Forecasting using Decomposition and Combinations of Experts

Pankaj Gulhane Bernard Menezes Timma Reddy Kalam Shah S. A. Soman

Kanwal Rekhi School of Information Technology
IIT Bombay
Powai, Bombay, 400 076
INDIA.
Telephone no.:+91-22-25767906
{pankaj, bernard, kunduru, kalam}@it.iitb.ac.in, soman@ee.iitb.ac.in

Abstract – We study the effect of decomposing a series into multiple components and performing forecasts on each component separately. The focus here is on sales data - most of the series considered display both seasonality and trend. Hence the original series is decomposed into trend, seasonality and an irregular component. Multiple forecasting 'experts' are used to forecast each component series. These range from different feedforward neural network topologies to Holt-Winter, ARIMA (of various orders) and double exponential smoothing. We compare the forecast errors with and without decomposition. We study the result of combining using the mean/median of all expert forecasts. Since our space of composite experts runs into the thousands, we experiment with more limited cardinalities using greedy elimination and best expert pair.

Index Terms: Time series forecasting, Neural Networks, Genetic Algorithms, Decomposition, Combining Techniques

I. INTRODUCTION

Sales forecasting is an important part of supply chain management – both at the retailer end and at the distributors, manufacturers and suppliers. Timely and accurate sales forecasts are crucial in bridging the gap between supply and demand, thereby decreasing inventory holding costs while maintaining a negligible probability of stock-out. Much work in sales forecasting has centered around the comparison of linear statistical models such as ARIMA [4][5], the Holt-Winter approach (exponential smoothing of level, trend and seasonality) [19] and the use of artificial neural networks (ANNs) [1][7][21].

While neural networks have been proposed for forecasting for well over a decade, there is no clear consensus on how one may proceed to determine the number of input neurons, hidden layer neurons or even size of the training set. For example, [16] use 1 and 10 neurons in the input and hidden layers to forecast a certain retail department stores' data but 4 and 2 neurons respectively for a time series of durable goods. Even in the area of pre-processing the data set, there is no clear verdict with some researchers favoring the view that ANNs can infer trends, seasonality and cyclicity in data [10] while others advocating some form of pre-processing [17].

One of the goals of this study is to investigate the effect of a specific form of pre-processing - series decomposition. Very recent work in this area [16] confirms our findings that de-trending and deseasonalizing the data greatly help in improving forecasts. Our work differs from theirs in that we choose and combine [2][8] a multitude of experts – both neural and statistical to forecast each component series. The neural experts are feedforward ANNs with different topologies and the statistical experts are ARIMA models of assorted orders. We perform forecasts for each series separately which are then used to derive the forecast for the original series.

In [7] for example, data over a 10-year period is used to train the ANN. The model is validated over the course of the following one year. In our work, all models use the first 2.5 years (30 months) exclusively for model training/validation and testing commences in the 31st month. Since our goal is to build and deploy a forecast engine for use in centralized supply chain management [11], we cannot assume that our clients maintain accurate and complete data over longer periods of time.

Also, model refinement does not stop but continues with every new data point. So, for example the parameters of a given ARIMA model change every month unless stated otherwise. The actual parameters are decided using a genetic algorithm (GA) [9].

The paper is organized as follows. Section II lists the basic forecasting techniques employed by us. In Section III we present our main strategy – decomposition and the use of multiple experts to forecast each component. We mention ways in which we have combined experts for our study. Section IV contains the main results – a comparison with and without decomposition. The effect of using mean/median of all experts (neural and statistical), greedy elimination of experts and the best performance of a pair of experts are compared. Section V contains the main conclusions.

II. Forecasting Techniques and Procedures

A. Holt-Winter Forecasting

The basic Holt-Winter[14] forecasting method with multiplicative seasonality (exponential smoothing of level (S_t), trend (T_t) and seasonal index (I_t)) is described by

$$S_t = \alpha\,(D_t / I_{t-p}) + (1-\alpha)\,(S_{t-1} + T_{t-1}) \quad (1)$$

$$T_t = \beta\,(S_t - S_{t-1}) + (1-\beta)\,T_{t-1} \quad (2)$$

$$I_t = \gamma\,(D_t / S_t) + (1-\gamma) I_{t-p} \quad (3)$$

Here p is the number of observation points in a cycle (p = 4 for quarterly data). α, β and γ are the smoothing constants. The forecast at time t for time $t+i$ is $(S_t + iT_t)\, I_{t-p+i}$.

B. Seasonal ARIMA

The general multiplicative seasonal *ARIMA (p, d, q) X (P,D,Q)* model is expressable as

$$\Phi(B)\, \emptyset(B^s)(1-B^d)\{(1-B^s)^D X(t) = C_0 + \theta(B)\Theta(B^s)\varepsilon(t) \quad (4)$$

$t=1,2,3,....$

Here X(t) is a time series, ε(t) is a sequence of i.i.d. (zero mean and normally distributed) errors, d and D are the orders of non-seasonal and seasonal differencing for the time series and $\Phi(B)$, $\emptyset(B^s)$, $\theta(B)$ and $\Theta(B^s)$ operators are polynomials in B with the following general forms

$$\Phi(B) = 1 - \Phi_1(B) - \Phi_2(B^2) - ... - \Phi_p(B^p) \quad (5)$$

$$\emptyset(B^s) = 1 - \emptyset_1(B^s) - \emptyset_2(B^{2s}) - ... - \emptyset_P(B^{Ps}) \quad (6)$$

$$\theta(B) = 1 - \theta_1(B) - \theta_2(B^2) - ... - \theta_q(B^q) \quad (7)$$

$$\Theta(B^s) = 1 - \Theta_1(B^s) - \Theta_2(B^{2s}) - ... - \Theta_Q(B^{Qs}) \quad (8)$$

The polynomials $\Phi(B)$ and $\theta(B)$ capture the non-seasonal behavior and $\emptyset(B^s)$ and $\Theta(B^s)$ capture the seasonal behavior of the series. The differencing orders d and D typically have a value 0 or 1. For non-seasonal ARIMA, $P=D=Q=0$.

C. Using GA for ARMA Parameter Estimation

We used GAlib genetic algorithm package [24] for finding the parameters of a particular *ARMA(p, q)* model. Each chromosome contains $p+q$ real genomes, and each genome represents one parameter of the model. We used value encoding for the parameters, Roulette wheel [12] (with the combination of elitism) as the selection method, and Mean Absolute Percentage Error (MAPE) as the fitness measure.

The following algorithm was used for finding the *ARMA(p, q)* parameters:

1. Randomly generate initial population of size (P_s).
2. For each chromosome, find the predictions for the training data set.
3. Compute the fitness value of each chromosome.
4. Generate new population from the old population by applying the genetic operators.
5. Repeat steps 2 to 4 for 'number of generations' (N_g) times.
6. Select chromosome having least fitness value.

Our experiments used the following GA parameters,

- Crossover probability = 0.9
- Mutation probability = 0.1
- Population size (P_s) = 200
- Number of generations (N_g) = 1000.

D. Using neural networks

We adapted an open-source package[13] in Java which trains a multi-layer neural network using back propagation. The main parameters we experimented with were:

- The number of input neurons (each corresponds to a lag from the point being forecast)
- The number of hidden layers
- The number of neurons in a hidden layer
- The topology (pattern of connections between consecutive layers)

We used batch training[18] with the standard back-propagation algorithm. Based on our experiments, we decided to use *tanh* and a *linear* activation function for the hidden and output layer neurons respectively. The training sets were divided into 2 parts, viz. 80% for training and 20% for cross-validation. After each routine of weight updates in the network, the Cross-Validation Error Ratio (CVER) is calculated by the package as follows:

$$CVER = \frac{\sum_{i=1}^{N} | P_i.output - P_i.\text{target} |}{N \times CVDev} \tag{9}$$

where

$P_i, 1 \le i \le N$, are the N Cross-Validation patterns

$P_i.output$ is the output of the neural network for the i^{th} pattern

$P_i.target$ is the expected output for the i^{th} pattern

$$CVDev = \frac{2 \times \sum_{i=1}^{N} | P_i.\text{target} - avgOut |}{N}$$

$$avgOut = \frac{\sum_{i=1}^{N} P_i.\text{target}}{N}$$

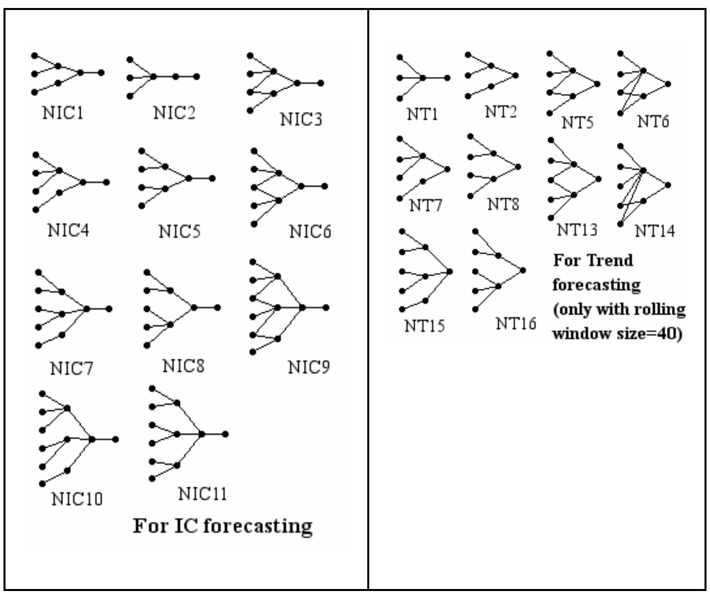

FIGURE 1: NEURAL NETWORK TOPOLOGIES USED

The training is stopped if any of the following conditions becomes true:

1. The number of epochs crosses *20000*.
2. $CVER < 0.04$
3. *CVER* increases with an increment $< 10^{-7}$

The neural network was retrained after forecasting each point in the series. We tried a number of values of the rolling window size for training and finally selected window size=100 for the Irregular Component and windows of sizes 40 and 70 for the Trend Component. (These components are described in section III.A) Thus we had 11 topologies for forecasting the Irregular Component and 10x2(window sizes) neural forecasters for the Trend Component. The chosen topologies are shown in Figure 1. For the Trend Component, topologies numbered NT3, NT4, NT9, etc. are identical to NT1, NT2, NT5, etc., but with a different rolling window size=70.

III. OUR STRATEGY

For our forecasting experiments, we used a total of 11 series from the Time Series Library [22] representing monthly sales (Table 1). Except for Hsales, these represent sales of Fast Moving Consumer Goods (FMCGs). We employed decomposition and expert combining for each series as explained below.

A. Series Decomposition

We use a form of pre-processing by decomposing a series, D, into three component series - Trend (T), Seasonality (S) and the Irregular component (IC). Assuming that series can be combined using the operators, + and ×, we have eight ways of combining these series (i.e. $S+T \times I$, $S+T+I$, etc). The pure multiplicative model ($D = T \times S \times I$) seemed to be superior in representing sales data and so we used it in the experiments reported here.

Let D_t, T_t, S_t and I_t denote the t^{th} point of the respective series. We define the component series below in a manner similar to [15].

$$T_t = \frac{\sum_{i=0}^{11} D_{(t-i)}}{12} \tag{10}$$

$$S_t = avg\left(\frac{D_t}{T_t}, \frac{D_{t-p}}{T_{t-p}}, \frac{D_{t-2p}}{T_{t-2p}}, \frac{D_{t-3p}}{T_{t-3p}}, \frac{D_{t-4p}}{T_{t-4p}}, \cdots\right) \tag{11}$$

where p is the seasonality period

$$I_t = \frac{D_t}{T_t \times S_t} \tag{12}$$

B. Combining Experts' Forecasts

We have employed the services of 40, 2 and 30 experts for forecasting T, IC and S respectively. The names and notations for the experts for each component are listed below:

TABLE 1: TIME SERIES USED FOR ANALYSIS IN THIS PAPER

ABRAHAM12	Monthly gasoline demand Ontario gallon millions 1960 - 1975. Source: Abraham & Ledolter (1983).
DRYWHITE	Monthly Australia sales of dry white wine: thousands of litres. Jan 1980 - Jul 1995. Source: ABS.
FORTIF	Monthly Australian sales of fortified wine: thousands of litres. Jan 1980 - Jul 1995. Source: ABS.
HSALES	Monthly sales of new one-family houses sold in the USA since 1973. Source: Makridakis, Wheelwright and Hyndman (1998).
PAPER	Jan 1963 – Dec 1972 printing and writing paper (10-year monthly)
REDWINE	Monthly Australian sales of red wine: thousands of litres. Jan 1980 - Jul 1995. Source: ABS.
ROSE	Monthly Australian sales of rose wine: thousands of litres. Jan 1980 - Jul 1995. Source: ABS.
SPAPER	CFE specialty writing papers monthly sales. Source: Makridakis & Wheelwright (1989).
SPARKLING	Monthly Australian sales of sparkling wine: thousands of litres. Jan 1980 - Jul 1995. Source: ABS.
SWEETWHITE	Monthly Australian sales of sweet white wine: thousands of litres. Jan 1980 - July 1995. Source: ABS.
WINE.DAT	Monthly Australian wine sales: thousands of litres. Jan 1980 - July 1995 (total wine). Source: ABS.

Trend experts

20 Neural Network: {NT1, NT2, …., NT20}
8 AR: {AR(2), AR(3), ….., AR(9)}
8 AR2[1]: {AR2(2), AR2(3), …., AR2(9)}
2 Holt (non adaptive, yearly adaptive): HNA, HYA
2 Double Exponential (non adaptive, yearly adaptive): DENA, DEYA

Seasonality Experts

Previous Value: RW (Random Walk)
Winter (yearly adaptive): W

IC experts

11 Neural Network: {NIC1, NIC2, … , NIC11}
8 AR2: {AR2(2), AR2(3), ….., AR2(9)}
11 ARMA: {ARMA(0,1), ARMA(0,2), ARMA(1,0), ARMA(1,1), ARMA(1,2), ARMA(2,0), ARMA(2,1), ARMA(2,2), ARMA(3,0), ARMA(3,1), ARMA(3,2)}

[1] The parameters take only positive values.

Experts, both neural and statistical, use the first 30 months exclusively for training. The weights in the NN get updated every subsequent month. In the case of statistical models, 3 options may be used – once set after the 30th month, they are not changed, or they may be updated either monthly or yearly. The MAPE is calculated starting from the 31st month. We then combined all these experts to obtain a total of 40 x 2 x 30 = 2400 expert combinations. We used MAPE as the error measure since it is often used in sales forecasting and is one of the better measures investigated[3]. Finally, the forecast horizon=1 in all experiments conducted here.

There are several approaches to factoring in the opinion of experts (expert combinations). One approach is to take the mean/median value of their forecasts. The more sophisticated (but much harder) approach is to choose a "best set" based on their past performance perhaps assigning weights to their individual forecasts. Even the intermediate strategy of using a static best set for a given series is of exponential complexity. To start with, we used a simple heuristic, Greedy Elimination, to find a suitable combination of experts starting with all experts. Then, at each step we eliminate the expert whose absence decreases the MAPE by the largest amount (or increases it by the smallest amount). This is repeated until all except the last expert have been eliminated.

TABLE 2: MAPES WITH AND WITHOUT DECOMPOSITION (USING GENETIC ALGORITHM ON ARMA)

Series	Without Decomposition (GA)			With decomp (GA)
	MAPE	p	q	
Abraham	3.001	3	0	3.200
Dry	9.821	1	0	8.888
Fortif	9.012	0	1	7.951
Hsales	9.552	3	3	7.672
Paper	6.546	0	1	4.708
Red	11.278	0	1	9.781
Rose	17.168	1	0	13.148
Spaper	10.639	3	0	8.670
Spark	14.410	0	1	13.226
Sweet	18.631	1	1	15.004
Wine	8.277	1	0	7.387

IV. RESULTS

To study the effect of the above decomposition, we fit the *ARMA(p,q)* model for the difference series at lag 12 (because all the series have a seasonality period of 12).

TABLE 3: BEST MODEL, BEST PAIR AND GREEDY ELIMINATION MAPES USING 2400 EXPERTS

Series	Top 2 experts	Best pair	Greedy Algorithm
Abraham	2.870 (NT15, W, ARMA(3,2)) 2.873 (AR(7), W, ARMA(3,2))	2.846 (AR(7), RW, NIC1) + (AR(7), W, ARMA(3,2))	2.738 (9)
Dry	8.724 (NT8, W, NIC3) 8.746 (AR(9), W, NIC3)	8.606 (NT8, W, NIC8) + (AR(9), W, NIC4)	8.502 (8)
Fortif	7.889 (AR(5), W, AR2(6)) 7.904 (AR(5), RW, AR2(6))	7.834 (AR(9), W, AR2(7)) + (HNA, RW, NIC1)	7.737 (17)
Hsales	7.588 (AR(5), W, AR2(6)) 7.593 (AR(2), W, AR2(6))	7.587 (AR(5), W, AR2(7)) + (AR(5), W, AR2(9))	7.582 (14)
Paper	4.634 (NT6, W, AR2(7)) 4.639 (NT19, W, AR2(7))	4.525 (NT18, W, AR2(3)) + (NT19, W, AR2(6))	4.489 (11)
Red	9.311 (AR2(8), W, ARMA(1,1)) 9.331 (AR(8), W, ARMA(2,1))	9.194 (NT13, W, AR2(5)) + (AR(2), RW, AR2(2))	9.076 (12)
Rose	12.772 (AR(9), W, NIC6) 12.908 (AR(9), W, NIC5)	12.030 (NT2, RW, AR2(3)) + (NT15, W, ARMA(0,2))	11.663 (27)
Spaper	8.000 (NT19, W, ARMA(1,1)) 8.008 (DENA, W, ARMA(1,1))	7.955 (NT19, W, ARMA(2,1)) + (DENA, W, ARMA(1,2))	7.900 (11)
Spark	12.957 (AR2(9), W, AR2(7)) 13.000 (AR2(9), RW, AR2(7))	12.776 (NT2, W, ARMA(2,0)) + (AR2(9), W, AR2(8))	12.480(15)
Sweet	14.701 (AR(6), W, AR2(7)) 14.814 (NT15, W, NIC10)	13.915 (NT4, W, NIC11) + (AR(6), W, AR2(9))	13.290 (20)
Wine	7.261 (NT14, W, ARMA(0,1)) 7.266 (NT14, W, ARMA(1,0))	7.040 (NT9, W, NIC6) + (NT15, W, ARMA(3,0)	6.892 (11)

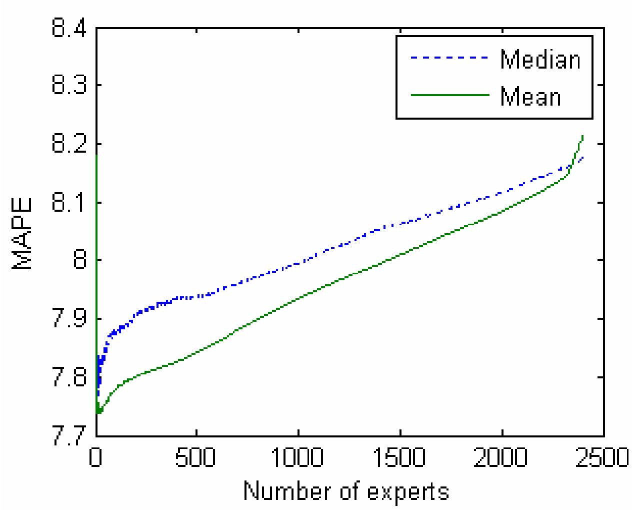

FIGURE 2: PROGRESS OF GREEDY ELIMINATION ALGORITHM USING MEAN AND MEDIAN (SERIES: FORTIF)

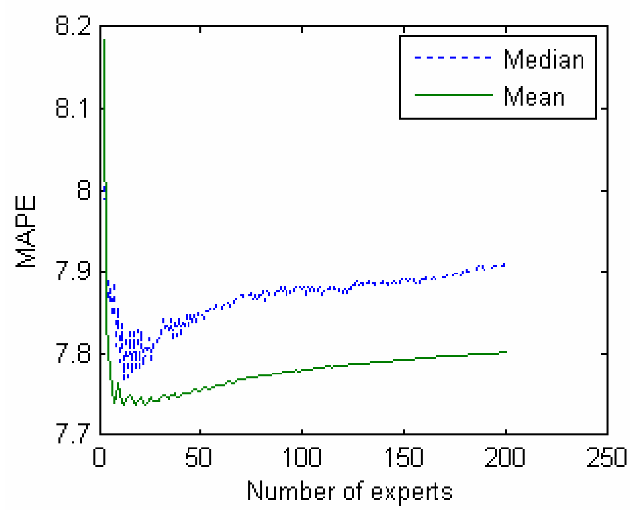

FIGURE 3: PROGRESS OF GREEDY ELIMINATION ALGORITHM USING MEAN AND MEDIAN (ZOOMED IN) WHEN NUMBER OF EXPERTS GO DOWN FROM 200 TO 1 (SERIES: FORTIF)

We experimented with $p=0,1,2,3$ and $q=0,1,2$ using GA as explained earlier. We used a yearly adaptive model (i.e. the parameters of the *ARMA* model were recalculated using GA every year but p and q of the ARMA model remained the same).

Table 2 shows the least MAPE obtained for each series using this method and the values of p and q that had given least MAPE. Except for one series (Abraham), it is clearly evident that decomposition helps significantly in reducing the MAPE.

The next set of experiments involves using the entire congregation of 2400 experts. Table 3 shows the top two experts for each series and their MAPEs. It also shows the best combination of two experts. While there is negligible improvement in combining two experts in the

Hsales series, the improvement for Rose and Sweet is 5-6% and averages 2% in the other cases. One immediate observation is the diversity of experts that yield best forecasts across series. Also, it is usually the case that neither of the experts in the best pair combination is the best in its own right.

TABLE 4: MAPES USING HOLT-WINTER AND MEAN/MEDIAN OF ALL COMBINATIONS OF EXPERTS

Series	HW monthly adaptive	Median	Mean
Abraham	3.286	3.073	3.035
Dry	8.324	8.937	9.054
Fortif	7.641	8.177	8.213
Hsales	10.453	8.220	8.322
Paper	5.065	4.882	5.013
Red	9.231	9.469	9.428
Rose	14.473	12.712	12.568
Spaper	8.568	8.696	8.843
Spark	12.872	13.456	13.452
Sweet	16.880	15.314	15.220
Wine	7.568	7.381	7.458

Greedy elimination improves MAPE to a far greater extent – nearly 4% on the average and about 10% for Rose and Sweet. Table 3 also shows in parentheses the number of experts in the lowest-MAPE combination using greedy elimination. The MAPE of the residual combination after each expert removal is plotted in Figure 2. Figure 3 zooms into the last 200 iterations of greedy elimination. Note that the oscillations are more prominent when using the median forecast of the combination of experts because of the consecutive removal of possibly balancing experts.

In a practical setting, the entire series will not be available before hand and the model parameters or the underlying model itself may change over time. One option is to obtain forecasts using the monthly adaptive Holt-Winter method wherein the smoothing constants are updated each month. Table 4 shows that even a naïve combining method using the median of all experts' forecasts does better in over 50% of the series compared with monthly adaptive Holt-Winter forecasting.

In the definition of S (Section III.A), all previous years were equally weighted. In our recent work, we give more weight to recent years using, for example,

$$S_t = 0.4 \times \frac{D_t}{T_t} + 0.3 \times S_{t-1} + 0.2 \times S_{t-2} + 0.1 \times S_{t-3} \tag{13}$$

TABLE 5: BEST MODEL MAPES (FORECASTED USING NN) COMPARING NEW AND OLD DECOMPOSITION METHODS

Series	Old decomposition	New decomposition
Abraham	3.149	2.868
Dry	8.724	8.404
Fortif	8.241	7.621
Hsales	8.181	9.193
Paper	4.738	4.937
Red	9.429	9.073
Rose	12.853	12.469
Spaper	8.416	8.234
Spark	13.226	12.163
Sweet	14.814	15.263
Wine	7.304	7.201

where D_t and T_t are as defined in section III.A

The results of the best neural network model for the old method of decomposition and this new method are shown in Table 5. These results were obtained by using neural networks for forecasting Trend and Irregular Component but Winter's method was used for forecasting the seasonality component. Early results indicate a substantial improvement with the new decomposition method.

V. CONCLUSIONS AND FUTURE WORK

We found that decomposition improves forecasts greatly though there is considerable room for improvement. We used a myriad of forecasting techniques – both neural and statistical for each component series. A simple strategy like using the median of all forecasts is about 4% worse on average than the best model while greedy elimination is about 4% better than the best model. No single technique is good across all series or even over every part of a single series. Hence, combining helps minimize the risk caused by an expert making outrageous forecasts.

Our approach in this project is iterative – the results reported here are those of the first iteration. In the second iteration which has begun, we hope to refine many of our practices – the definitions of the decomposed series, neural network training procedures, assignment of rewards to more parsimonious ARIMA models in the fitness function, ranking of experts and more sophisticated combining strategies.

REFERENCES

[1] I.Alon, M. Qi and R.J.Sadowski: "Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods". *Journal of Retailing and Consumer Services* 8, pp. 147-156, 2001.

[2] J. S. Armstrong. "Combining forecasts: The end of the beginning or the beginning of the end?" *International Journal of Forecasting*, 1990.

[3] J. Scott Armstrong, Fred Collopy. "Error Measures for Generalizing about forecasting methods: Empirical comparisons" *International Journal of Forecasting*, 8, pp. 69-80, 1992.

[4] G.Box, G.Jenkins and G.Reinsel "Time Series Analysis – Forecasting and Control" Third Edition, Prentice Hall, 1994, Englewood Cliffs, NJ.

[5] P. J. Brockwell and R. A. Davis. "Introduction to Time Series and Forecasting". Springer, 2001.

[6] S. Chopra and P. Meindl "Supply Chain Management - Strategy, Planning and Operation", Prentice Hall, 2001.

[7] C-W. Chu and G.P.Zhang "A comparative study of linear and non-linear models for aggregate retail sales forecasting". *International Journal of Production Economics,* 2003.

[8] R. T. Clemen. "Combining forecasts: A review and annotated bibliography". *International Journal of Forecasting*, 1989.

[9] Paulo Cortez, Miguel Rocha, Jose Neves, "Genetic and Evolutionary Algorithms for Time Series Forecasting", *Proc. of 14th International conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems*, pp. 393-402, 2001.

[10] P. H. Franses, G. Draisma, "Recognizing changing seasonal patterns using artificial neural networks". *Journal of Econometrics* 81, pp. 273-280, 1997.

[11] G.Gangadharan, L.Khandelwal, and B.L.Menezes, "Forecasting in Centralized Supply Chain Management", *Smart E-Business International Workshop* (part of First Indian International Conference on Artificial Intelligence), Hyderabad, pp. 95-107, Dec. 20, 2003.

[12] D.E.Goldberg, Genetic algorithms in search, optimization, and machine learning. Addison Wesley, 1989.

[13] Aydin Gurel. "Feed forward neural networks – A Java implementation v2.0". Website: http://aydingurel.brinkster.net/neural/

[14] L.Krajewski and L.Ritzman, "Operations Management – Strategy and Analysis". 5th Edn, Addison Wesley Publishing Company, 1999.

[15] R.Pindyck and D.L.Rubinfeld, "Econometric Models and Economic Forecasts", 4th Edition McGraw Hill International, 1998.

[16] Min Qi G. Peter Zhang. "Neural network forecasting for seasonal and trend time series". *European Journal of Operational Research*, pp. 501–514, 2005.

[17] M. Nelson, T. Hill, T. Remus, M. O'Connor, "Time series forecasting using NNs: Should the data be deseasonalized first?". *Journal of forecasting* 18, pp. 359-367, 1999.

[18] Warren S. Sarle, Neural Network FAQ Part part 2 (of 7). website: ftp://ftp.sas.com/pub/neural/FAQ2.html

[19] R.Snyder, A.Koehler, J.K.Ord, "Forecasting for inventory control with exponential smoothing". *International Journal of Forecasting*, 18, pp. 5-18, 2002.

[20] L.Tao, "Shifting Paradigms with the Application Service Provider Model," *IEEE Computer*, pp. 32-39, 2001.

[21] Frank M. Thiesing and Oliver Vornberger. "Forecasting sales using neural networks". *Proceedings of the International Conference on Computational Intelligence, Theory and Applications*, pp. 321–328, April 1997.

[22] Time Series Data Library. Website: http://www-personal.buseco.monash.edu.au/~hyndman/TSDL/

[23] Francesco Virili and Bernd Freisleben. "Nonstationarity and data preprocessing for neural network predictions of an economic time series". *International Joint Conference on Neural Networks*, 5:129–134, July 2000.

[24] Matthew Wall, Massachusetts Institute of Technology, GAlib, A C++ Library of Genetic Algorithm Components, http://lancet.mit.edu/ga/

[25] K. Walsh, "Analyzing the Application ASP Concept: Technologies, Economies and Strategies," *Communications of the ACM,* Vol. 46, No. 8, (2003) 103-107.

[26] G.Zhang, B.E.Patuwo and M.Y.Hu (1998) "Forecasting with artificial neural networks: The state of the art", *International Journal of Forecasting*, 14, pp. 35-62.

[27] H. Zou and Y. Yang. "Combining time series models for forecasting". *International Journal of Forecasting*, 2004.