# A Framework for Modeling Efficient Demand Forecasting Using Data Mining in Supply Chain of Food Products Export Industry

Pongsak Holimchayachotikul[1] and Nuanlaor Phanruangrong[2]

[1] College of Arts, Media and Technology
Chiang Mai University, Chiang Mai 50200 Thailand
g490631075@gmail.com

[2] Department of Industrial Engineering, Faculty of Engineering Chiang Mai University, Chiang Mai 50200 Thailand
pookka29@yahoo.com

**Abstract.** According to the Hamburger effect, food products export industry sector, especially cooked chicken products export to Japan of Thai industry, effort has been spent in the supply chain management (SCM) of internal efficiency, solely aiming at competitiveness survival in terms of cost reduction, better quality. To meet the customer satisfaction, the company must work towards a right time and volume of demand delivery. Therefore, forecasting technique is the crucial element of SCM. The more understanding how their company use the right forecasting based on information sharing in their SCM context; the more reducing inventory and capacity planning cost increase their company competitiveness. Presently, most of companies, in this sector, do not have a right knowledge to implement the suitable forecasting system to sustain their business; furthermore, they only use top management judgment and some of economical data for forecasting decision making to production. Because the complex, stochastic, dynamic nature and multi-criteria of the logistics operations along the food products exporting to Japan of Thai industry supply chain, the existing time series forecasting approaches cannot provide the information to operate demand from upstream to downstream effectively. The objective of the paper is how to develop a conceptual framework for an innovative and simplified forecasting system implementation for this industry based on data mining including time series factors and causal factors. Then we discuss a methodology to determine appropriated implementation guideline.

**Keywords:** Conceptual Framework, Data Mining, Food products export industry, Forecasting, SCM.

## 1 Introduction

The supply chain has increased in complexity over the years, resulting in traditional forecasting methods being unsuitable for current situations. An inaccurate forecast in one place of the supply chain can affect the whole supply chain due to the bullwhip

G. Huang et al. (Eds.): DET2009 Proceedings, AISC 66, pp. 1387–1397.
springerlink.com 

effect, which also affects the manufacturing plans and any plans to conform to customer demands in terms of time and quantity - a significantly difficult problem in business.

The mentioned problem has occurred in the case study factory which is a food products export to Japan of Thai industry sector. We have found out that the customer demands are highly uncertain, causing current forecasting methods not being accurate enough and can lead to various problems in production plans such as over purchasing clutch food and planning clutch production , which results in high costs on reserved goods. Other effects are excess overtime for employees and increased costs in bringing chick up . Purchasing insufficient materials can also result in production holds, leading to loss in sales opportunities. This is why demand forecasting is very important in management of the supply chain. Leading companies have agreed that customer demand can be managed only to a certain degree. Demand fluctuation is an important reason that causes conflicts in demand and supply chain which in ideal conditions; everyone desires to know the demand and the value chain to be constant. All this is to allow companies to be able to satisfy customer requirements with lowest costs. The ideal supply chain is a supply chain that conforms to the proper amount of goods, has high certainty of work conditions, no redundant workings, no reserved goods, and also lowest cost in shipping in which goods will be manufactured, delivered, and used according to the forecasted demand within a certain timeframe. Owing to the bullwhip effect in the supply chain, regardless of where it occurs, the supply chain is affected in whole. The variance in the supply chain highly burdens companies in terms of production cost and complexity of the increased workload. The bullwhip effect management is thus necessary for maximum efficiency of the supply chain.

All said before is why running a business should consider the significance of the actual demand of customers by reducing the bullwhip effect and causing customer satisfactory. This research is appointed to apply back-propagation artificial neural networks in the demand forecasting of the supply chain to be able to realize the true demand of customers even more accurately by analyzing historical data. Today, forecasting techniques have played an important role in supply chain management because precisely forecasting customer demand results in reduction of the bullwhip effect. The forecasting techniques used in supply chain management are classified into three groups, which are 1) Traditional techniques such as moving average, exponential smoothing and ARIMA, 2) Computational intelligence techniques such as artificial neural network and support vector regression, and 3) Hybrid methods such as applying traditional techniques such as ARIMA with artificial neural networks or support vector regression, or applying fuzzy system techniques with ARIMA which presents us with techniques more complex but of higher accuracy, as shown in Table 1. The differences among the three groups are that computational intelligence and adapted techniques can import factors from time series and economical factors from the industry of interest while traditional techniques can only use time series data.

Leung (1995) has introduced an approach to applying artificial neural networks to various aspects of the supply chain management for instance, forecasting, optimum parameter achievement, model and simulation creation, and decision support. Later, Tseng (2001) presented a Fuzzy ARIMA approach to forecasting the U.S. Dollar (USD) to Taiwan Dollar (TWD) exchange rate, which demonstrates the application of

**Table 1.** Forecasting Techniques in Supply Chain Management from 90's to 2008

| Author (year) | Detail | Forecasting Method | | The best method |
|---|---|---|---|---|
| | | Traditional Techniques | Advanced Techniques | |
| Leung (1995) | A general forecasting concept of ANN in SCM | - | MLP | - |
| Tseng (2001) | The exchange rate of NT dollars to US dollars forecasting | - | Fuzzy ARIMA | - |
| Chu and Zhang (2003) | Aggregate retail sales forecasting | ARIMA | MLP | MLP |
| G. Peter Zhang (2003) | Hybrid ARIMA and ANN based on three well-know data set -the Wolf's sunspot data, the Canadian lynx data and the British pound/US exchange rate data | - | ARIMA+MLP | ARIMA+MLP |
| Chiu and Lin (2004) | Collaborative Supply Chain Planning of an alliance of small firms | - | MLP | - |
| Pai and Lin (2005) | Production value forecasting of the machinery industry in Taiwan | SARIMA ,GRNN | SVM | SVM |
| Pai and Lin (2005) | Stock price forecasting in Taiwan | ARIMA | ANN,SVM,ARIMA+SVM | ARIMA+SVM |
| Ediger *et al.* (2005) | Forecasting production of fossil sources in Turkey | Regression, ARIMA,SARIMA | - | Depend on Products |
| Lee *et al.* (2006) | Production quantity allocation for order fulfillment in scm | - | MLP | - |
| Perez (2006) | Bankruptcy forecasting review based on ANN | - | MLP | - |
| Yujun *et al.* (2006) | Short term daily load forecasting at Hebei province of China | - | ARIMA+SVM | ARIMA+SVM |
| Co and Boonsarawondse (2007) | Thailand's rice export forecasting | ARIMA, Exp Smt | MLP | MLP |
| Zou *et al* (2007) | Chinese food grain price forecasting | ARIMA | MLP | MLP |
| Arburto and Weber (2007) | Chilean supermarket demand forecasting in Chile | Naïve, Seasonal Naïve, ARIMA ,Unconditional average ARIMA | MLP, MLP with SARIMAX | MLP with SARIMAX |
| Aslanargun *et al.* (2007) | Tourist arrival forecasting to Turkey | - | MLP, RBFN | ARIMA+ MLP |
| Tannock *et al.* (2007) | Aerospace sector supply chain data-driven simulation | Exp Smt | SCMB | - |
| Bayraktar *et al.* (2008) | Role of forecasting on bullwhip effect for E-SCM application | Naïve, Average, Moving | - | - |
| Carbonneau *et al.* (2008) | Distorted demand signal forecasting based on The foundries monthly sales data was obtainedfrom the Statistics Canada table 0304-0014 and simulated data from MATLAB | Average, Trend, MLR Exp Smt, Croston's method, Syntetos-Boylan approximation | MLP, RNN, SVM | RNN, SVM |
| Gutierrez *et al.* (2008) | Lumpy demand forecasting using neural networks | - | MLP | MLP |
| Saini (2008) | Up to 7 days ahead electrical peak load forecasting | - | MLP | - |
| Vahidinasab (2008) | Day-ahead price forecasting in restructured power systems in Pennsylvania-New Jersey – Maryland (PJM) market | - | MLP integrated with fuzzy c-mean | - |

**Noted:** ANN: RNN: Recurrent Neural Network, SVM: Support Vector Machine, MLR: Multiple Linear Regression, ARIMA: Auto Regressive Integrated Moving Average, MLP: Multilayer Perceptron learning with back-propagation , SARIMA : The seasonal time series auto regression integrated moving average , GRNN : general regression neural network, SCMB: Supply-Chain Model Builder, RBFN : Radial basis function network, Exp Smt : Exponential Smoothing.

statistical techniques with computational intelligence for accurate forecasting. A variety of forecasting techniques has been applied to the complex supply chain management, leading researchers to uncover techniques for their industry field of interest. For example, Carbonneau *et al.* (2008) has applied machine learning to forecast supply chain demand by neural network comparison, recurrent neural networks, and support vector machine with traditional forecasting techniques consisting of Naïve Forecasting, average forecasting, average moving forecasting, tendency forecasting, and multiple regression. By using actual data from Canadian Foundries datasets and simulated data, RNN and LS-SVM yielded best results compared to traditional methods.

From the year 2000, ARIMA forecasting techniques have been widely used for application of forecasting and at the same time, computational intelligence has also gone widespread, which led to the comparison of both techniques in industrial data forecasting. The results have shown that computational intelligence techniques have higher precision than traditional ones such as ARIMA as displayed in Table 1. Pai and Lin (2005) demonstrated that the accuracy of computational intelligence techniques are more appropriate for complicated industries than traditional techniques based on manufacturing value of the machine industry in Taiwan. Later on, Co and Boonsarawondse (2007) have forecasted the rice export of Thailand using artificial neural networks that imports time series, smoothing exponential time series and ARIMA. The imported data was dated from January 1996 to December 2004 and January 2005 to December 2005. The results of the trial were that artificial neural networks were rated best with ARIMA in second.

Due to the uncertainty of the supply chain and both internal industrial factors and external economical factors, researchers need to develop forecasting techniques that support this uncertainty by combining traditional techniques such as ARIMA with computational techniques. For instance, G. Peter Zhang (2003) introduced a combined technique of ARIMA and artificial neural networks using three standard data sets in the experiment which are Wolf's sunspot, Canadian lynx, and the exchange rate of Dollars and British Pounds. The result was that the selected approach was able of accurate forecasting. On the other hand, Pai and Lin (2005) have introduced an applied model of ARIMA and support vector regression since both techniques are powerful in both linear and non-linear familiarization of Taiwan's stock market. Results when compared to individual techniques such as artificial neural networks, support vector regresstion, and ARIMA, are notably better. In 2008, Vahidinasab (2008) introduced a price forecasting system for electronic appliance repair and maintenance in the Pennsylvania-New Jersey-Maryland using artificial neural networks trained by various algorithms together with C-Mean fuzzy logic, which were able to forecast more precisely than original methods.

## 2 Background

### 2.1 Neural Network

Multilayer perceptron (MLP) was wildly used in many researches. It comprises of processing elements called neurons located in layers. Some or all of the outputs in each layer are connected to one or more inputs in the next layer. The input layer is the first layer where the MLP receives input parameters. The output layer is the final layer where the outputs are provided to the user. The hidden layers are located between input and output layer. The task of an individual neuron is to take inputs from the outside, or from other neurons connected to it, and sum these inputs according to their weight or the strength of the connection of each input. A transfer function is then adopted to produce the output. Quickprop (BP) is the most extensively adopted learning algorithm. The output layer provides a response depending on training history of the network. The trained network should be able to correctly predict outputs for unseen input conditions.

### 2.2 Recurrent Neural Network Support Vector Regression

The main different process of Recurrent neural networks (RNN) from general ANN is the input for the neurons of the same layer or those of the previous layers coming from the pervious outputs. When time series data included in RNN, the approach called ''back-propagation through time'' could be opted for train a RNN on a given training set. Figure. 1 shows schematically the structure of RNN for the supply chain demand forecasting problem.

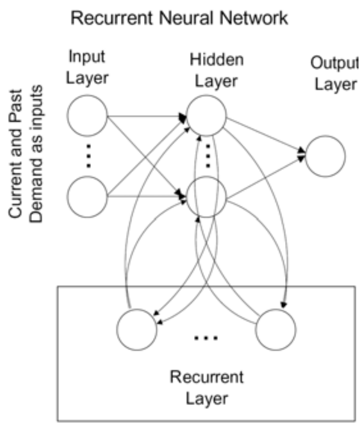


**Fig. 1.** Recurrent neural network for demand forecasting (MathWorks, 2009)

### 2.3 Support Vector Regression

Support vector regression (SVR) is a novelty universal function approximators, based on the structural risk minimization principle from statistical learning theory (Vapnik, 1995).

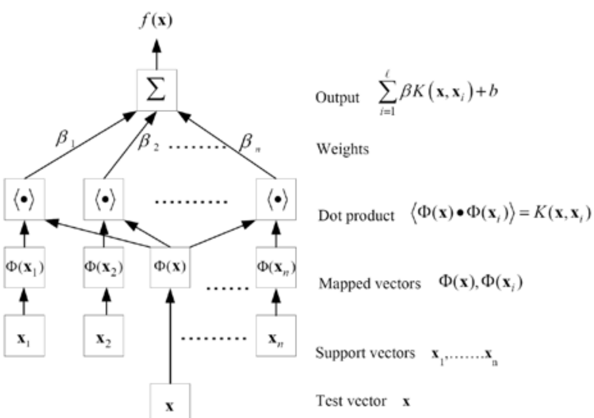


**Fig. 2.** Support Vector Regression Structure

It is superior to the empirical risk minimization principle on which ANN and linear regression Figure. 2 shows schematically the structure of SVR for the general forecasting problem.

### 2.4 Adaptive Neuro-fuzzy Inference System

A fuzzy inference system (FIS) employing fuzzy if-then rules can model the qualitative aspects of human knowledge and reasoning process without employing precise quantitative analyse (Jang, 1993). Moreover, ANFIS has been recognized as a potential tool that can facilitate the effective development of models by combining information from various sources, such as empirical models, heuristics and data. Hence, in most cases neuro-fuzzy models can be better used to elucidate solutions to users than completely black box models such as neural networks. Figure 3. shows the ANFIS architecture.

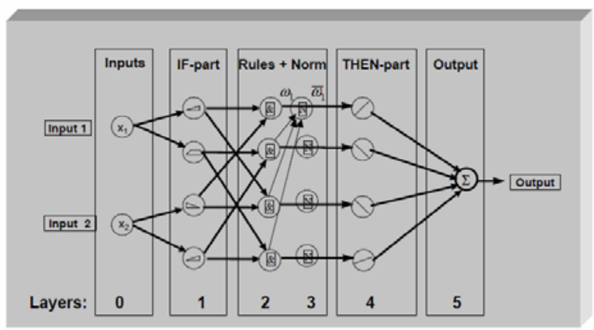


**Fig. 3.** The ANFIS architecture

## 3 Methodology

From documents and related researches, we can define the processing methodology. A schematic diagram of the proposed conceptual framework is shown in Figure. 4. This comprises of the combination of feature selection, ANN, SVR, and time series forecasting techniques applied to build the innovative and simplified forecasting system implementation for supply chain of cooked chicken products export to Japan of Thai industry based on data mining including time series factors and causal factors such as gold price, oil price, dollar-bath and yen-bath exchange rate.

The phases are as follows:

### 3.1 Exploring and Selection of Data

Current demand in case study company of cooked chicken products export to Japan of Thai industry supply chain Primary data analysis, following the concept of demand forecasting in SCM based on collaborative planning forecasting and replenishment (CPFR), of the case study factory are conducted by means of basic demand and

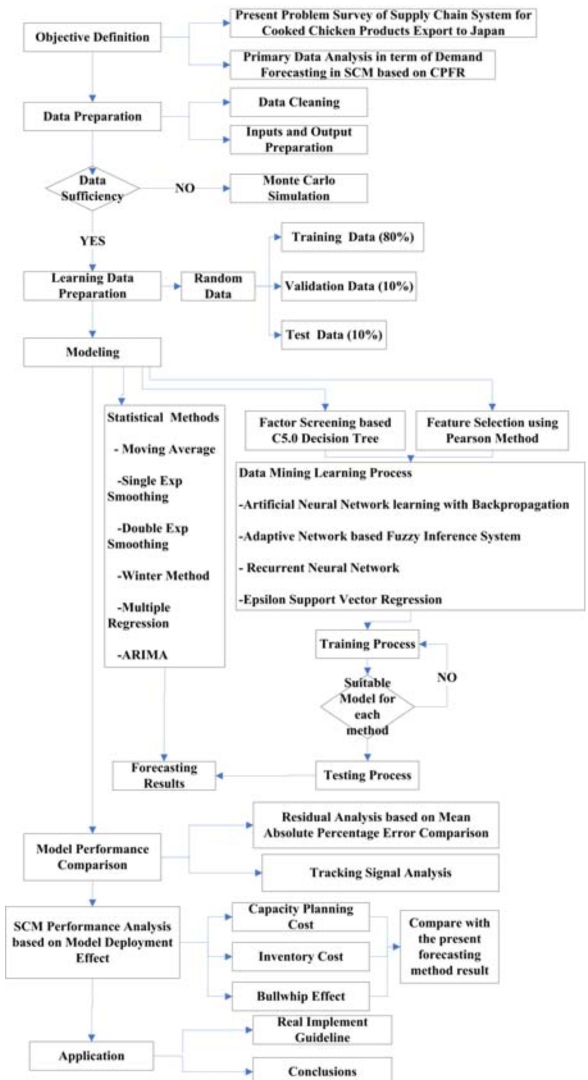


**Fig. 4.** Schematic diagram of the proposed framework

supply data of the supply chain to measure the efficiency of customer conformance and production and resource management to define the direction of solving the problems.

## 3.2 Data Preparation and Sufficiency

Data cleaning process was performed to eliminate noise data. Next, the cleaned data was rearranged following input and output format for each method. Data sufficiency was employed to indicate the suitable quantity of data for data mining learning process. Moreover, monte carlo simulation will generate the data in case of insufficiency of data

### 3.3 Learning Data Preparation

The data from 3.2 can used for forecast modeling by randomly dividing the data sets into three groups for training, validation and test set were 90, 10 and 10 percentage of all data, respectively

### 3.4 Time Series Based on Statistical Method

#### 3.4.1 Basic Data Testing

A graph plot of demand, depicted in Figure 5, and number of cooked chicken export to Japan of case study company in term of tons was used to easier describe the tendency, product life cycle, and abnormal events.

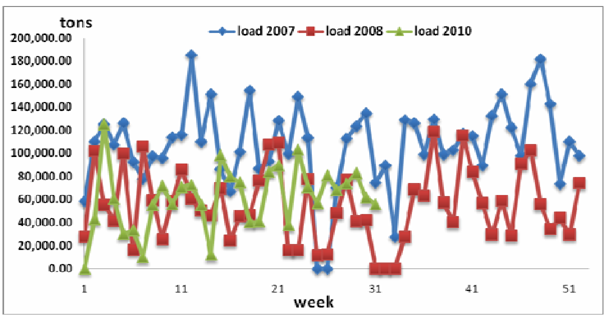


**Fig. 5.** Real demand of cooked chicken from Japan customer (2007- July 2009)

#### 3.4.2 Modeling

Moving Average, Single Exponential Smoothing, Double Exponential Smoothing, and Winter's Method were used to forecast the values of 12 periods after. Forecasting was done period by period.

#### 3.4.3 ARIMA Forecasting

The methods are as follows

1) Model Identification – Historical data analysis was used for format selection by statistical calculations to assist in the decisions. First, the time series data must be considered that it is in a balanced state, and has scattering around stationary average values consisting of:
   i. The series is stationary, having stationary average values and variation.
   ii. Formatting of the time series with ACF and PACF graphs.
2) Parameter Estimation – An estimation of coefficients of the model from the selected format in phase 1 using the least square method. Initial estimates were fed to the computer to process until final estimates were obtained. The calculations were iterative until the least amount was returned.

3) Model Diagnostics – Inspection of appropriateness of the model was carried out using Modified Box – Pierce ( Ljung - Box ) Chi – Square Statistic
4) Forecasting

- Forecasting based on the appropriate model.
- Reasonable forecasting and must be explainable.
- Statistical graphs and confidence intervals are used in the consideration of appropriateness of the model.

5) Accuracy testing of the estimation of power usage in the metal basis and woven product industry using ARIMA was performed. The index used for measuring error is the mean absolute percentage square error (MAPE)

## 3.5 Data Mining Learning Process

### 3.5.1 Input Feature Selection Process

Both Pearson's chi-square and C.50 decision tree were used for indicating significant input factors related with the next week customer demand at the 95 percentage of confident interval.

### 3.5.2 Develop the Suitable Structure of Each Data Mining Model

Support vector regression, artificial neural network learning with backpropagation, recurrent neural network and ANFIS. The data from 3.3 can be used for forecast modeling development. In addition, MAPE of test set was used as the threshold for indicating the appropriated model based on over fitting checking with residual analysis for each data mining technique. If the MAPE of test set was less than or equal five percentage, it can imply that the structure of this method will be the suitable structure.

### 3.5.3 Model Performance Analysis and Comparison

MAPE of each the suitable model was employed for comparison to find the best technique for demand forecasting of case study company. Moreover, the tracking signal of each the suitable model was performed for double checking in term of real usage of each model.

### 3.5.4 Bullwhip Effect Analysis

As a forecasting result from suitable data mining model, the bullwhip effect analysis was conducted to adjust production planning, inventory management. Then the new forecasting system performance compared with the present forecasting of case study company.

### 3.5.5 Real Implementation

To develop the innovative and simplified forecasting system implementation for supply chain of cooked chicken products export to Japan of Thai industry based on data mining including time series factors and causal factors guideline, technology transfer in term of modeling will conduct to apply in case study company based on CPFR concept.

## 4 Research Results

The test results of demand forecast simulation from time series and data mining in supply chain of the case study company are proposed to real forecasting system implementation using data mining technique. Business processes in food products export to Japan industry of Thailand sectors are designed to add value for the customer and should not include unnecessary factor. To obtain the appropriated forecasting system from this framework, we define the most relevant factor that include in future model. To this aim, the following factor should be considered market share, common stock of Thailand and Japan market in term fresh food, operation times, representing the ability to respect the decided deadlines, number of export products produced (finished goods)and customer order arrival information.

## 5 Conclusion

The concept of forecasting system implementation integrates with data mining approach to develop a conceptual framework for the exported food products industry is presented.

Developing optimal forecasting system for production system is a complex problem. For analyzing complicated systems, forecasting based on modern data mining plays a dramatically important role. Feature selection tools can help for removing of unnecessary input factors in the data mining learning process. For the better system forecasting system, data mining such as support vector regression, artificial neural network learning with backpropagation, recurrent neural network and ANFIS were used to model this forecasting system as the candidate technique in order to determine the suitable method for case study company SCM. Further research in each step of a framework will develop depended on internal, external economical factors based on exported food products SCM expert. Next, the real research implementation incorporated with some data to see the real effect should be done.

## Acknowledgments

The authors would like to acknowledge Asst.Prof.Dr Apichat Sopadang and Miss. Pachinee Payongyam for the data of supply chain system for cooked chicken products export to Japan used in the case study.

## References

Aburto, L., Weber, R.: Improved supply chain management based on hybrid demand forecasts. Applied Soft Computing Journal 7, 136–144 (2007)

Aslanargun, A., Mammadov, M., Yazici, B., Yolacan, S.: Comparison of ARIMA, neural networks and hybrid models in time series: Tourist arrival forecasting. Journal of Statistical Computation and Simulation 77, 29–53 (2007)

Bayraktar, E., Lenny Koh, S.C., Gunasekaran, A., Sari, K., Tatoglu, E.: The role of forecasting on bullwhip effect for E-SCM applications. International Journal of Production Economics 113, 193–204 (2008)

Carbonneau, R., Laframboise, K., Vahidov, R.: Application of machine learning techniques for supply chain demand forecasting. European Journal of Operational Research 184, 1140–1154 (2008)

Chiu, M., Lin, G.: Collaborative supply chain planning using the artificial neural network approach. Journal of Manufacturing Technology Management 15, 787–796 (2004)

Chu, C.W., Zhang, G.P.: A comparative study of linear and nonlinear models for aggregate retail sales forecasting. International Journal of Production Economics 86, 217–231 (2003)

Co, H.C., Boosarawongse, R.: Forecasting Thailand's rice export: Statistical techniques vs. artificial neural networks. Computers and Industrial Engineering 53, 610–627 (2007)

Ediger, V.S., Akar, S., Uğurlu, B.: Forecasting production of fossil fuel sources in Turkey using a comparative regression and ARIMA model. Energy Policy 34, 3836–3846 (2006)

Gutierrez, R.S., Solis, A.O., Mukhopadhyay, S.: Lumpy demand forecasting using neural networks. International Journal of Production Economics 111, 409–420 (2008)

Jang, J.S.R.: ANFIS: Adaptive network based fuzzy inference system. IEEETransactions on Systems, Man, and Cybernetics 23(3), 665–684 (1993)

Lee, Y.H., Jung, J.W., Eum, S.C., Park, S.M., Nam, H.K.: Production quantity allocation for order fulfilment in the supply chain: A neural network based approach. Production Planning and Control 17, 378–389 (2006)

Leung, H.C.: Neural networks in supply chain management (1995)

Khaw, J.F.C., Lim, B.S., Lim, L.E.N.: Optimal design of neural network using the Tagushi method. Neurocomputing 7, 245–255 (1999)

MathWorks, I.: Using MATLAB, Version 7. MathWorks, Inc. (2009)

Pai, P.F., Lin, C.S.: A hybrid ARIMA and support vector machines model in stock price forecasting. Omega 33, 497–505 (2005)

Pai, P.F., Lin, C.S.: Using support vector machines to forecast the production values of the machinery industry in Taiwan. International Journal of Advanced Manufacturing Technology 27, 205–210 (2005)

Perez, M.: Artificial neural networks and bankruptcy forecasting: A state of the art. Neural Computing and Applications 15, 154–163 (2006)

Saini, L.M.: Peak load forecasting using Bayesian regularization, Resilient and adaptive backpropagation learning based artificial neural networks. Electric Power Systems Research 78, 1302–1310 (2008)

Tannock, J., Cao, B., Farr, R., Byrne, M.: Data-driven simulation of the supply-chain-Insights from the aerospace sector. International Journal of Production Economics 110, 70–84 (2007)

Tseng, F.M., Tzeng, G.H., Yu, H.C., Yuan, B.J.C.: Fuzzy ARIMA model for forecasting the foreign exchange market. Fuzzy Sets and Systems 118, 9–19 (2001)

Vahidinasab, V., Jadid, S., Kazemi, A.: Day-ahead price forecasting in restructured power systems using artificial neural networks. Electric Power Systems Research 78, 1332–1342 (2008)

Vapnik, V.N.: An overview of statistical learning theory. IEEE Transactions on Neural Network 10(5), 988–999 (1999)

Wang, Y.M., Elhag, T.M.S.: A comparison of neural network, evidential reasoning and multiple regression analysis in modelling bridge risks. Expert Systems with Applications 32, 336–348 (2007)

Yujun, H., Youchan, Z., Dongxing, D.: Research on hybrid ARIMA and support vector machine model in short term load forecasting (2006)

Zou, H.F., Xia, G.P., Yang, F.T., Wang, H.Y.: An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. Neurocomputing 70, 2913–2923 (2007)