.1%) of such simple methods as smoothing, average and naïve performed mainly in Excel spreadsheets being the most common type of software used [10]. In these cases, the forecasting error may vary to around 40% [5]. More sophisticated methods have emerged in the last decades [11-13].Using system dynamics modelling approach to forecasting problems allow creating models to track decease progressing over time and create loops in the forecast models. Simulation provides effective tools for performing validation of predictive models. Data visualisation leads to better collaborative decision making and forecasting. Intelligent demand forecasting for pharmaceutical products by exploring technologies from Artificial Intelligence improves the forecasting process and accuracy of demand forecasts. But the industry still prefers time-series models sprinkled with a judgmental (qualitative) approach [4]. The benchmarking studies also show that although there is plenty of data useful for more accurate demand forecasting (e.g., promotion activities and their effects on sales), data usage is limited due to various aspects, e.g., different data formats; lack of data integration tools; data collection times and data actuality; and lack of new models for increasing forecasting intelligence.

3. Case study

The *Pharma_Log* acts as a wholesaler and being a part of a supply chain for pharmaceutical products provides a link between manufacturing companies in EU and CIS and distributors in Central Asia (see Fig. 1). The affiliated distributor *Pharma_Dis* is located in the emerging market and resells only in-market products. While the *Pharma_Dis* has direct access to market and sales data records, the *Pharma_Log* has access to data provided by manufacturers (historical sales data, product availability, future product changes, stockouts, planned marketing activities, etc.).

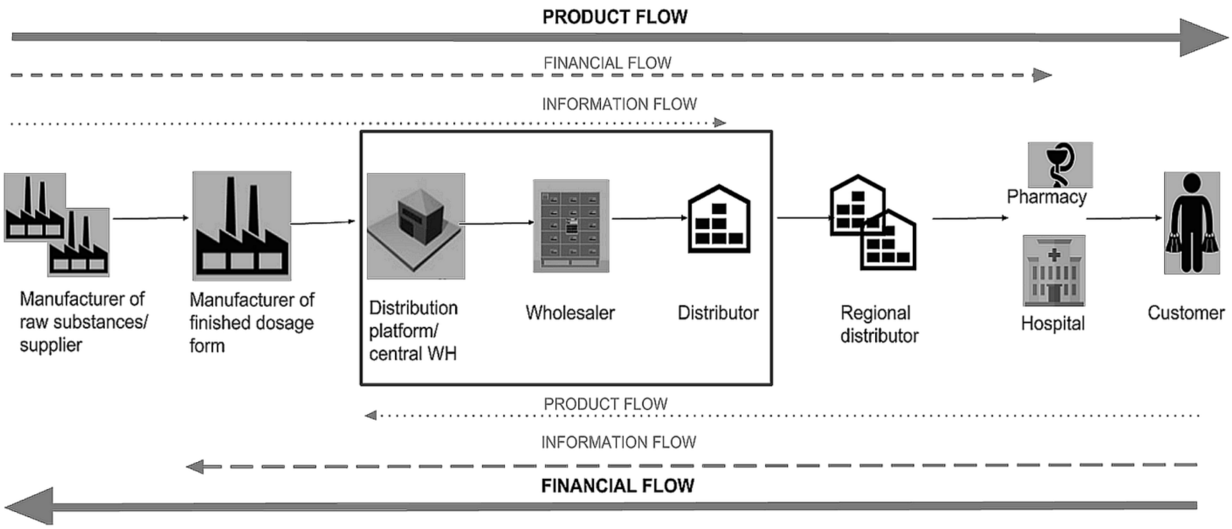

Fig. 1. Supply chain of pharmaceutical products.

The *Pharma_Dis* is a startup company and faced with increasing challenges in demand forecasting. On the one hand, there is lack of historical data sets as it is operating during the period less than a year. On the other hand, volatility comes in with an emerging market environment.

As a result, available population data for demand forecasting has relatively high fluctuations (see Fig. 2) which are difficult to model and hard to fit within such a short data range.

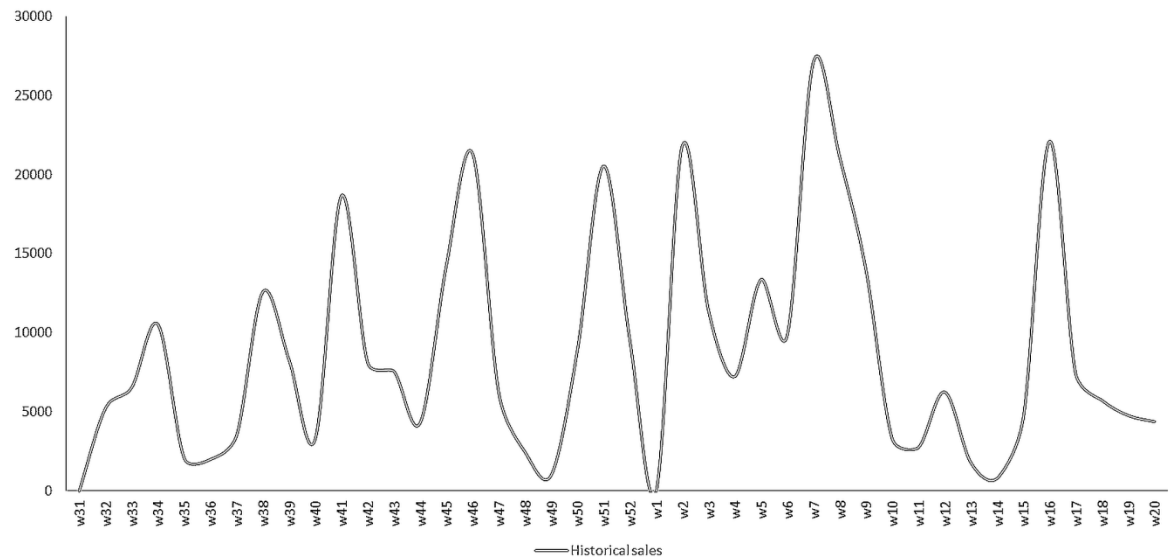

Fig. 2. Historical sales for the product family.

Currently, purchase orders are generated based on product demand projected by regional distributors based on the sales data and stocks available at the distribution platform. The demand forecasting function is performed by the wholesaler using market sales data from the distributor. Weekly forecast is calculated based on the moving average of the last 13 weeks of sales. No additional data is considered while historical sales of the product provided by a manufacturer might be of a high value and can be potentially integrated into demand forecasting and purchase order processing.

Availability of market data and access to manufacturer's data provides sufficient ground for setting up collaborative forecasts as a joint project of two companies. In addition, collaborative demand forecasting will allow introducing mechanisms for sharing demand information and incorporating the jointly derived demand forecasts into replenishment decisions in a supply chain.

4. Approach used

The general algorithm used for in-market product forecasting [9] includes the following main steps:

- Trending historical data sets to examine market and product performance
- Applying the effects of ex-trend events which occur but are not reflected in the historical data
- Converting trended data into the forecast outputs. In the simplest case, trending of historical data is performed by drawing a projected line (trend) into future

Methods for obtaining market and product baseline functions via forecast calculations are analysed in the next section. To identify ex-event and quantify their effects on the forecast appear to be the challenging tasks for forecasters. Converting trended data to forecasts outputs allows integration of supply and demand data. Further details can be found in [9].

The above mentioned steps have been applied to create an integrated procedure for in-market pharmaceutical product demand forecasting and purchase orders (PO) generation (see Fig. 3).

The functional components of the procedure define the process workflow, i.e.,selection of historical sales data and their conversion to demand data, baseline trend calculations and analysis, application of ex-trend events and productdemand quantification followed by calculations of purchase orders and their adjustment based on the product availability and product shelf life limitations.

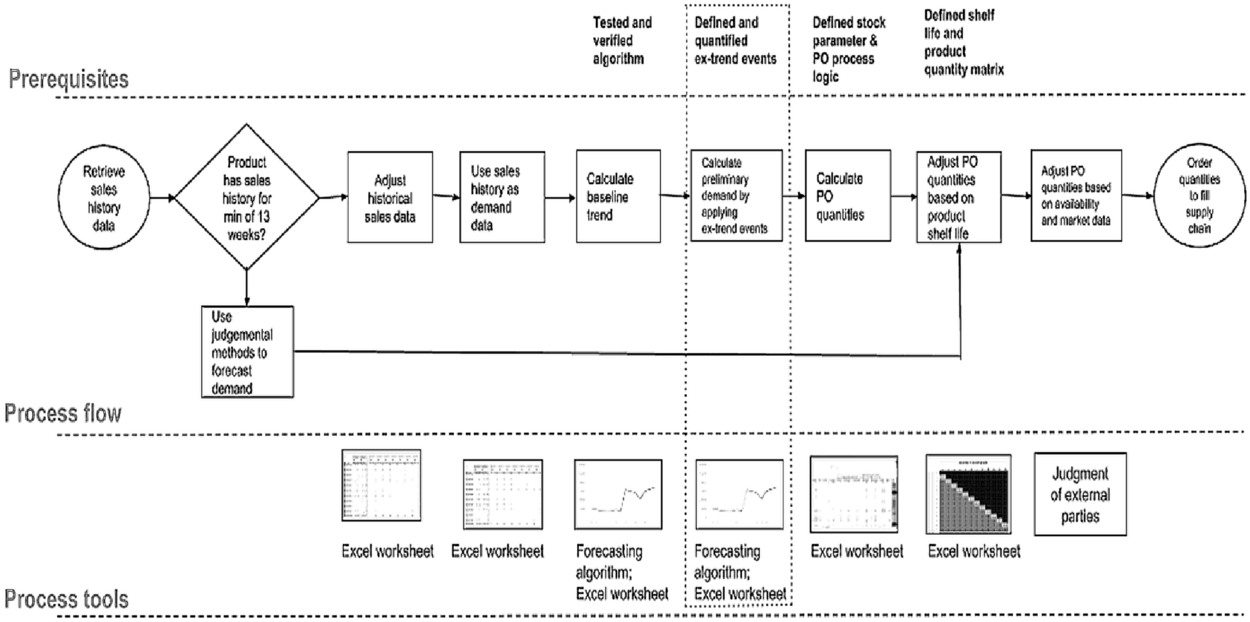

Fig. 3. Integrated procedure for in-market product demand forecast and purchase order generation.

5. Experiments

In the case study, demand forecasting experiments have been performed for a specific pharmaceutical product ACT0002UZ01. The historical weekly sales product data contains 41 data points. Three experimental scenarios based on application of different forecasting methods are investigated applying the simple moving average method, multiple linear regression, and symbolic regression with genetic programming. For regression scenarios, the following factors potentially affecting the product demand have been taken into consideration: a distributor price-list; the discounted selling price of the product; a week number of sales in a month; and weekly average currency rate. For each scenario, forecasting output and forecast errors are analysed and applications feasibility and implications are provided.

5.1. Scenario 1

The forecasting scenario based on the simple moving average (SMA) method is introduced as the initial or current scenario used in practice. Forecast calculations have been performed for a number of periods equal to 4 and to 13 weeks taking into account the average frequency in data fluctuations and based on general practice applied by manufacturing companies operating in the emerging market. In both cases, the absolute error is high comparing against the total demand observed (see, Table 1). With the average of absolute deviations over all periods (MAD), the estimated standard deviation of the forecast error is large in relation to the actual data. Finally, the forecast appears to be biased and under-estimated as a tracking signal value range (TS) falls below -6 and reaches its minimum values, -8.3 and -6.2, respectively.

The SMA model falls to adaptive forecasting thatenables updating the moving average value of demand after each observation appeared, and therefore it reacts on the latest changes in the demand pattern. In the experiment, historical demand data are overcomplicated and do not clearly correspond to periodic demand fluctuations. That is why the SMA model generates inconsistent and inaccurate sales forecasts (see, also Fig. 4).

5.2. Scenario 2

In the second series of experiments, the correlation and linear regression analysis have been conductedto determine the causal relationship between product demand and demand influencing factors.Two sequential regression models were constructed. Each model is represented by a linear algebraic equation with more than one independent variable. All calculations have been performed using MS Excel built-in tools.

In iteration 1, a linear regression model with 3 independent variables such as base price, discounted price and a week number of sales for a month (WeekNoM) has been built. In this case, the results for the correlation show a strong relationship between product demand and product discounted price, while WeekNoM variable is taken into consideration just based onthe expert judgment. As a result, the demand regression equation looks as follows:

$$Demand \approx 17\,660 - 0{,}089 * Baseprice - 0{,}75 * Discountedprice + 42.824 * WeekNoM \quad (1)$$

Because the P value is equal to 0,678 in the ANOVA table, the effect of the base price is not considered as statistically significant and excluded from further consideration.

In iteration 2, the demand regression model with two independent variables is described by the following linear equation:

$$Demand \approx 15\,876 - 0{,}752 * Discountedprice + 43.128 * WeekNoM \quad (2)$$

Compared to the SMA model, the absolute error and mean absolute deviation are more than 50% lower in case of regression based forecasts (see Table 1). Forecasting charts calculated based on sequential regression models (1) and (2) are shown in Fig. 4. So, multiple criteria linear regressions show better results than the SMA model and have better capability to reproduce behavior of the demand pattern. However it still lacks the ability to accurately predict demand peak sales. The TS range fall below -6, and the forecasts still are considered as consistently underestimated. Further in depth analysis is needed to explore linear regression-based forecasts and check if non-linear regression functions (e.g., polynomial, exponential, logarithmical) can better fit to the historical dataset. Besides, non-price factors that can influence product demand such as population size, population ageing, male and female ratio, etc., have to be taken into consideration and systematically examined.

5.3. Scenario 3

In the last series of experiments, the symbolic regression-based forecasting model is built and its performance is analysed. Symbolic regression enables to find mathematical expressions in a symbolic form which better fits problem data and predicts a dependent variable from explanatory variables with the smallest error [13-15]. The structure of the model is not predefined, and both the symbolic form of the model and coefficients for model variables should be determined.

Symbolic regression-based forecasting experiments have been performed using the preconfigured tree-based (Koza-style) genetic programming (GP) algorithm for producing symbolic regression models in HeuristicLab software [13]. The problem data set is the same data set being analysed in the linear regression scenario (iteration 2).

The maximum number of generations is experimentally defined equal to 50 generations. A population size is equal to 1000 individuals. Fitness function is evaluated by Pearson R^2 coefficient. Available tree nodes are real value constant in a range of [-20; 20], explanatory variables, arithmetic functions, exponential and logarithmic functions. The maximum tree depth and length are defined by 10 and 25 nodes, respectively. The first third of historical sale records is used to train the model, and the remaining ones to test it.

In this scenario, 25 GP experiments have been performed, The model with the best found fitness is expressed in an exponential form:

$$Demand \approx EXP(EXP((1/((1/(-0.889 * WeekNoM) + 0.362)) + 27444.607)/(0.413 * Discountedprice))) * 4.885E - 10 + 212.491) \quad (3)$$

The model fits data with high coefficient values: $R^2 \approx 0.934$ for the training set and $R^2 \approx 0.824$ for the test set. It has higher accuracy than a linear model (2) obtained from the same data. The model performance parameters are given in Table 1. Particularly, the symbolic regression with an application of genetic programming showed the lowest absolute error and mean absolute deviation values across all forecasting scenarios and experiments.

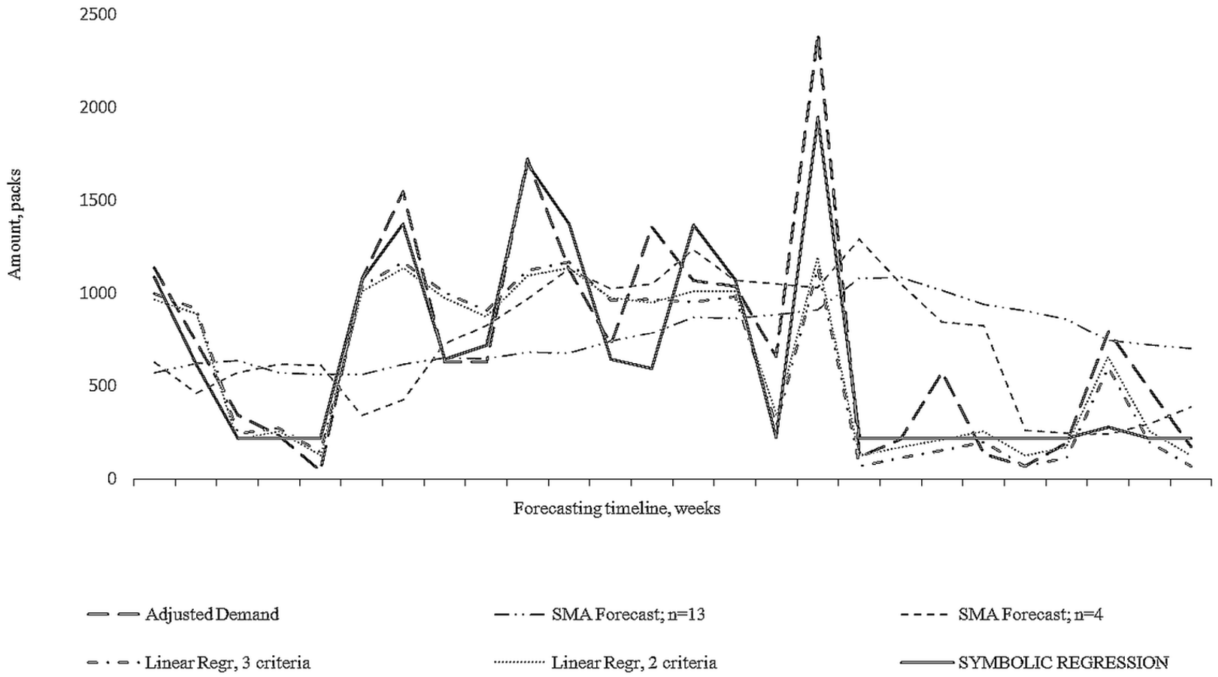

Fig. 4. Historical demand data and experimental demand forecasts.

Table 1. Summarized error estimates for demand forecasting for all experiments.

Forecasting method	Total Demand	Absolute Error	MAD	Standard deviation of forecast error	TS Interval
SMA for n =13	19 754	13 577	12 361	15 451	[-8.3, 4.8]
SMA for n =4	19 754	11 964	12 465	15 581	[-6.2, 1.1]
Linear regression,3 cr.	19 754	6 224	5 725	7 156	[-16.0, 1.8]
Linear regression,2 cr.	19 754	5 718	5 606	7 007	[-15.2, 0.4]
Symbolic Regression	19 754	4704	3 741	4 676	[-12.7, 0.3]

The results of demand forecasting experiments were discussed by invited experts operating in the field of logistics and supply chain management in the pharmaceutical company. Symbolic regression based forecasting has been selected as the most appropriate method in the study.

6. Conclusion

The supply chain of pharmaceutical products is characterized by high complexity which is considered as one of the main barriers to improving performance of a pharmaceutical supply chain. Demand forecasts form basis for all strategic and planning in pharmaceutical logistics and supply chain management. Benchmarking studies show that forecasting is a relatively new task for the pharmaceutical industry that might explain the dominant position of simple methods performed mainly in Excel spreadsheets. More sophisticated forecasting techniques have emerged in sales and demand forecasting applications, e.g., system dynamics modelling approach to track decease progressing over time and create loops in the forecast models; simulation to perform validation of predictive demand models; data visualisation to support collaborative forecasts and intelligent forecasting to increase forecasting intelligence in supply chain applications.

The case study given in the paper is based on the real life example within the pharmaceutical field. The results of experimental analysis of three forecasting scenarios show that symbolic regression-based forecasting model provides the best fitting curve to history demand data, lower error estimates across all scenarios and performed experiments, and the ability to more accurately predict demand peak sales in the study.

References

[1] Durbha, D. (2016) “State of the Pharmaceutical Supply Chain: Key Takeaways from LogiPharma US 2016“. 21st Century Supply Chain Blog. Available: https://blog.kinaxis.com/2016/09/state-pharmaceutical-supply-chain-key-takeaways-logipharma-us-2016/. [Accessed April 30, 2018].

[2] Preparing the Supply Chain Pharma Needs. (2014) “A.T.Kearney Pharma Supply Chain Panel 2014“. A.T. Kearney, Inc. Available: https://www.atkearney.com/web/the-purchasing-chessboard/article/-/asset_publisher/9AutfSQfJm6Y/content/preparing-the-supply-chain-pharma-needs/20152. [Accessed April 30, 2018].

[3] PWC (2011) “Pharma 2020: Supplying the Future, Which pass you take”. *Pharmaceuticals and Life Sciences*. Available: https://www.pwc.ch/en/publications/2016/pharma-2020-supplying-the-future.pdf. [Accessed February 31, 2018].

[4] Lukas, U., and M. Mencer. “A Cure for Pharmaceutical Supply Chain Complexity”. Patheon, *OneSource*™. Available: https://www.patheon.com/onesource/pdf/A-Cure-For-Pharmaceutical-Supply-Chain-Complexity-Patheon-OneSource-WhitePaper.pdf. [Accessed February 31, 2018].

[5] Chaman, L. Jain, Jack Malehorn. (2006) “Benchmarking Forecasting Practices: A Guide to Improving supply chain performance”, Graceway Publishing Co., 119.

[6] Measuring Forecast Accuracy: The Complete Guide. RELEX Solutions. Available: https://www.relexsolutions.com/measuring-forecast-accuracy/. [Accessed November 10, 2018].

[7] S. Chopra, S, and P. Meindl. (2007) “Supply Chain Management: Strategy, Planning, and Operation”. Pearson Education, Inc., 536.

[8] Jain, C.L. (2003) “Benchmarking Forecasting Practices in Pharmaceutical Industry”, St. John's University, Jamaica, New York, 6.

[9] Cook, A.G. (2006) “Forecasting for the Pharmaceutical Industry – Models for New Products and In-Market Forecasting and How to use them”. Gower Publishing Co., 141.

[10] Weller, M.,and S. Crone. (2012) “Supply Chain Forecasting: Best Practices & Benchmarking Study”. Lancaster Centre for Forecasting, 41.

[11] Candan, G., M. F. Taşkin, and H. R. Yazgan. (2014) “Demand Forecasting in Pharmaceutical Industry Using Artificial Intelligence: Neuro-Fuzzy Approach“, *Journal of Military and Information Science*, 2(2): 41-49.

[12] Pharmaceutical & Biotech Supply Chain Software Solutions. Demand Solutions. Available: https://www.demandsolutions.com/industry-pharmaceutical-supply-chain-biotech-supply-chain.html/. [Accessed November 10, 2018].

[13] Wagner, S., G. Kronberger, and M. Affenzeller et al. (2014) “Architecture and Design of the HeuristicLab Optimization Environment”. Advanced Methods and Applications in Computational Intelligence, Topics in Intelligent Engineering and Informatics Series, Springer, 197-261.

[14] Bolshakov, V. (2013) “Regression-based Daugava River Flood Forecasting and Monitoring”. *Scientific Journal of RTU*, *Series 5, Computer Science:* 137-142.

[15] Merkuryeva, G., Y. Merkuryev, and B. Sokolov et al. (2015) “Advanced river flood monitoring, modelling and forecasting. *Journal of Computational Science*”, **5(10)**: 75-85.

Galina Merkurjeva is Professor in the Department of Modelling and Simulation, Faculty of Computer Science and Information Technology at the Riga Technical University. She holds Dr.sc.ing. from the Institute of Electronics and Computer Science of the Latvian Academy of Sciences (Latvia) and Dr habil. (DSc.) from the Institute of Control Sciences of the Russian Academy of Sciences (Russia). Her professional interests and experiences are in the fields of discrete-event simulation, simulation metamodelling and optimisation, artificial intelligence, logistics and supply chain management. She is the author of more than 200 publications including 6 books in the field. Contact information: Riga Technical University, 1, Kalku Street, Riga, LV-1658, Latvia. Contact her at Galina.Merkurjeva@rtu.lv.