Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

USING SIMULATION TO ASSESS THE RELIABILITY OF FORECASTS IN HIGH-TECH INDUSTRY

Bhoomica M. Nataraja
Nitish Singh
Ivo Adan

Dept. of Industrial Engineering and Innovation Sciences
Eindhoven University of Technology
PO Box 513
Eindhoven, 5600MB, THE NETHERLANDS

Tanmay Aggarwal

Lambda Function, Inc.
1960 Decanter Circle
Brentwood
California, 94513, USA

Koen Herps

KMWE Headquarters
Brainport Industry Campus (BIC) 1
Eindhoven, 5657BX, THE NETHERLANDS

ABSTRACT

In a high-tech production environment, capacity investment and production planning are often based on the demand information from manufacturers within a supply chain. A supplier solicits forecast information from a manufacturer, and the manufacturer provides demand forecasts that are updated on a rolling horizon basis. Problems arise with this setup if the manufacturer provides volatile forecast quantities due to the market's fluctuating demand or internal bias. As a result, suppliers' mistrust regarding forecast quantities grows, leading to adjusted production plans based on planners' anecdotal experience. The paper presents a decision model to determine the reliability of forecasts provided by manufacturers to facilitate better production planning. The study also suggests alternate forecasting techniques in case of low reliability. To evaluate the effectiveness of the proposed approach, a simulation study is conducted for different manufacturers and scenarios. Our experiments showed an average cost reduction of 14% across all instances.

1 INTRODUCTION

Several manufacturing companies (suppliers) manage their supply chains based on the demand forecasts provided by their customers (manufacturers), especially when there are contractual agreements to meet the forecasted demand. The customer's forecast is usually transmitted and updated in a rolling horizon using Electronic Data Interchange (EDI). However, it does not suffice only to have access to customers' forecasts. A supplier must perceive the forecast information to fulfill the expectations. However, the interpretation of forecast data depends on the quality of the forecast information. The forecast could, for example, be changed so drastically that the supplier does not trust it. Such situations could lead to under-production or over-production. The supplier incurs a penalty in both cases. To avoid the risks in supply, it is necessary to estimate the reliability of the forecasts based on which production and capacity decisions are taken.

In the context of forecasting, reliability refers to the consistency and stability of a forecast over time or the ability of a forecast to be replicated with similar results if the forecasting process were to be repeated. Reliability is desirable as it ensures the forecast can be used to make decisions. Reliability differs from accuracy. Accuracy refers to how closely a forecast matches the actual outcome. Common accuracy measures used in forecasting include mean squared error (MSE), mean absolute deviation (MAD), and

979-8-3503-6966-3/23/$31.00

mean absolute percentage error (MAPE). Accuracy does not consider the forecast's consistency and stability over time, which is why we consider reliability a separate but important concept. An example is a supplier using a forecasting model to predict sales for the upcoming quarter. If the forecasted sales are consistently within a certain range of the actual sales, it demonstrates the reliability of the model. On the other hand, if, even though the forecasted sales are very close to the actual sales, the forecasted sales are inconsistent and unstable over time, it can be considered that the model is not reliable.

This work originates from a real-world case study at KMWE Precision, henceforth referred to as KMWE. The company, headquartered in Eindhoven, The Netherlands, is an aerospace parts manufacturer of structural and engine components. KMWE serves a multitude of customers, making its product portfolio highly diversified. Product family architecture (PFA) is one of the applied methods to face a strictly heterogeneous production environment (Tseng et al. 1996). Products are classified based on a wide range of production characteristics, and the manufacturing of these parts is planned accordingly. In this case, a reliable forecast is essential for production planning, i.e., to allocate jobs to a predefined time window. Otherwise, an unreliable forecast (such as changes in quantities, delivery dates, and product types) diminishes the possibility of applying PFA procedures effectively. Moreover, KMWE faces long lead times on purchased materials (such as certified materials for the aerospace market) from its suppliers. When the forecast changes within the lead time of this supply, either back-orders or abundant inventory is inevitable.

In this study, long-term demand forecasts are considered in a make-to-stock environment. For example, KMWE (supplier) seeks forecast information from Airbus (manufacturer) to make specialized aerospace components. Unlike the consumer goods industry, the demand for high-tech products has high fluctuations. Therefore, manufacturers are encouraged to place their orders in advance to reduce demand uncertainty. Thus, the manufacturer provides a long-term forecast to the supplier on a rolling horizon with periodic updates (e.g., every month). The supplier then has multiple forecasts, one from each period, for the same product. The supplier should be able to make an informed decision based on actual scientific facts and not rely on anecdotal experience. Thus, it is helpful if the supplier is aware of the reliability of the forecast for each period offered by the manufacturer.

In this paper, we introduce a decision model to evaluate the reliability of the forecast offered by the manufacturer. First, a reliability metric can be calculated for each period's forecast leading up to a given point in time. Then, actual historical data can be used to generate a new forecast. This approach generates a quantitative measure that can be used to make decisions based on the reliability of the forecast. We employ different forecasting techniques and use the data set of KMWE.

The remainder of this paper is organized as follows. Section 2 gives an overview of relevant literature. In Section 3, the proposed approach is described. Section 4 presents the implementation and results of our model. Finally, conclusions and directions for future research are discussed in Section 5.

2 LITERATURE REVIEW

In the make-to-stock industry, procure and produce decisions are often based on some interpretation of future demand or demand forecasts from downstream customers. Demand forecast provides critical input for various strategic decisions for supply chain management, including capacity planning, inventory control, and demand fulfillment (Cakanyildirim and Roundy 2002). Bueno et al. (2020) provided a detailed review of studies on the integration of demand forecasting with capacity and inventory management.

Several studies have emphasized the importance of sharing forecast information between firms (Forslund and Jonsson 2007; Ganesh et al. 2014; Shamir and Shin 2016). Supply chains are composed of independent firms with private information. Firms downstream (closer to the end customer) organize their production plans in anticipation of market fluctuations and share forecast information with their upstream suppliers. This information is only an intention of the firm and cannot be easily verified or employed in practice. Many of the suppliers have contractual agreements to fulfill these forecasted demands and place their customer order decoupling point further downstream to store items in their inventory holding establishments (Olhager 2010). There exist contracts that align incentives in the supply chain and induce the buyer/firm to reveal demand information truthfully, often as two-part tariff quadratic contracts (Desai and Srinivasan 1995), capacity commitment, and option contracts (Özer and Wei 2006). Contractual agreements create a 'win-win' situation for buyers and suppliers. However, in practice, linear pricing contracts are more prevalent (Ren et al. 2010). The suppliers thus use forecast information from downstream customers to organize their production plans to meet the firm's demand. When the forecasts are not shared truthfully or are inaccurate, the suppliers face the problem of overproduction and excessive stock holding costs or inefficient use of available capacity (Chen and Chien 2018). Bullwhip Effect also makes demand forecasting problems more difficult for upstream suppliers (Lee et al. 1997).

Forecasting demand has been a long-standing issue for many years. Traditional time-series methods, including moving average, exponential smoothing, and ARIMA, attempt to identify forecasting parameters such as trend cycle, seasonality, and irregularity, then extrapolate these components to produce forecasts. However, these trend-cycle and seasonal data components of a time forecast tend to evolve and need to be continuously revised for higher accuracy in forecasts. In addition, these models consider a linearity assumption, i.e., the activities responsible for influencing the past will continue to influence the future. This assumption is valid in forecasting short-term demand but falls short when attempting to forecast for a long-term analysis (Chase 2013). Recent studies have focused on using big-data analytics and machine learning approaches to estimate demand forecasts (Andersson and Jonsson 2018; Lee and Liang 2018; Fu and Chien 2019; Gonzalez-Vidal et al. 2019). With improvements in computation speeds and computing algorithms, machine learning approaches have gained traction as a solution approach for traditional data-driven problems in recent years. Huber and Stuckenschmidt (2020) used artificial neural networks and gradient-boosted decision trees to solve the forecasting problem. They show that machine learning methods are viable alternatives to traditional methodologies.

Studies in the past have highlighted the tendency of manufacturers to inflate their forecast quantities leading to increased supplier mistrust in provided forecasts (Cachon and Lariviere 2001). Simulation is used in many studies to evaluate the performance of the forecast models. Steinmann and de Freitas Filho (2013) showed how simulation can be used to generate data that can be used to evaluate incoming call forecasting algorithms. Zeiml et al. (2020) used a discrete event simulation model to evaluate the performance of the developed dynamic forecast correction model. The authors further used a simulation study to identify and correct outliers in the forecasts to improve forecast accuracy (Seiringer et al. 2021). However, little or no research has been done, either with or without the use of simulation, to quantify and measure the reliability of the forecasts provided by manufacturers periodically.

3 MODEL DESCRIPTION

In this section, we describe the reliability score estimation, the inference of the score, and alternate forecasting techniques. The combined method can serve as a decision-support tool for suppliers to address the unreliability of forecasts, which often extend to several end-customers and end-markets such as Aerospace, Semiconductors, and Medical.

3.1 Reliability Analysis

This section seeks to address the reliability problem by studying the forecast variability for each product across several time periods that lead up to the forecast month. The forecast data used in this model has the following structure. The forecasts obtained by the suppliers are on a rolling horizon and are periodically updated. A time period of one month offset from the actual forecasted month is defined as a *lag period*. For example, if the end-customer provides an 18-month forward-looking forecast for each product, we will have forecast data for 18 lag periods associated with the actual forecasted month. Thus, we have a finite time series of the forecast data successively taken each month for 18 lag periods. In Figure 1, lag periods and their August and September forecasts for a random product are presented. For instance, the demand for August was anticipated to be 37, in the previous period. Similarly, the demand was 46,

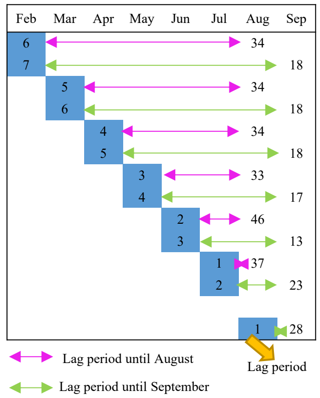


Figure 1: Illustration of lag periods for August and September forecasts.

33, 34, 34, and 34 in the forecasts sent two, three, four, five, and six periods ago, respectively. For each product/product family, a *forecast reliability score* is calculated for each lag period leading up to the actual forecasted month. Figure 2 displays a flowchart that includes the set of sequential steps undertaken for reliability analysis. After data pre-processing, the steps can be grouped into two major categories: time series forecasting analysis and calculation of the forecast reliability score.

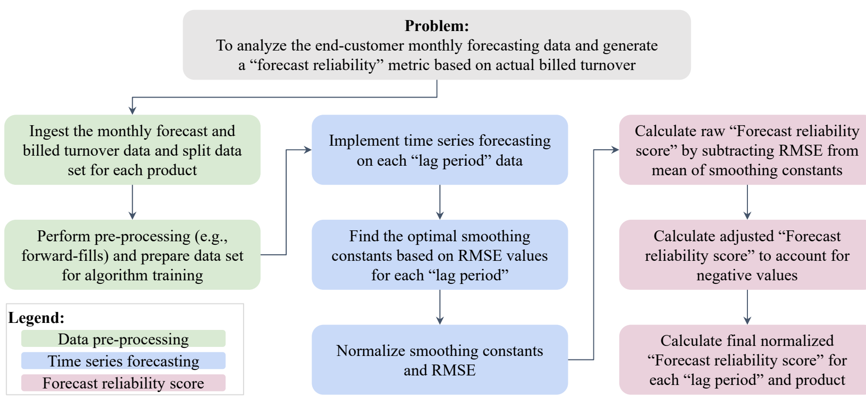


Figure 2: A flowchart displaying the steps of Reliability Analysis.

3.1.1 Time Series Forecasting

Let $x_{d,t}$ be the forecasted order quantity for any given month d stated t periods in advance (stated in t^{th} lag period). The first forecast arrives at $t = T$ periods before d, where T is the total number of lag periods. A time-series forecast approach is proposed to quantify the reliability, and the following principles guide it.

- Use a double exponential smoothing algorithm (LaViola 2003) on each lag period to generate a forecast and compare the forecast relative to the actual order quantity for all months. Exponential smoothing is a well-known forecasting method that weighs past values with exponentially decreasing weights to forecast future values, and it also adapts quickly to variability in forecast values. The forecast is,

$$P_t = \alpha x_{d,t} + (1-\alpha)(P_{t-1} + b_{t-1}) \tag{1}$$

$$b_t = \gamma(P_t - P_{t-1}) + (1-\gamma)b_{t-1}\ , \tag{2}$$

where $t \in \{1, 2, \ldots, T\}$, P_t is the forecast at lag period t, $x_{d,t}$ is the time series observation at lag period t for month d, b_t is the trend value obtained for lag period t, and $0 \leq \alpha, \gamma \leq 1$ are the smoothing constants for the level and trend, respectively.

- Use Root Mean Square Error (RMSE) as an evaluation metric to determine the forecast accuracy. RMSE measures the standard deviation of the difference between the forecast and actual billed turnover value. The RMSE is calculated for several α and γ values pairs for the previous T months, using the following equation (Willmott and Matsuura 2005).

$$\text{RMSE} = \left(\frac{1}{T} \sum_{t=1}^{T} (P_t - y_d)^2 \right)^{1/2}, \tag{3}$$

where P_t is the forecast value from the smoothing algorithm for the lag period t, y_d is the corresponding actual order quantity value for that month d.

3.1.2 Forecast Reliability Score

The forecast reliability score is determined for each lag period of every product using the optimal α, γ, and RMSE values resulting from the time series forecast analysis. The optimal values of α and γ for each lag period are those value pairs that minimize the RMSE value. The calculation of the score consists of the following steps.

- Scale the optimal α, γ and RMSE values using Min-max normalization. This normalization technique transforms the values into the range $[0, 1]$ so that the difference in scale does not create a bias.
- Calculate the mean of α and γ values over all lag periods (to be used as the single metric for smoothing constants).
- Subtract the RMSE from the mean of α and γ values to generate an initial value of the raw forecast reliability score.
- Remove negative score values by subtracting all values by the minimum value to obtain the adjusted forecast reliability score.
- Normalize the adjusted forecast reliability score to get the final forecast reliability score.

High forecast reliability is indicated by a small RMSE value and large smoothing constant values (i.e., large α and γ values). Low forecast reliability is indicated by a large RMSE value and small smoothing constant values (i.e., small α and γ values). Smoothing constants govern the level of influence of previous observations on the forecast. Thus, the reliability score is inversely and directly related to RMSE and smoothing constants, respectively. Smoothing constants close to one indicate a minimal change to the original forecast values, while those close to zero indicate significant changes.

3.2 Alternate forecasting techniques

Accurate quantification of end-customer demand forecast reliability is crucial for optimal production planning decisions. Equally important is the precise forecasting of order quantities based on historical data. When faced with low reliability in end-customer demand forecasts, it is essential for suppliers to rely on scientific evidence rather than anecdotal experience. In this section, we propose an alternative approach to address the challenge of low reliability scores. We aim to enhance the supplier's confidence by forecasting monthly order quantities for each product and product family based on actual historical sales volume data.

Figure 3 illustrates the sequence of steps taken for forecasting based on 'Billed Turnover Data', the actual realized demand in this study. After pre-processing, the data is normalized and split into training and test sets. Further, the following steps are carried out.

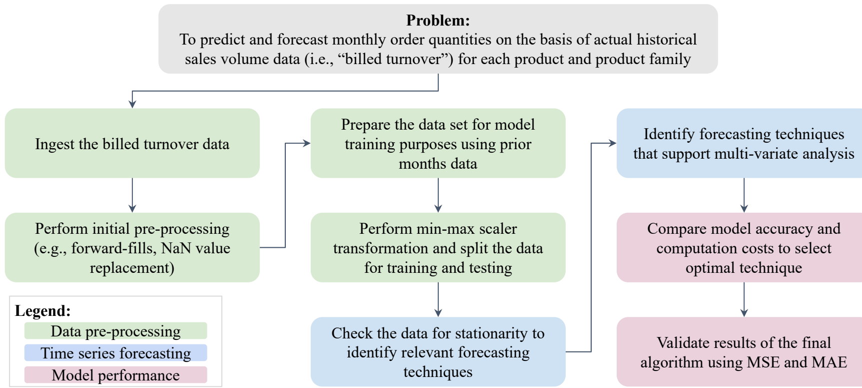


Figure 3: A flowchart displaying the steps of our proposed forecasting method.

Given the time series nature of the data, we consider some well-known methods to achieve the analytical objective of generating a monthly forecast based on historical billed turnover data. They are Autoregressive Integrated Moving Average (ARIMA), Prophet, Support Vector Regression (SVR) and Long Short-Term Memory (LSTM) (Drucker et al. 1996; Hochreiter and Schmidhuber 1997; Taylor and Letham 2018). The following principles guide this time series approach. In our study, we opted for established models such as ARIMA, Prophet, SVR, and LSTM over advanced models like transformers. This choice was motivated by the need for transparency and interpretability in the high-tech industry. Unlike transformers, which can be challenging to interpret due to their black-box nature, the selected models strike a balance between accuracy, interpretability, and applicability for our study's objectives in the high-tech industry.

- Train the models ARIMA, Prophet, SVR, and LSTM by fitting them on the training set.
- Test the models for accuracy by predicting results on the test set.
- Use Mean Absolute Error (MAE) and Mean Squared Error (MSE) as the metric of evaluation to determine model accuracy (Willmott and Matsuura 2005). MAE (Eq. 4) is the average of the absolute difference between the forecast and actual billed turnover value. MSE (Eq. 5) measures the variance of the difference between the two values.

$$\mathrm{MAE} = \frac{1}{T}\sum_{t=1}^{T} |x_{d,t} - y_d| \tag{4}$$

$$\mathrm{MSE} = \frac{1}{T}\sum_{t=1}^{T} (x_{d,t} - y_d)^2 \,, \tag{5}$$

  where t is the lag period equivalent to any given month d, $x_{d,t}$ is the forecast value predicted by the model, y_d is the corresponding actual order quantity value for that month, and T is the total number of lag periods.

Finally, the performance and accuracy of all the models are evaluated by comparing their MSE and MAE values. These values are quantitative measures that determine the proximity of predicted forecast values to the actual realized values. Therefore, low MAE and MSE values indicate a high model accuracy.

3.3 Simulation model

To evaluate the performance of the proposed methods, a discrete event simulation model was developed using the Simpy framework, a process-based discrete-event simulation framework based on standard Python. The simulation model replicates the production plant of KMWE, implementing the production process according to the PFA methodology. Each product in the simulation has a production lead time and needs to be released onto the production floor within the remaining lead time until the due date. The simulation model incorporates workforce allocation and resource management rules based on industry practices, calculating the required number of workers and machines based on the input forecast order quantity and due date. At each lag period, the reliability score is calculated, triggering changes in the forecast within the simulation model using empirical data. Various costs, including inventory holding costs, labor costs, operating costs, and back-ordering costs, are recorded throughout the simulation.

The evaluation of the proposed methods is conducted using three input scenarios:

1. Scenario A: The original forecast of the lag period at the start of production is used as input.
2. Scenario B: The lag period forecast with the highest reliability score is selected as input.
3. Scenario C: The forecast provided by the SVR model is used as input.

4 IMPLEMENTATION AND DISCUSSION

In Section 3, we presented a data-driven support tool that helps the suppliers in planning and making production decisions. In the subsequent sections, we discuss the implementation of the proposed methods and discuss the results of our empirical study.

4.1 Reliability Analysis

4.1.1 Implementation

Initially, a single exponential smoothing algorithm was applied to generate time series forecasts for each lag period of unique products. However, this approach did not accurately capture the underlying data, as indicated by the large RMSE values observed. Upon further investigation, it was identified that a systematic trend in the lag period forecast data was causing the poor performance of the single exponential smoothing algorithm. To address this issue, the double exponential smoothing algorithm was implemented, resulting in improved forecast accuracy.

The time series forecasting analysis of the dataset is performed according to the principles listed in Section 3.1. Setting the initial values of P_t and b_t, in Equations (1) and (2), is crucial to compute their subsequent values. To address this, we choose P_1 to be equal to first observation x_1, and explore two commonly used approaches for setting the value of b_1:

(1) $b_1 = x_2 - x_1$, where x_1 and x_2 are first and second observations, respectively.

(2) $b_1 = (x_n - x_1)/(n-1)$, where x_1 and x_n are first and last observations, respectively and n is the total number of observations.

The second critical step in the double exponential smoothing model is to set the best smoothing constant values, namely α and γ. An incremental method is used to solve this problem. This is an iterative procedure where we set the parameters: α in $[0.1, 0.9]$ with the increment 0.01 and γ in $[0.1, 0.9]$ with the increment 0.01. For each α and γ value pair, the corresponding RMSE value, using Eq. (3), between the predicted forecast and actual realized values is calculated. The best or optimal values of alpha α and γ, for a single lag period of a unique product, are those that yield the least RMSE value.

This time series analysis is performed repeatedly to identify the optimal α and γ values for all lag periods present across all products of the dataset. The first three columns of Table 1 show the optimal α, γ, and RMSE values calculated for each lag period of a randomly selected product. Then, the final forecast reliability score is determined for each lag period of all unique products. This final score is thus a quantitative measure of forecast reliability.

4.1.2 Discussion

The normalized reliability score in Table 1 displays the final forecast reliability score generated for each lag period of a randomly selected product. We observed an increasing trend of the final forecast reliability score for most products as the lag periods decreased. If the number of lag periods considered decreases,

Table 1: Forecast reliability scores for a randomly selected product for each lag period and their corresponding smoothing constants and RMSE values.

	Pre-normalization			Post-normalization						
Lag Periods	Optimal α	Optimal γ	Optimal RMSE	Optimal α	Optimal γ	Minimum RMSE	Mean of α and γ	Reliability Score	Adjusted Reliability Score	Normalized Reliability Score
18	0.29	0.25	1.53	0.88	1.00	0.94	0.94	0.00	0.08	4.53%
17	0.29	0.25	1.53	0.88	1.00	0.94	0.94	0.00	0.08	4.53%
16	0.30	0.25	1.56	1.00	1.00	1.00	1.00	0.00	0.08	4.60%
15	0.29	0.25	1.53	0.88	1.00	0.94	0.94	0.00	0.08	4.53%
14	0.27	0.24	1.42	0.63	0.80	0.71	0.71	0.00	0.08	4.50%
13	0.27	0.24	1.42	0.63	0.80	0.71	0.71	0.00	0.08	4.50%
12	0.27	0.24	1.42	0.63	0.80	0.71	0.71	0.00	0.08	4.50%
11	0.22	0.21	1.16	0.00	0.20	0.18	0.10	-0.08	0.00	0.00%
10	0.22	0.21	1.16	0.00	0.20	0.18	0.10	-0.08	0.00	0.00%
9	0.22	0.20	1.07	0.00	0.00	0.00	0.00	0.00	0.08	4.60%
8	0.22	0.20	1.07	0.00	0.00	0.00	0.00	0.00	0.08	4.60%
7	0.22	0.20	1.07	0.00	0.00	0.00	0.00	0.00	0.08	4.60%
6	0.22	0.20	1.07	0.00	0.00	0.00	0.00	0.00	0.08	4.60%
5	0.24	0.22	1.22	0.25	0.40	0.31	0.33	0.02	0.10	5.64%
4	0.24	0.22	1.20	0.25	0.40	0.27	0.33	0.06	0.14	7.88%
3	0.25	0.22	1.22	0.38	0.40	0.31	0.39	0.08	0.17	9.07%
2	0.25	0.22	1.20	0.38	0.40	0.27	0.39	0.12	0.21	11.32%
1	0.26	0.21	1.14	0.50	0.20	0.14	0.35	0.21	0.29	15.99%

say from 18 to 8, higher reliability scores are observed. In practice, the production process involves long lead times and mandates the supplier to take (or even execute) planning decisions before a single lag period. Considering a month's cut-off time for making decisions, we observed that the fifth and sixth lag periods have higher reliability of forecasts for the given dataset, as shown in Figure 4.

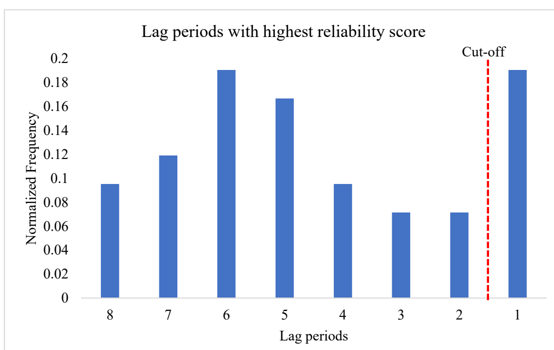


Figure 4: Frequency of lag periods with highest reliability score.

Based on our findings, we infer that the double exponential smoothing algorithm is an effective and practical approach to generate the forecast reliability metric for each lag period of every unique product. The incremental approach further helps in the identification of optimal α and γ values. Finally, we combine the RMSE values with the optimal smoothing constants to generate the forecast reliability metric. Our results indicate that the supplier can use the final forecast reliability scores as a quantitative measure to support their decisions when evaluating the reliability of new forecasts from the end-customer.

4.2 Forecasting Order Quantities

4.2.1 Implementation

In this part of the study, the historical actual data is used to forecast quantities for the upcoming months. A look-back period of twelve months is considered reasonable based on the supplier's feedback, and the lag period columns are generated after further pre-processing. Finally, a min-max scalar transformation is performed to normalize the data, and the dataset is split in an 80:20 ratio across training and test sets.

The considered models, ARIMA, Prophet, SVR and LSTM, are trained and tested for accuracy by determining MAE and MSE values for each product and product family. Given the non-stationary attribute of our data, the ARIMA model was de-prioritized relative to other techniques (Huber and Stuckenschmidt 2020). Figure 5 presents the order quantities for one product over time. The non-stationary behavior of the data is exhibited, along with an increasing trend seen over several months.

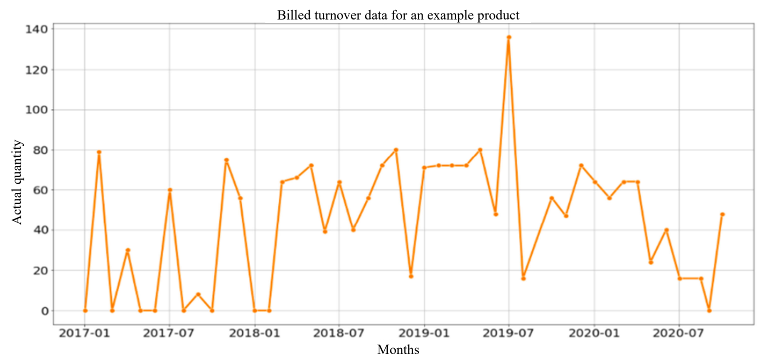


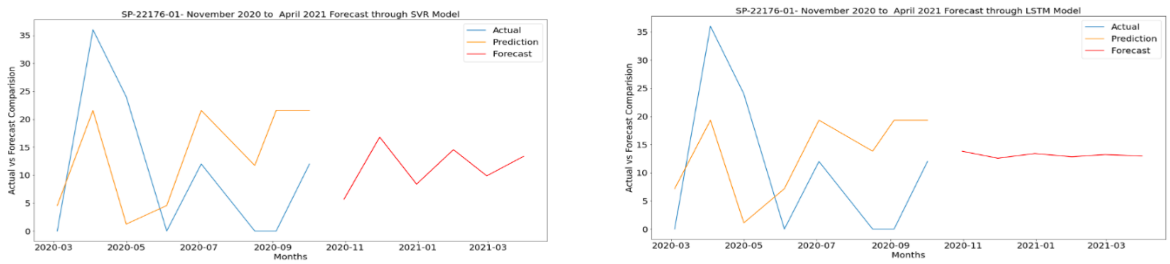
Figure 5: Variation of order quantities of a random product exhibiting non-stationary attribute of the dataset.

The other time series forecasting techniques, Prophet, LSTM, and SVR, demonstrated promising results with low MSE and MAE values. However, this gain in accuracy is offset by the complexity of the model and the ensuing computation cost. The Prophet model was observed to perform well for forecasting an individual product rather than multiple products concurrently. This is anticipated, given that the dataset is multi-variate, with each product varying over time. The Prophet model is primarily designed to work with uni-variate data (Taylor and Letham 2018). A regression model like SVR and the deep-learning-based LSTM model are effective for multi-variate analysis. A radial basis function (rbf) is used as the kernel for the SVR model. After performing a hyperparameter search, the rbf parameters are determined to be $C = 10$ and $\varepsilon = 0.001$. The performance of these two models is then compared, and an inference is made on their multi-variate time series forecast.

4.2.2 Discussion

To compare the performance and accuracy of SVR and LSTM models, their respective MAE and MSE values for every product and product family are determined. It is observed that the SVR model captures the underlying fluctuations present in the actual data in detail and generates an accurate forecast (of the test dataset) that is comparable to the LSTM model. The SVR model algorithm uses a maximal margin principle for regression, similar to that used for classification by Support Vector Machines (SVM). While its accuracy is comparable to the LSTM model, it has a much lower computation time. The training runtime of the SVR model, referring to the total time taken for fitting the training data, is only a few minutes, significantly shorter than the LSTM model, which necessitates 200 training epochs for each unique product.

Figures 6(a) and 6(b) display the forecast for a randomly selected product, as predicted by the SVR and LSTM models, respectively. Similar plots are generated for all products and product families, along with their average MAE and MSE values for each model. Table 2 gives a summary of the performance metrics of the two models. We can infer that while the two models have comparable accuracy, SVR consistently outperforms the more complex LSTM in terms of computational time.



(a) SVR model (b) LSTM model

Figure 6: Outputs of SVR and LSTM models for a randomly selected product.

Table 2: Summary of SVR and LSTM model performance at product and product family levels.

	Product level		Product family level	
	LSTM	SVR	LSTM	SVR
Average MSE	0.18	0.22	0.12	0.13
Average MAE	0.31	0.32	0.27	0.26
Computation Time(seconds)	1877	11	728	4.12

4.3 Simulation Output

We examined the demand forecasts of three customers, denoted as C_{low}, C_{med}, and C_{high}. These customers have different production needs, with C_{high} requiring the largest production quantities and C_{low} requiring the smallest. We also assign a reliability attribute to each customer, represented by the values L, M, or H. L indicates that the customer frequently changes their order quantities, making it more likely that the actual demand will differ significantly from the forecast. On the other hand, H represents a customer whose forecasts are less variable and more accurate. The attributes of M lie between those of L and H. Note that due to confidentiality purposes, actual production quantities and costs cannot be revealed.

We explore how different customer reliability attributes affect the results and obtain 27 scenarios, as presented in Table 3. We observed that 100 iterations are sufficient to obtain a half-width of around 2-5% of value in all our experiments. All experiments have been run on a computer with Intel Core i7-8750H CPU @ 2.2GHz and 16GB RAM with Windows 10 OS. During the evaluation of Scenarios B and C, the customer reliability attribute (L, M, and H) is latent (hidden from the approach). We calculate the gap percentage based on the average costs of scenarios A, B, and C. Scenarios B and C exhibit a negative gap, indicating cost savings and an improvement over scenario A. As the reliability of forecasts increases, the gap decreases. Transitioning from 'L' to 'H' in $C_{h}igh$ results in an average increase of 15.72%. Similarly, with $C_{m}ed$, the average increase is 4.975%, and with $C_{l}ow$, it is 1.69%. These findings highlight the impact of reliability on the output. When the forecasts for the C_{low} customer become highly reliable, there are no significant cost savings due to their low production quantities. Conversely, when the demand forecast for the high-demand customer becomes highly reliable, the cost savings are relatively higher.

Table 3: Sensitivity analyses on simulated scenarios

Forecast reliability			Gap[%] (vs. Scenario A)		Forecast reliability			Gap[%] (vs. Scenario A)	
C_{high}	C_{med}	C_{low}	Scenario B	Scenario C	C_{high}	C_{med}	C_{low}	Scenario B	Scenario C
L	L	L	-25.23	-22.16	M	M	H	-11.87	-13.42
L	L	M	-23.67	-22.55	M	H	L	-10.65	-11.24
L	L	H	-23.83	-21.29	M	H	M	-10.54	-10.98
L	M	L	-22.13	-22.78	M	H	H	-9.32	-10.15
L	M	M	-21.34	-22.90	H	L	L	-9.95	-9.23
L	M	H	-21.76	-21.45	H	L	M	-7.13	-7.96
L	H	L	-20.88	-21.32	H	L	H	-8.24	-8.19
L	H	M	-19.96	-19.12	H	M	L	-6.32	-7.11
L	H	H	-17.67	-19.25	H	M	M	-5.21	-5.90
M	L	L	-18.23	-19.39	H	M	H	-4.97	-5.19
M	L	M	-17.21	-17.10	H	H	L	-4.15	-4.43
M	L	H	-15.23	-14.89	H	H	M	-3.04	-3.15
M	M	L	-13.12	-14.96	H	H	H	-2.90	-3.20
M	M	M	-12.29	-12.16					

5 CONCLUSION

In this study, we consider the problem of assessing the reliability of forecasts provided by a manufacturer to its supplier. We, first, study the forecast variability by product across time periods to provide the supplier with a forecast reliability score for quantifying the extent of reliability of the forecast provided by the manufacturer, and second, make use of time series analyses and machine learning models to generate forecasts based on the actual data. In combination, both methods aid and assist a supplier in executing a more informed production planning process. While the first method indicates the level of confidence that a supplier can accord to forecasts provided by the manufacturer, the second method proposes production volumes based on historical data. We further validate the proposed approach on a simulation model comprising customers with a latent reliability characteristic. The ensuing sensitivity analyses show an average cost reduction of 14% across all instances. Further research may explore the relevance of incorporating additional internal and external information from data sources such as macroeconomic indicators, market trends, geopolitical events, and supply chain information to improve forecast reliability and the usage of ensemble forecasting techniques.

REFERENCES

Andersson, J., and P. Jonsson. 2018. "Big Data in Spare Parts Supply Chains: The Potential of Using Product-in-Use Data in Aftermarket Demand Planning". *International Journal of Physical Distribution & Logistics Management* 48(5):524–544.

Bueno, A., M. Godinho Filho, and A. G. Frank. 2020. "Smart Production Planning and Control in the Industry 4.0 Context: A Systematic Literature Review". *Computers & Industrial Engineering* 149:106774.

Cachon, G. P., and M. A. Lariviere. 2001. "Contracting to Assure Supply: How to Share Demand Forecasts in a Supply Chain". *Management Science* 47(5):629–646.

Cakanyildirim, M., and R. O. Roundy. 2002. "SeDFAM: Semiconductor Demand Forecast Accuracy Model". *IIE Transactions* 34(5):449–465.

Chase, C. W. 2013. *Demand-Driven Forecasting: A Structured Approach to Forecasting*. New Jersey: John Wiley & Sons.

Chen, Y.-J., and C.-F. Chien. 2018. "An Empirical Study of Demand Forecasting of Non-Volatile Memory for Smart Production of Semiconductor Manufacturing". *International Journal of Production Research* 56(13):4629–4643.

Desai, P. S., and K. Srinivasan. 1995. "Demand Signalling Under Unobservable Effort in Franchising: Linear and Nonlinear Price Contracts". *Management Science* 41(10):1608–1623.

Drucker, H., C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik. 1996. "Support Vector Regression Machines". *Advances in neural information processing systems* 9.

Forslund, H., and P. Jonsson. 2007. "The Impact of Forecast Information Quality on Supply Chain Performance". *International Journal of Operations & Production Management* 27(1):90–107.

Fu, W., and C.-F. Chien. 2019. "UNISON Data-Driven Intermittent Demand Forecast Framework to Empower Supply Chain Resilience and an Empirical Study in Electronics Distribution". *Computers & Industrial Engineering* 135:940–949.

Ganesh, M., S. Raghunathan, and C. Rajendran. 2014. "The Value of Information Sharing in a Multi-product, Multi-level Supply Chain: Impact of Product Substitution, Demand Correlation, and Partial Information Sharing". *Decision Support Systems* 58:79–94.

Gonzalez-Vidal, A., F. Jimenez, and A. F. Gomez-Skarmeta. 2019. "A Methodology for Energy Multivariate Time Series Forecasting in Smart Buildings Based on Feature Selection". *Energy and Buildings* 196:71–82.

Hochreiter, S., and J. Schmidhuber. 1997. "Long Short-Term Memory". *Neural Computation* 9(8):1735–1780.

Huber, J., and H. Stuckenschmidt. 2020. "Daily Retail Demand Forecasting Using Machine Learning with Emphasis on Calendric Special Days". *International Journal of Forecasting* 36(4):1420–1438.

LaViola, J. J. 2003. "Double Exponential Smoothing: An Alternative to Kalman Filter-Based Predictive Tracking". In *Proceedings of the Workshop on Virtual Environments 2003*, EGVE '03, 199–206. New York, USA: Association for Computing Machinery.

Lee, C.-Y., and C.-L. Liang. 2018. "Manufacturer's Printing Forecast, Reprinting Decision, and Contract Design in the Educational Publishing Industry". *Computers & Industrial Engineering* 125:678–687.

Lee, H. L., V. Padmanabhan, and S. Whang. 1997. "Bullwhip Effect in Supply Chains". *Sloan Management Review* 38(3):93–102.

Olhager, J. 2010. "The Role of the Customer Order Decoupling Point in Production and Supply Chain Management". *Computers in Industry* 61(9):863–868.

Özer, Ö., and W. Wei. 2006. "Strategic Commitments for an Optimal Capacity Decision Under Asymmetric Forecast Information". *Management Science* 52(8):1238–1257.

Ren, Z. J., M. A. Cohen, T. H. Ho, and C. Terwiesch. 2010. "Information Sharing in a Long-Term Supply Chain Relationship: The Role of Customer Review Strategy". *Operations Research* 58(1):81–93.

Seiringer, W., K. Altendorfer, and T. Felberbauer. 2021. "Simulation of Stochastic Rolling Horizon Forecast Behavior with Applied Outlier Correction to Increase Forecast Accuracy". In *Proceedings of the 2021 Winter Simulation Conference*, edited by S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, 1–12. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Shamir, N., and H. Shin. 2016. "Public Forecast Information Sharing in a Market with Competing Supply Chains". *Management Science* 62(10):2994–3022.

Steinmann, G., and P. J. de Freitas Filho. 2013. "Using Simulation to Evaluate Call Forecasting Algorithms for Inbound Call Centers". In *Proceedings of the 2013 Winter Simulations Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. R. Hill, and M. E. Kuhl, 1132–1139. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Taylor, S. J., and B. Letham. 2018. "Forecasting at Scale". *The American Statistician* 72(1):37–45.

Tseng, M. M., J. Jiao, and M. E. Merchant. 1996. "Design for Mass Customization". *CIRP Annals* 45(1):153–156.

Willmott, C. J., and K. Matsuura. 2005. "Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance". *Climate Research* 30(1):79–82.

Zeiml, S., U. Seiler, K. Altendorfer, and T. Felberbauer. 2020. "Simulation Evaluation of Automated Forecast Error Correction Based on Mean Percentage Error". In *Proceedings of the 2020 Winter Simulation Conference*, edited by K.-H. G. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, and Z. Zheng, 1572–1583. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

BHOOMICA M. NATARAJA is a Ph.D. candidate in the Department of Industrial Engineering at Eindhoven University of Technology. Her research interests include simulation modeling and optimization techniques, with a focus on multi-echelon inventory systems and warehouse processes. Her email address is b.m.nataraja@tue.nl and her website is bhoomica.com.

TANMAY AGGARWAL is the founder and CEO of Lamda Function Inc (lambdafunction.ai). His research interests include application of AI solutions to accelerate autonomous manufacturing. His email address is tanmay@lambdafunction.ai.

NITISH SINGH is a Ph.D. candidate in the Department of Industrial Engineering at the Eindhoven University of Technology. His research interests include simulation-based optimization, and data-driven operations management, focusing on applications in supply chain planning. His email address is n.singh1@tue.nl and his website is nitishsingh.com.

KOEN HERPS is an industrial engineer at the KMWE Group and a Ph.D. candidate in the Department of Industrial Engineering at the Eindhoven University of Technology. He is responsible for the operational execution of the Advanced Manufacturing Logistics project conducted at the Brainport Industries Campus (BIC). His email address is k.herps@kmwe.com or k.herps@tue.nl.

IVO ADAN is a Professor in the Department of Industrial Engineering at Eindhoven University of Technology. His current research interests include modeling and design of manufacturing and warehousing systems and analysis of multi-dimensional Markov processes and queueing models. His email address is i.adan@tue.nl and his homepage is tue.nl/en/research/researchers/ivo-adan.