

O.R. Applications

# Application of machine learning techniques for supply chain demand forecasting

Real Carbonneau, Kevin Laframboise, Rustam Vahidov \*

*Department of Decision Sciences & MIS, John Molson School of Business, Concordia University, 1455 de Maisonneuve Blvd., W, Montreal, Que., Canada H3G 1M8*

Received 3 August 2004; accepted 1 December 2006  
Available online 29 January 2007

## Abstract

Full collaboration in supply chains is an ideal that the participant firms should try to achieve. However, a number of factors hamper real progress in this direction. Therefore, there is a need for forecasting demand by the participants in the absence of full information about other participants' demand. In this paper we investigate the applicability of advanced machine learning techniques, including neural networks, recurrent neural networks, and support vector machines, to forecasting distorted demand at the end of a supply chain (bullwhip effect). We compare these methods with other, more traditional ones, including naïve forecasting, trend, moving average, and linear regression. We use two data sets for our experiments: one obtained from the simulated supply chain, and another one from actual Canadian Foundries orders. Our findings suggest that while recurrent neural networks and support vector machines show the best performance, their forecasting accuracy was not statistically significantly better than that of the regression model.

© 2007 Elsevier B.V. All rights reserved.

*Keywords:* Supply chain management; Forecasting; Neural networks; Support vector machines; Bullwhip effect

## 1. Introduction

Recently, firms have begun to realize the importance of sharing information and integration across the stakeholders in the supply chain (Zhao et al., 2002). Although such initiatives reduce forecast errors, they are neither ubiquitous nor complete

and forecast errors still abound. Collaborative forecasting and replenishment (CFAR) permits a firm and its supplier-firm to coordinate decisions by exchanging complex decision-support models and strategies, thus facilitating integration of forecasting and production schedules (Raghunathan, 1999). However, in the absence of CFAR, firms are relegated to traditional forecasting and production scheduling. As a result, the firm's demand (e.g., the manufacturer's demand) appears to fluctuate in a random fashion even if the final customer's demand has a predictable pattern. Forecasting the manufacturer's demand under these conditions becomes a challenging task due to a well-known

\* Corresponding author. Tel.: +1 514 848 2424x2974; fax: +1 514 848 2824.

*E-mail addresses:* [rcarbonneau@jmsb.concordia.ca](mailto:rcarbonneau@jmsb.concordia.ca) (R. Carbonneau), [Lafrak@jmsb.concordia.ca](mailto:Lafrak@jmsb.concordia.ca) (K. Laframboise), [rvahidov@jmsb.concordia.ca](mailto:rvahidov@jmsb.concordia.ca) (R. Vahidov).

“bullwhip effect” (Lee et al., 1997a) – a result of information asymmetry.

The objectives of this research are to study the feasibility and perform a comparative analysis of forecasting the distorted demand signals in the extended supply chain using non-linear machine learning techniques. More specifically, the work focuses on forecasting the demand at the upstream end of the supply chain. The main challenge lies in the distortion of the demand signal as it travels through the extended supply chain. Our use of the term extended supply chain reflects both the holistic notion of supply chain as presented by Tan (2001) and the idealistic collaborative relationships suggested by Davis and Spekman (2004).

The value of information sharing across the supply chain is widely recognized as the means of combating demand signal distortion (Lee et al., 1997b). However, there is a gap between the ideal of integrated supply chains and the reality (Gunasekaran, 2004). A number of factors could hinder such long-term stable collaborative efforts. Premkumar (2000) lists some critical issues that must be addressed in order to permit successful supply chain collaboration. These issues include:

- alignment of business interests;
- long-term relationship management;
- reluctance to share information;
- complexity of large-scale supply chain management;
- competence of personnel supporting supply chain management;
- performance measurement and incentive systems to support supply chain management.

In most companies these issues have not yet been addressed in any attempts to enable effective extended supply chain collaboration (Davis and Spekman, 2004). Moreover, in many supply chains there are power regimes and power sub-regimes that can prevent supply chain optimization (Cox et al., 2001; Watson, 2001). “Hence, even if it is technically feasible to integrate systems and share information, organizationally it may not be feasible because it may cause major upheavals in the power structure” (Premkumar, 2000, p. 62). Furthermore, it has been mathematically demonstrated, that while participants in supply chains may gain certain benefits from advance information sharing, it actually tends to increase the bullwhip effect (Thonemann, 2002).

Another complicating factor is the possibility of the introduction of inaccurate information into the system. Inaccurate information too would lead to demand distortion, as was the case reported in a study of the telecom industry demand chain where some partners were double forecasting and ration gaming (Heikkilä, 2002), despite the fact that there was a collaborative system in place and a push for the correct use of this system.

Finally, with the advance of E-business there is an increasing tendency towards more “dynamic” (Vakharia, 2002) and “agile” (Gunasekaran and Ngai, 2004; Yusuf et al., 2004) supply chains. While this trend enables the supply chains to be more flexible and adaptive, it could discourage companies from investing into forming long-term collaborative relationships among each other due to the restrictive nature of such commitments.

The above reasons are likely to impede extended supply chain collaboration, or may cause inaccurate demand forecasts in information sharing systems. In addition, the current realities of businesses today are that most extended supply chains are not collaborative all the way upstream to the manufacturer and beyond, and, in practice, are nothing more than a series of dyadic relationships (Davis and Spekman, 2004). In light of the above considerations, the problem of forecasting distorted demand is of significant importance to businesses, especially those operating towards the upstream end of the extended supply chain.

In our investigation of the feasibility and comparative analysis of machine learning approaches to forecasting manufacturer’s distorted demand, we will use concrete tools, including Neural Networks (NN), Recurrent Neural Networks (RNN), and Support Vector Machines (SVM). The performance of these machine learning methods will be compared against baseline traditional approaches including naïve forecasting, moving average, linear regression, and time series models. We have included two sources of distorted demand data for our analysis. The first source is based on a simulation of the extended supply chain and the second one derives from the estimated value of new orders received by Canadian Foundries (StatsCan, 2003).

The remainder of the paper reviews the background and related work for demand forecasting in the extended supply chain; introduces the machine learning approaches included in our analysis; reviews the data sources; and describes and compares the results of our experiments with different

forecasting methods. The paper concludes with the discussion of findings and directions for future research.

**2. Background**

One of the major purposes of supply chain collaboration is to improve the accuracy of forecasts (Raghunathan, 1999). However, since, as discussed above, it is not always possible to have the members of a supply chain work in full collaboration as a team, it is important to study the feasibility of forecasting the distorted demand signal in the extended supply chain in the absence of information from other partners. Therefore, although minimizing the entire extended supply chain’s costs is not the primary focus of this research, we believe that improved quality of forecasts will ultimately lead to overall cost savings. The use of simulation techniques has shown that genetic algorithm-based artificial agents can achieve lower costs than human players. They even minimize costs lower than the “1–1” policy without explicit information sharing (Kimbrough et al., 2001). Analysis of forecasting techniques is of considerable value for firms, as it has been shown that the use of moving average, naïve forecasting or demand signal processing will induce the bullwhip effect (Dejonckheere et al.,

2003). Autoregressive linear forecasting, on the other hand, has been shown to diminish bullwhip effects, while outperforming naïve and exponential smoothing methods (Chandra and Grabis, 2005). In this paper, we will analyze the applicability of machine learning techniques to demand forecasting in supply chains.

The primary focus of this work is on facilitating demand forecasting by the members at the upstream end of a supply chain. The source of the demand distortion in the extended supply chain simulation is demand signal processing by all members in the supply chain (Forrester, 1961). According to Lee et al. (1997b), demand signal processing means that each party in the supply chain does some processing on the demand signal, transforming it, before passing it along to the next member. As the end-customer’s demand signal moves up the supply chain, it is increasingly distorted because of demand signal processing. This occurs even if the demand signal processing function is identical in all parties of the extend supply chain. The phenomenon could be explained in terms of chaos theory, where a small variation in the input could result in large, seemingly random, behavior in the output of the chaotic system (Kullback, 1968).

The top portion of Fig. 1 depicts a model of the extended supply chain with increasing demand dis-

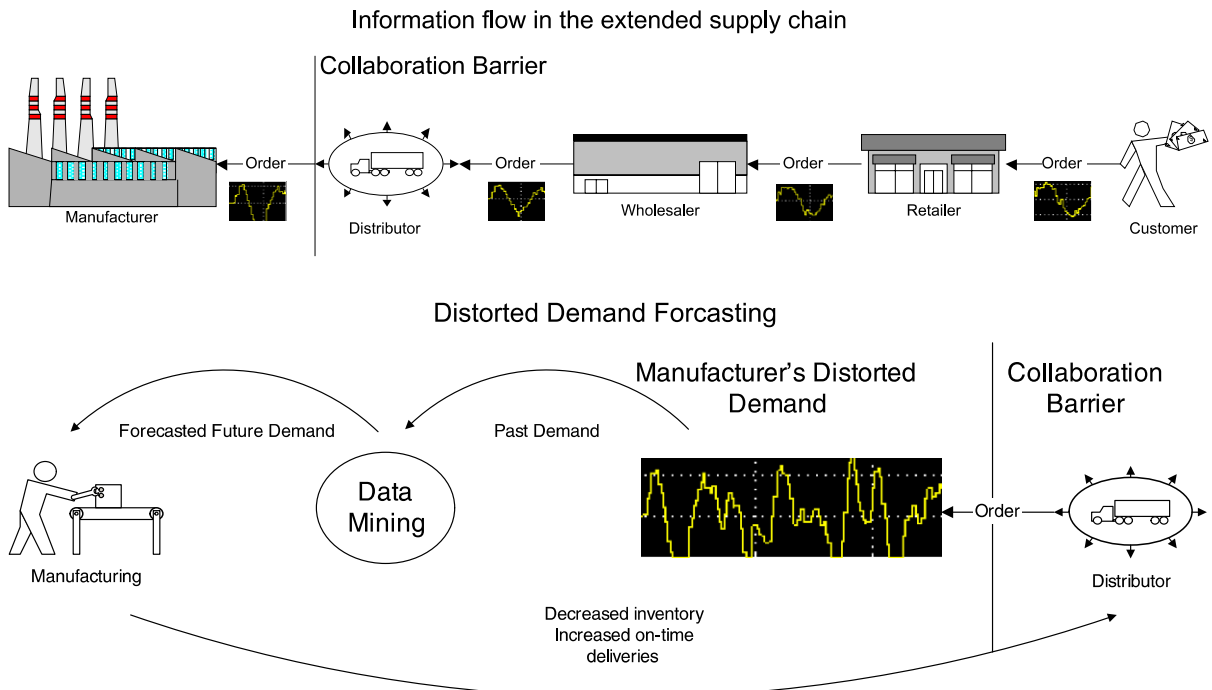


Fig. 1. Distorted demand signal in an expanded supply chain.

tortion and which includes a collaboration barrier. The latter could be defined as the link in the supply chain at which no explicit forecast information-sharing occurs between the partners. Thus, our objective is to forecast future demand (far enough in advance to compensate for the manufacturer's lead time) based only on past manufacturer's orders. To this end, we will investigate the utility of advanced machine learning techniques. Consequently, we suggest that if an increase in forecasting accuracy can be achieved, it will result in lower costs because of reduced inventory as well as increased customer satisfaction that will result from an increase in on-time deliveries. In the bottom portion of Fig. 1, we depict the effect of distorted demand forecasting.

Basic time series analysis (Box, 1970) will be used in this research as one of the "traditional" methods against which the performance of other advanced techniques will be compared. The latter include Neural Networks, Recurrent Neural Networks, and Support Vector Machines. Neural Networks (NN) and Recurrent Neural Networks (RNN) are frequently used to predict time series (Dorffner, 1996; Herbrich et al., 2000; Landt, 1997; Lawrence et al., 1996). In particular, RNN are included in the analysis because the manufacturer's demand is considered a chaotic time series. RNN perform back-propagation of error through time that permits learning patterns through an arbitrary depth in the time series. This means that even though we provide a time window of data as the input dimension to the RNN, it can match pattern through time that extends further than the provided current time window because it has recurrent connections. Support Vector Machines (SVM), a more recent learning algorithm that has been developed from statistical learning theory (Vapnik, 1995; Vapnik et al., 1997), has a very strong mathematical foundation, and has been previously applied to time series analysis (Mukherjee et al., 1997; Rüping and Morik, 2003).

### 3. Supply chain demand forecasting techniques

The forecasting techniques used in our analysis include:

- Naïve Forecast
- Average
- Moving Average
- Trend
- Multiple Linear Regression

- Neural Networks
- Recurrent Neural Networks
- Support Vector Machines

The naïve forecast is one of the simplest forecasting methods and is often used as a baseline method against which the performance of other methods is compared. It simply uses the latest value of the variable of interest as a best guess for the future value. The moving average forecast uses the average of a defined number of previous periods as the future forecasted demand. Trend-based forecasting is based on a simple regression model that takes time as an independent variable and tries to forecast demand as a function of time. The multiple linear regression model tries to predict the change in demand using a number of past changes in demand observations as independent variables. Thus, it is an autoregressive model.

We consider these first five of the techniques as "traditional", and the remaining as more "advanced". We expect that the advanced methods will outperform the more traditional ones, since:

- The advanced methods incorporate non-linear models, and as such could serve as better approximations than those based on linear models;
- We expect to have a significant level of non-linearity in demand behavior as it exhibits complex behavior.

In the remainder of the section, we will briefly overview the advanced forecasting techniques.

#### 3.1. Neural networks

Although artificial neural networks could include a wide variety of types, here we refer to most commonly used *feed-forward error back-propagation type neural nets*. In these networks, the individual elements ("neurons") are organized into layers in such a way that output signals from the neurons of a given layer are passed to all of the neurons of the next layer. Thus, the flow of neural activations goes in one direction only, layer-by-layer. The smallest number of layers is two, namely the input and output layers. More layers, called hidden layers, could be added between the input and the output layer. The function of the hidden layers is to increase the computational power of the neural nets. Provided with sufficient number of hidden units, a

neural network could act as a “universal approximator”.

Neural networks are tuned to fulfill a required mapping of inputs to the outputs using training algorithms. The common training algorithm for the feed-forward nets is called “error back-propagation” (Rumelhart et al., 1986). This is a supervised type of training, where the desired outputs are provided to the NN during the course of training along with the inputs. The provided input–output couplings are known as training pairs, and the set of given training pairs is called the training set. Thus, neural networks can learn and generalize from the provided set of past data about the patterns in the problem domain of interest.

### 3.2. Recurrent neural networks

Recurrent neural networks allow output signals of some of their neurons to flow back and serve as inputs for the neurons of the same layer or those of the previous layers. RNN serve as a powerful tool for many complex problems, in particular when time series data is involved. The training method called “back-propagation through time” could be applied to train a RNN on a given training set (Werbos, 1990). Fig. 2 shows schematically the structure of RNN for the supply chain demand-forecasting problem. As may be noted in Fig. 2, only the output of the hidden neurons is used to serve as the input to the neurons of the same layer.

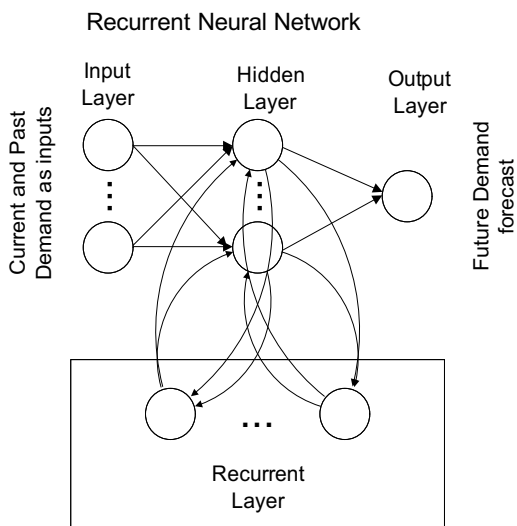


Fig. 2. Recurrent neural network for demand forecasting (MathWorks, 2000).

### 3.3. Support vector machines

Support vector machines (SVMs) are a newer type of universal function approximators that are based on the structural risk minimization principle from statistical learning theory (Vapnik, 1995) as opposed to the empirical risk minimization principle on which neural networks and linear regression, to name a few, are based. The objective of structural risk minimization is to reduce the true error on an unseen and randomly selected test example as opposed to NN and MLR, which minimize the error for the currently seen examples. Support vector machines project the data into a higher dimensional space and maximize the margins between classes or minimize the error margin for regression. Margins are “soft”, meaning that a solution can be found even if there are contradicting examples in the training set. The problem is formulated as a convex optimization with no local minima, thus providing a unique solution as opposed to back-propagation neural networks, which may have multiple local minima and, thus cannot guarantee that the global minimum error will be achieved. A complexity parameter permits the adjustment of the number of error versus the model complexity, and different kernels, such as the Radial Basis Function (RBF) kernel, can be used to permit non-linear mapping into the higher dimensional space.

## 4. Data

In order to examine the effectiveness of various advanced machine learning techniques in forecasting supply chain demand, two data sets were prepared. The first one represents results of the simulations of an extended supply chain, and the second one contains actual Foundries data provided by Statistics Canada.

### 4.1. Supply chain simulation data

Our simulations included four parties that make up the supply chain that delivers a product to a final customer. Based on this extended supply chain model, in accordance with our earlier discussion, the simulated demand signal processing is introduced as the source of demand distortion. In essence, demand signal processing is modeled by a simple linear regression that calculates the trend over the past 10 days, and, which is then used to forecast the demand in 2 days. This results in



demand signal processing sufficient to cause significant distortion at the end of the extended supply chain. In essence, the more “nodes” the demand signal has to travel through, the more chaotic its behavior becomes. For the distorted demand signal, that we know has been generated using a well-defined pattern, we can attempt forecasting using various machine learning techniques.

The extended supply chain simulation is developed in MATLAB Simulink (MathWorks, 2000). The macro structure of the supply chain defines the one-day delay for the order to move to the next party, the two-day delay for the goods to be delivered, as well as all the display points required for monitoring the simulation and the data collection points that collect data to be used by the forecasting methods. The one-day delay for the orders models the communication lag between the two parties since the customer must first determine the purchasing requirements, create the purchase order and have it approved and then place the order with the vendor which then would submit it for further processing. The two-day delay for the goods to be received into inventory by the customer models a delivery lag because of the processing time of the goods issue, the delivery time and then subsequent processing time for goods receipt. In the very last step of the supply chain, these delays are almost identical but are for the internal production process, such as processing the production order and manufacturing the good.

The end-customer demand is generated (see Fig. 3) using a sine wave pattern that varies between 800 and 1200 final product units representing the effects of seasonality. The pattern repeats itself approximately every month. White noise is superim-

posed on this pattern to simulate randomness in customer demand. The noise is normally distributed random numbers with a noise power of 1000 which in practice gives a variation of approximately plus or minus 100 every 1000 periods.

Each partner in the extended supply chain is modeled as having the same structure and behavior. Although this is arguably a simplification of reality, it nevertheless results in demand distortion. A partner is divided into the following functions; demand forecasting, purchase-order calculation, and order processing that includes goods receipt and goods delivery. The mechanism for demand signal processing is represented diagrammatically in Fig. 4. As can be noted, the demand signal, after having been processed by the demand forecasting function, feeds into the purchase-order calculation function. The Delivery, Inventory and Backlog signals are all part of the simulated partner model in order to enable simulation monitoring and data extraction. In the simulation schemas, the ovals containing a number indicate the modules interfaces, the inputs and outputs are numbered starting at 1 and they are also named.

Demand forecasting is the business function that is the source of the demand signal distortion. In our simulation, it is based on a simple linear regression of the past 10 days of demand. The forecast is used for the estimation of demand in three following days. This permits the ordering today of the estimated quantity required in three days. The combined effect of small random variations in end-customer demand and demand signal processing distorts the initial demand, which initially has a simple pattern but the demand changes into a chaotic pattern once it reaches the manufacturer.

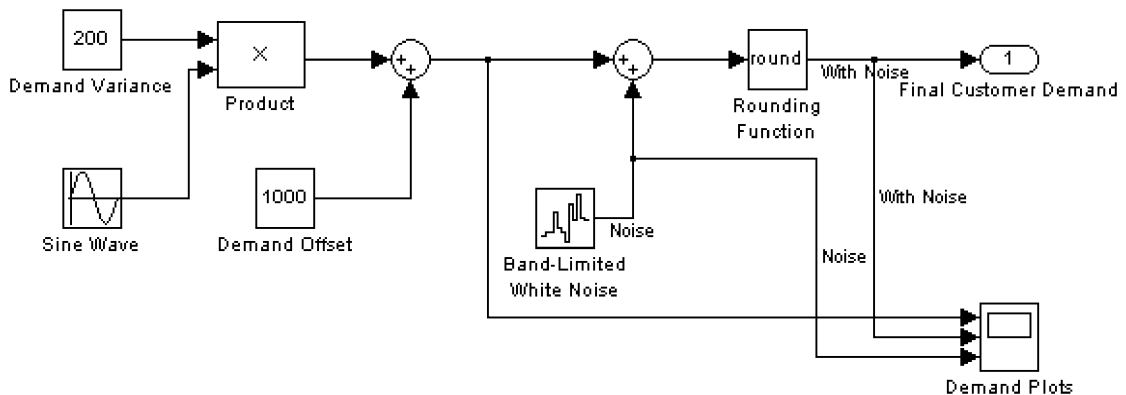


Fig. 3. End-customer demand simulation.

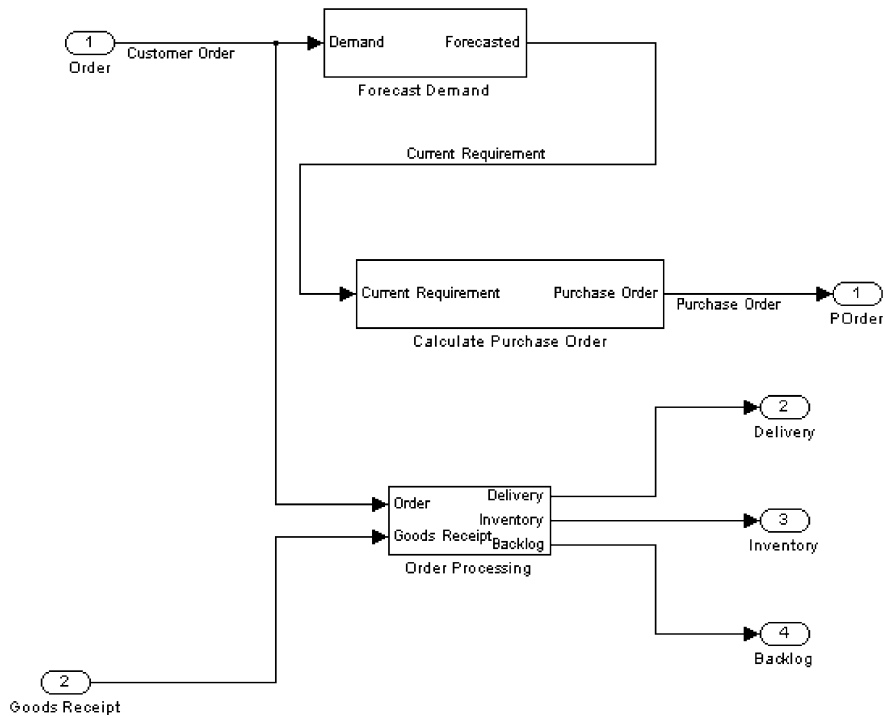


Fig. 4. Model of a partner.

Once the demand signal has been processed by the demand forecasting function, it is then used to calculate the purchase order that will be submitted to the next partner. We assume that the ordering party keeps track of the current total non-received quantities so that it does not keep ordering more for the items that are in the backlog. Deliveries received from the supplier, plus the quantity that is currently in inventory, are used to fulfill any backlog order. Only then do regular orders and the remaining quantity replenish the inventory.

An initial simulation run of 200 days is used to review the results of the extended supply chain. Fig. 5 shows the order history for all partners. As can be observed, the simple demand signal processing can cause significant distortions in the manufacturer’s orders and production orders. This distorted demand will eventually have adverse effects on deliveries, inventories and backlogs for the retailer (Fig. 6).

#### 4.2. Foundries monthly sales data

The foundries monthly sales data was obtained from the Statistics Canada table 0304-0014, which is defined as “Manufacturers’ shipments, invento-

ries, orders and inventory to shipment ratios, by North American Industry Classification System (NAICS), Canada, monthly” (StatsCan, 2003). From this table, the monthly sales for all foundries are used. The classification of foundries is as Durable Goods Industries, Primary Metal Manufacturing. The industry is described as follows: “this Canadian industry comprises establishments primarily engaged in pouring molten steel into investment moulds or full moulds to manufacture steel castings.” (StatsCan, 2003). Therefore, it is far upstream from the final customer and is expected to be subject to much demand signal distortion. The value (in Canadian dollars) of the monthly orders by the Foundries is shown in Fig. 7.

### 5. Experimental results

This section describes the results of experiments using various forecasting techniques on the two data sets described in the previous section.

#### 5.1. Data preparation

The exact same training and testing data sets have been used for all of the forecasting models

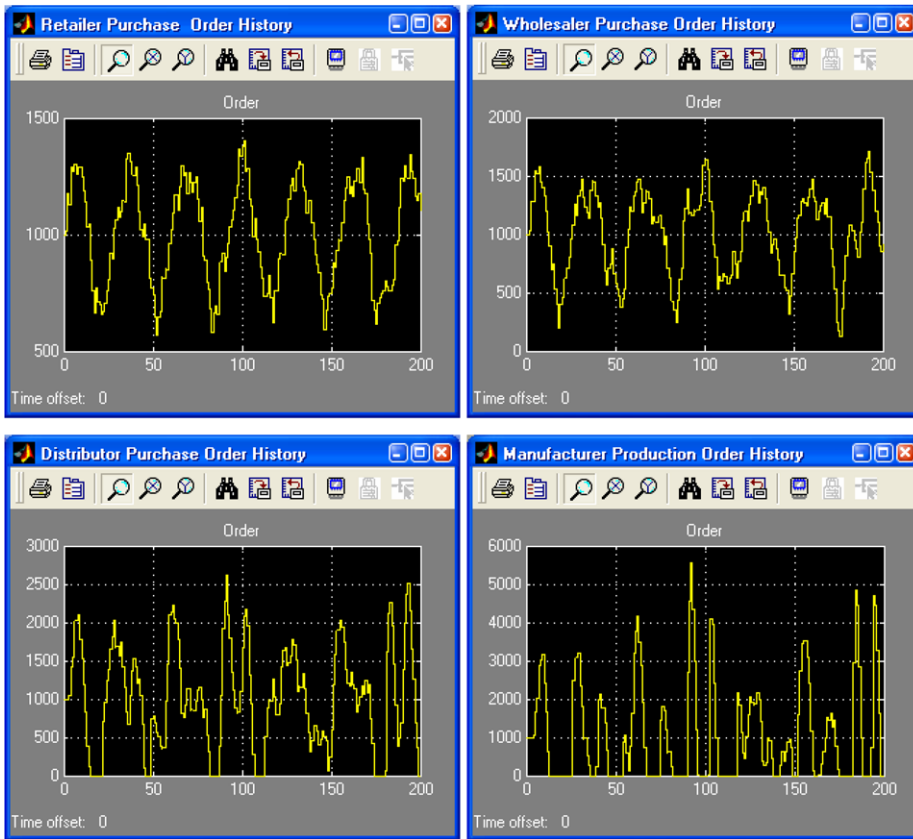


Fig. 5. Progressive distortion of demand due to demand signal processing.

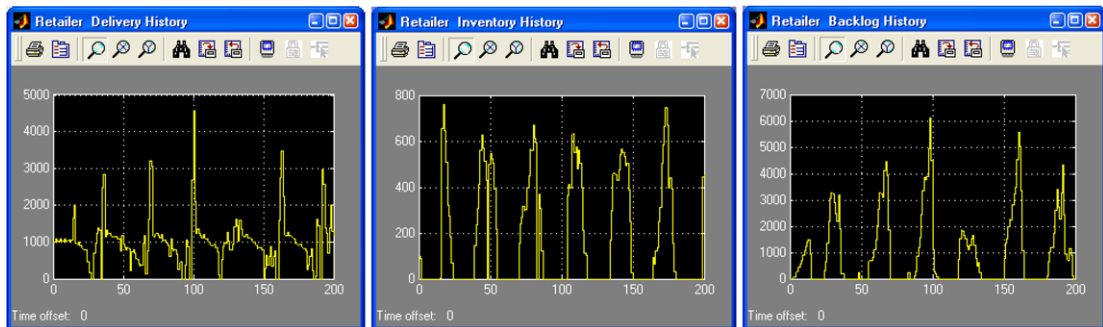


Fig. 6. Retailer delivery history, inventory history, and backlog history.

for valid comparisons. The daily manufacturer orders, obtained from our supply chain simulation, are first used as a source of data for building forecasting models. The first 300 days of this data can be seen in Fig. 8.

The input variables for the models are the change in demand for each of the past four periods plus that for the current period, and the output variable is the total change over the next three periods. The

exceptions are the Naïve, Average and Trend forecast. Since this data pre-processing creates null values at the beginning and end of the prepared data set, these observations are dropped from consideration. We separate the simulation demand into two sets: training set and testing set. The training set has 600 days and the testing set has 600 days. A sample of the first 15 observations and the prepared data set are given in Table 1.



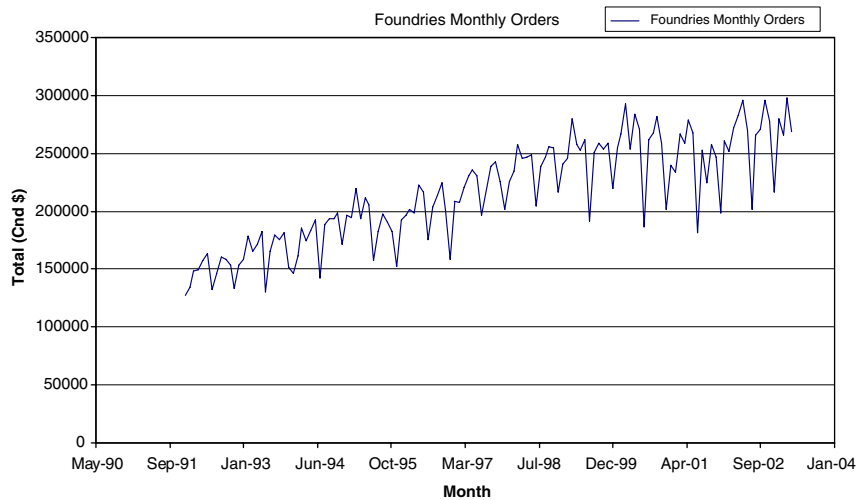


Fig. 7. Foundries monthly orders.

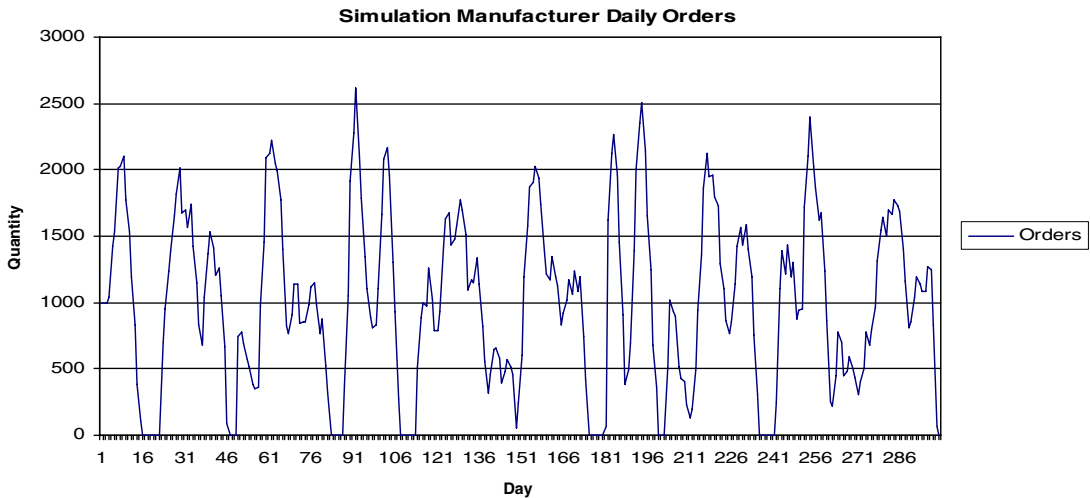


Fig. 8. Data obtained from simulation.

There are 136 months of estimated sales data for Canadian Foundries. The observations start on January 1992 and end on April 2003. Because of the data preparation, which causes null values, five observations from the beginning of the data set and one observation from the end have been dropped. This gives a total of 130 months of observations.

Input variables for all the models is the percentage change in demand for each of the past four months plus this month and the output variable is the percentage change over the next month. Because of the increasing trend in total demand and increasing demand oscillation amplitude, percentage

change is preferable to just the change. Table 2 shows a sample of data set for foundries.

The training set for foundries data includes 100 months (77%) and the testing set has 30 months (23%).

### 5.2. Neural networks

A three-layer feed-forward back-propagation neural network was used to build a forecasting model (Demuth and Beale, 1998). A neural network with a hyperbolic tangent (tanh) transfer function, a learning rate of 0.1 and a momentum of 0.7 was trained to capture relationships between the five

Table 1  
Sample of observations for simulated data forecasting

Day	Demand	Change over the next 3 days	Lag 4 days	Lag 3 days	Lag 2 days	Lag 1 day	Today
1	1000	37.8					
2	1000	426.7					0
3	1000	537.1				0	0
4	1037.8	979.3			0	0	37.8
5	1426.7	595.2		0	0	37.8	388.9
6	1537.1	567.4	0	0	37.8	388.9	110.4
7	2017.1	-244.4	0	37.8	388.9	110.4	480
8	2021.9	-489.8	37.8	388.9	110.4	480	4.8
9	2104.5	-909.8	388.9	110.4	480	4.8	82.6
10	1772.7	-937.8	110.4	480	4.8	82.6	-331.8
11	1532.1	-1145.77	480	4.8	82.6	-331.8	-240.6
12	1194.7	-1084.94	4.8	82.6	-331.8	-240.6	-337.4
13	834.9	-834.9	82.6	-331.8	-240.6	-337.4	-359.8
14	386.33	-386.33	-331.8	-240.6	-337.4	-359.8	-448.57
15	109.76	-109.76	-240.6	-337.4	-359.8	-448.57	-276.57

Table 2  
Sample of foundries data set

Month	Demand	Relative change over next 1 month	Lag 4 months	Lag 3 months	Lag 2 months	Lag 1 month	This month
Jan-92	127712	0.055124					
Feb-92	134752	0.104154					0.055124
Mar-92	148787	0.002971				0.055124	0.104154
Apr-92	149229	0.054661			0.055124	0.104154	0.002971
May-92	157386	0.037024		0.055124	0.104154	0.002971	0.054661
Jun-92	163213	-0.188392	0.055124	0.104154	0.002971	0.054661	0.037024
Jul-92	132465	0.108678	0.104154	0.002971	0.054661	0.037024	-0.18839
Aug-92	146861	0.094797	0.002971	0.054661	0.037024	-0.18839	0.108678
Sep-92	160783	-0.016289	0.054661	0.037024	-0.18839	0.108678	0.094797
Oct-92	158164	-0.027579	0.037024	-0.18839	0.108678	0.094797	-0.01629
Nov-92	153802	-0.134933	-0.18839	0.108678	0.094797	-0.01629	-0.02758
Dec-92	133049	0.150929	0.108678	0.094797	-0.01629	-0.02758	-0.13493
Jan-93	153130	0.033063	0.094797	-0.01629	-0.02758	-0.13493	0.150929
Feb-93	158193	0.128356	-0.01629	-0.02758	-0.13493	0.150929	0.033063
Mar-93	178498	-0.072415	-0.02758	-0.13493	0.150929	0.033063	0.128356

inputs and one output for both data sets. We use a 20% cross-validation set (a subset of the training set) to stop the training once the error on this set starts increasing to prevent over-trained.

For the simulation data, 10 hidden layer neurons were used. This choice of 10 hidden layer neurons give a ratio of 8 training samples to one neural network weight:

$$\frac{S \cdot T}{H(I + O)} = \frac{600 \cdot 0.8}{10(5 + 1)} = 8.$$

Here  $S$  is the sample size,  $T$  is the proportion of the sample used for training, and  $I$ ,  $O$ , and  $H$  are the number of input, output and hidden neurons respectively.

For the foundries demand forecasting, the number of neurons in the hidden layer was set to 3 with a ratio of samples to weights of 5.7. A somewhat smaller ratio is due to the smaller size of the available training set. Figs. 9 and 10 show graphically the forecasted change in the demand for both of the data (simulated and foundries) using neural networks on the testing data sets.

### 5.3. Recurrent neural networks

Our recurrent neural network model was similar to the neural network model described above, but it additionally had recurrent connections for every

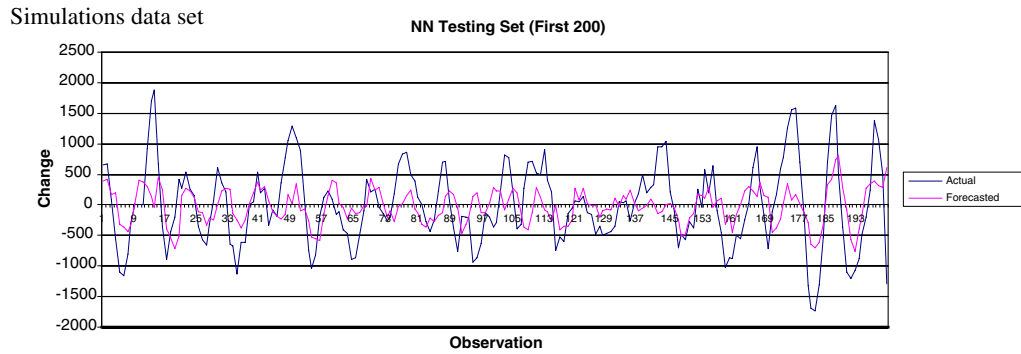


Fig. 9. Simulations testing data set results.

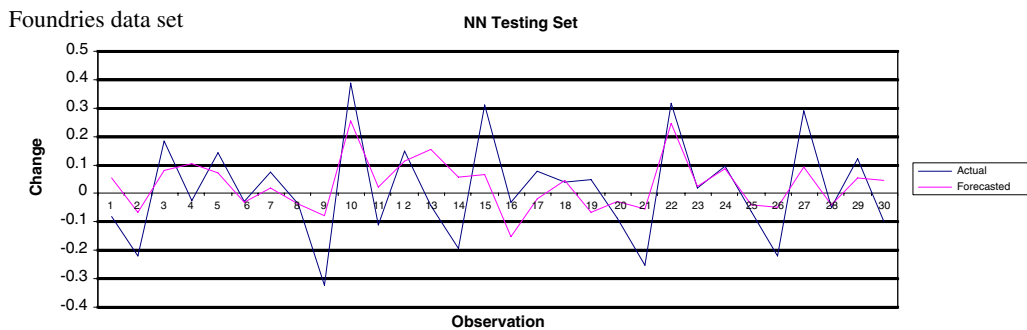


Fig. 10. Foundries testing data set results.

neuron in the hidden layer that fed back into all the neurons of the same layer at the next step. This allowed the RNN to learn patterns through time. Both RNNs were based on a hyperbolic tangent transfer function, the learning rate was 0.01 and the momentum was 0.7. For the RNN trained on the simulated data six neurons in the hidden layer were used. With the 62 recurrent connections, the final ratio of samples to network weights was 6.7. The RNN for foundries demand forecasting had two neurons in the hidden layer to achieve the ratio of samples to weights of 6.5. Figs. 11 and 12 show the performance of RNNs on the two testing sets.

#### 5.4. Support Vector Machine

A Least Squares Support Vector Machine with a Radial Basis Function kernel was used as our next forecasting method. The specific tool used is the LS-SVMLab1.5 for MATLAB (Pelckmans et al., 2002; Suykens et al., 2002). For choosing the best RBF kernel parameters, LS-SVM's 10-fold cross-validation based hyper parameter selection function

was used. Once the best parameters had been selected, then the final LS-SVM model has been trained. For Simulation Distorted Demand Forecasting the hyper parameter optimization function decided on a regularization parameter of 0.6206 and a RBF kernel parameter of 26.3740. For the foundries demand forecasting the hyper parameter optimization function decided on a regularization parameter of 16.3970 and a RBF kernel parameter of 2.1510. The LS-SVM could learn the training set extremely well. Figs. 13 and 14 show the performance of SVM on the testing data sets.

#### 5.5. Comparison of results

The comparison of results of the various forecasts is presented in Tables 3 and 4. They have been sorted in ascending order based on the test set Mean Average Error (MAE). The Recurrent Neural Network and the Least-Squares Support Vector Machine have the best results, but their superior accuracy is relatively larger for the Foundries' demand data series than the simulated demand data

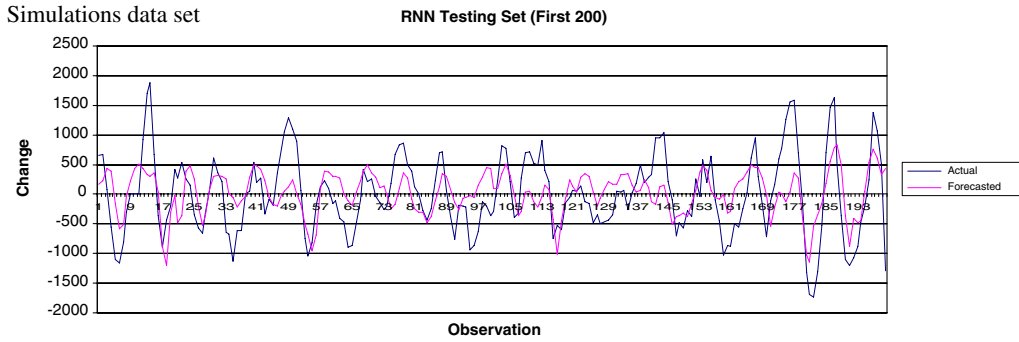


Fig. 11. Simulations testing data set results.

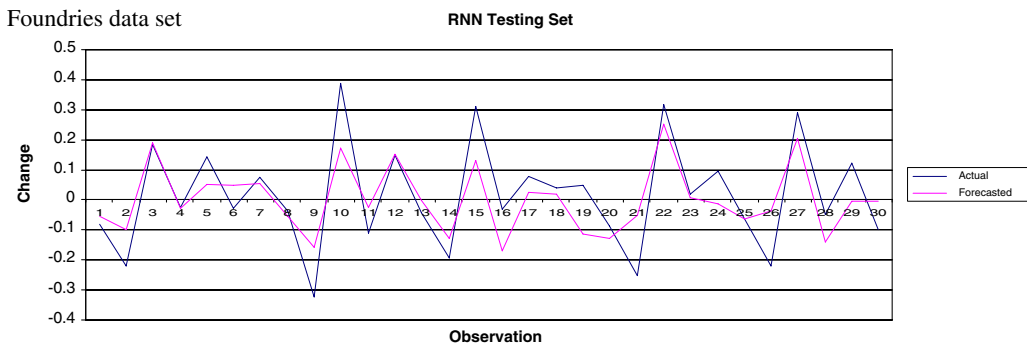


Fig. 12. Foundries testing data set results.

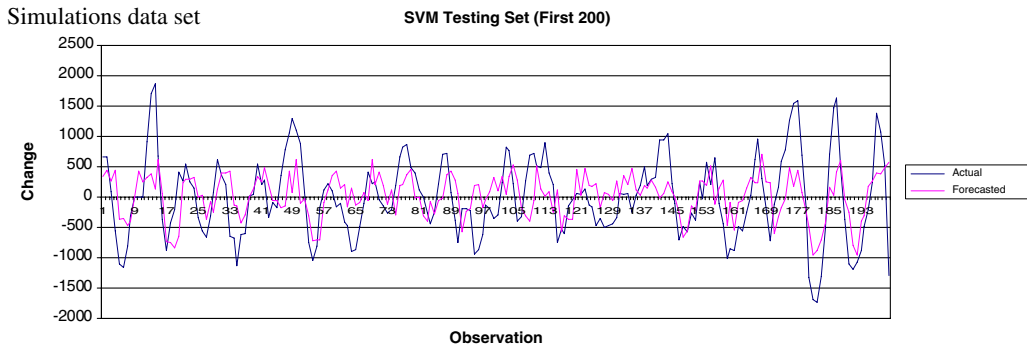


Fig. 13. Simulation testing data set results.

series. In both cases the SVM has the best training set accuracy with the least impact on its generalization ability seen in the testing set. We can also see that the trend estimation and the naïve forecast are the worst types of forecasting since they have the highest level of error.

We have used *t*-tests in order to compare the accuracy of different forecasting techniques. Tables 5 and 6 show the resulting *p*-values.

The results of the experiments suggest that the recurrent neural networks and support vector machines have been the most accurate forecasting techniques. Moving average, naïve and trend forecasting have been among the worst performers. However, statistical analysis of the results show that there was no statistically significant differences in terms of the accuracy of forecasts among RNN, SVM, NN, and MLR. Thus, the gains in the

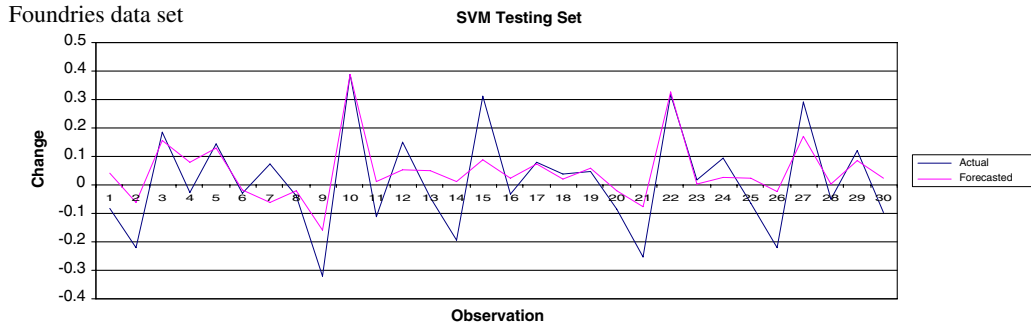


Fig. 14. Foundries testing data set results.

Table 3  
Comparison of the performance (MAE) of forecasting techniques on the simulation data set

Forecasting technique	Testing set		Training set	
	MAE	Std. dev.	MAE	Std. dev.
RNN	447.72	328.23	461.66	350.35
LS-SVM	453.04	341.88	449.32	365.01
MLR	453.22	343.65	464.62	375.62
NN	455.41	354.40	471.03	383.29
Naïve	520.53	407.29	536.45	435.05
Moving average	526.61	370.35	558.82	400.11
Trend	618.02	487.42	674.68	490.50

Table 4  
Comparison of the performance of forecasting techniques on the foundries data set

Forecasting technique	Testing set		Training set	
	MAE	Std. dev.	MAE	Std. dev.
RNN	20.352	16.203	15.521	12.334
LS-SVM	20.485	17.304	3.665	3.722
MLR	21.396	19.705	15.007	15.041
NN	25.260	19.733	12.855	12.057
Moving average	25.481	19.253	18.205	13.028
Trend	27.323	24.198	17.995	17.292
Naïve	32.591	23.485	20.263	17.380

Table 5  
Simulation dataset: Probability that the two samples come from the same population

	RNN	LS-SVM	MLR	NN	Moving average	Trend
RNN						
LS-SVM	0.2947					
MLR	0.2862	0.4824				
NN	0.2069	0.3354	0.2833			
Moving average	0.0000	0.0000	0.0000	0.0000		
Trend	0.0000	0.0000	0.0000	0.0000	0.0001	
Naïve	0.0000	0.0000	0.0000	0.0000	0.3451	0.0000

forecast accuracy of RNN and SVM over MLR have been marginal for both datasets. Moreover,

MLR has outperformed neural networks on both datasets. This result can be explained by the prob-

Table 6  
 Foundries dataset: Probability that the two samples come from the same population

	RNN	LS-SVM	NN	MLR	Moving average	Trend
RNN						
LS-SVM	0.4851					
NN	0.0739	0.0330				
MLR	0.3677	0.4089	0.1532			
Moving average	0.0603	0.0835	0.4770	0.0786		
Trend	0.0200	0.0387	0.2889	0.0638	0.3030	
Naïve	0.0010	0.0021	0.0240	0.0065	0.0513	0.0707

lem of overtraining of neural network. Recurrent neural network has shown a better performance presumably due to its ability to capture temporal patterns. SVM has shown the best performance on the training sets that did not extend to the testing set. This shows the limitations of the SVM on achieving the true generalization.

Overall, the performance of RNN, SVM, NN, and MLR have been significantly better than that of the simpler techniques including moving average, naïve, and trend methods.

## 6. Conclusion

The objective of this research was to study the effectiveness of forecasting the distorted demand signals in the extended supply chain with advanced non-linear machine learning techniques.

The results are important for situations where parties in the supply chain cannot collaborate for the reasons discussed in the beginning of the paper. In such cases, the ability to increase forecasting accuracy will result in lower costs and higher customer satisfaction because of more on-time deliveries.

Although showing better results overall, the advanced techniques did not provide a large improvement over more “traditional techniques” (as represented by the MLR model) for the simulation data set. However, for the real foundries data, the more advanced data mining techniques (RNN and SVM) provide larger improvements. Recurrent Neural Networks (RNN) and Support Vector Machines (SVM) provide the best results on the foundries test set. We can also see that the trend estimation and naïve forecast are the worst types of demand signal processing since they have the highest level of error.

Overall, we can conclude that the use of machine learning techniques and MLR for forecasting distorted demand signals in the extended supply chain

provide more accurate forecasts than simpler forecasting techniques (including naïve, trend, and moving average). However, we did not find that machine learning techniques show significantly better performance than linear regression. Thus, the marginal gain in accuracy of the RNN and SVM models should be weighed in practice against the conceptual and computational simplicity of the linear regression model.

Future research could be directed to investigating the impacts of information sharing on forecasting accuracy, e.g., using the Internet, and other e-business technologies, as a means of enabling firms to coordinate decisions with their various partners (Vakharia, 2002). To study the impact of collaborative forecasting, the current simulations, models and techniques can be used while providing additional information from down the supply chain to the data mining tools. This additional information would likely increase the model’s accuracy. However, as demonstrated by Frohlich (2002), as long as barriers remain, integration of decision making will continue to be a constraint. Such constraints might be alleviated with our model.

## References

- Box, G.E., 1970. *Time Series Analysis*. Holden-Day, San Francisco.
- Chandra, C., Grabis, J., 2005. Application of multi-steps forecasting for restraining the bullwhip effect and improving inventory performance under autoregressive demand. *European Journal of Operational Research* 166 (2), 337–350.
- Cox, A., Sanderson, J., Watson, G., 2001. Supply chains and power regimes: Toward an analytic framework for managing extended networks of buyer and supplier relationships. *Journal of Supply Chain Management* 37 (2), 28–35.
- Davis, E.W., Spekman, R., 2004. *Extended Enterprise*. Prentice-Hall, Upper Saddle River, NJ.
- Dejonckheere, J., Disney, S.M., Lambrecht, M.R., Towill, D.R., 2003. Measuring and avoiding the bullwhip effect: A control theoretic approach. *European Journal of Operational Research* 147 (3), 567–590.



- Demuth, H., Beale, M., 1998. In: Natick (Ed.), *Neural Network Toolbox for Use with MATLAB, User's Guide* (version 3.0). The MathWorks, Inc., Massachusetts.
- Dorffner, G., 1996. Neural Networks for Time Series Processing. *Neural Network World* 96 (4), 447–468.
- Forrester, J., 1961. *Industrial Dynamics*. Productivity Press, Cambridge, MA.
- Frohlich, M., 2002. Demand chain management in manufacturing and services: Web-based integration, drivers and performance. *Journal of Operations Management* 20 (6), 729–745.
- Gunasekaran, A., 2004. Supply chain management: Theory and applications. *European Journal of Operational Research* 159 (2), 265–268.
- Gunasekaran, A., Ngai, E.W.T., 2004. Information systems in supply chain integration and management. *European Journal of Operational Research* 159 (2), 265–527.
- Heikkilä, J., 2002. From supply to demand chain management: Efficiency and customer satisfaction. *Journal of Operations Management* 20 (6), 747–767.
- Herbrich, R., Keilbach, M.T., Graepel, P.B.-S., Obermayer, K., 2000. Neural networks in economics: Background, applications and new developments. *Advances in Computational Economics: Computational Techniques for Modeling Learning in Economics* 11, 169–196.
- Kimbrough, S., Wu, D.J., Fang, Z., 2001. Computers play the beer game: Can artificial agents manage supply chains? *Decision Support Systems* 33 (3), 323–333.
- Kullback, S., 1968. *Information Theory and Statistics*, second ed. Dover Books, New York.
- Landt, F.W., 1997. *Stock Price Prediction using Neural Networks*. Laiden University.
- Lawrence, S., Tsoi, A.C., Giles, C.L., 1996. Noisy time series prediction using symbolic representation and recurrent neural network grammatical inference. University of Maryland, Institute for Advanced Computer Studies.
- Lee, H.L., Padbanaban, V., Whang, S., 1997a. The Bullwhip Effect in Supply Chains. *Sloan Management Review* (Spring).
- Lee, H.L., Padbanaban, V., Whang, S., 1997b. Information distortion in a supply chain: The bullwhip effect. *Management Science* 43 (4), 546–558.
- MathWorks, I. 2000. *Using MATLAB, Version 6*, MathWorks, Inc.
- Mukherjee, S., Osuna, E., Girosi, F., 1997. Nonlinear prediction of chaotic time series using support vector machines. Paper presented at the Proceedings of IEEE NNSP.
- Pelckmans, K., Suykens, J.A.K., Van Gestel, T., De Brabanter, J., Lukas, L., Hamers, B., De Moor, B., Vandewalle, J., 2002. LS-SVMLab: A Matlab/C toolbox for Least Squares Support Vector Machines. ESAT-SISTA, K.U. Leuven, Leuven, Belgium.
- Premkumar, G.P., 2000. Interorganization systems and supply chain management: An information processing perspective. *Information Systems Management* 17 (3), 56–69.
- Raghunathan, S., 1999. Interorganizational collaborative forecasting and replenishment systems and supply chain implications. *Decision Sciences* 30 (4).
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning internal representations by error propagation. In: Rumelhart, J.L., McClelland, J.L. (Eds.), *Parallel distributed processing*, vol. 1. MIT Press, Cambridge, MA, pp. 318–362.
- Rüping, S., Morik, K., 2003. *Support Vector Machines and Learning about Time*. Paper presented at the Proceedings of ICASSP 2003.
- StatsCan, 2003. *Statistics Canada Table 304-0014*.
- Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J., 2002. *Least Squares Support Vector Machines*. World Scientific, Singapore.
- Tan, K.C., 2001. A framework of supply chain management literature. *European Journal of Purchasing & Supply Management* 7 (1), 39–48.
- Thonemann, U.W., 2002. Improving supply-chain performance by sharing advance demand information. *European Journal of Operational Research* 142 (1), 81–107.
- Vakharia, A.J., 2002. e-business and supply chain management. *Decision Sciences* 33 (4), 495–505.
- Vapnik, V., 1995. *The Nature of Statistical Learning Theory*. Springer, New York.
- Vapnik, V., Golowich, S.E., Smola, A., 1997. Support vector method for function approximation, regression estimation, and signal processing. *Advances in Neural information processing systems* 9, 287–291.
- Watson, G., 2001. Subregimes of power and integrated supply chain management. *Journal of Supply Chain Management* 37 (2), 36–41.
- Werbos, P.J., 1990. Backpropagation through time: What it does and how to do it. *Proceedings of IEEE* 78 (10), 1550–1560.
- Yusuf, Y.Y., Gunasekaran, A., Adeleye, E.O., Sivayoganathan, K., 2004. Agile supply chain capabilities: Determinants of competitive objectives. *European Journal of Operational Research* 159 (2), 379–392.
- Zhao, X., Xie, J., Wei, J.C., 2002. The impact of forecast errors on early order commitment in a supply chain. *Decision Sciences* 33 (2), 251–280.