

# A hybrid SARIMA wavelet transform method for sales forecasting

Tsan-Ming Choi <sup>\*</sup>, Yong Yu, Kin-Fan Au

Business Division, Institute of Textiles & Clothing, Faculty of Applied Science and Textiles, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

## ARTICLE INFO

### Article history:

Received 14 January 2010

Received in revised form 6 August 2010

Accepted 5 December 2010

Available online 15 December 2010

### Keywords:

Sales forecasting

SARIMA model

Wavelet transform

Decision support system

## ABSTRACT

Time series forecasting, as an important tool in many decision support systems, has been extensively studied and applied for sales forecasting over the past few decades. There are many well-established and widely-adopted forecasting methods such as linear extrapolation and SARIMA. However, their performance is far from perfect and it is especially true when the sales pattern is highly volatile. In this paper, we propose a hybrid forecasting scheme which combines the classic SARIMA method and wavelet transform (SW). We compare the performance of SW with (i) pure SARIMA, (ii) a forecasting scheme based on linear extrapolation with seasonal adjustment (CSD + LESA), and (iii) evolutionary neural networks (ENN). We illustrate the significance of SW and establish the conditions that SW outperforms pure SARIMA and CSD + LESA. We further study the time series features which influence the forecasting accuracy, and we propose a method for conducting sales forecasting based on the features of the given sales time series. Experiments are conducted by using real sales data, hypothetical data, and publicly available data sets. We believe that the proposed hybrid method is highly applicable for forecasting sales in the industry.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

Sales forecasting plays a very important role in business operations and it is essentially significant in businesses with a highly volatile sales pattern such as fashion apparels and consumer electronics. A proper selection of models for forecasting sales is one of the major research efforts over the past few decades [16].<sup>1</sup> Although traditional methodologies of sales forecasting such as auto-regression (AR), and integrated moving average model have been proven to be effective in many decision support applications [2,3,9,21], they still have certain shortcomings, and fail in an environment where the sales are more influenced by exogenous variables such as size, price, colour, climatic data, effect of media, and price changes. Recently, more and more research efforts have been spent on exploring hybrid forecasting models such as a remarkable method based on clustering and decision trees analysis [22] and some others [1,5,6].

It is well-known that artificial intelligence (AI) models [5,6] have more flexibility and can be used to estimate the non-linear relationship. Therefore, many researchers have applied AI models for forecasting problems. For example, in Ref. [11], the authors propose the use of artificial neural networks (ANN) approach to forecast women's apparel sales. Their experimental results show that ANN approach outperforms the other two conventional models, namely, single seasonal exponential smoothing model and Winters'

three parameter models. In Ref. [6], the authors propose a hybrid model for stock price forecasting by integrating multiple regressions, back propagation (BP) neural network and an autoregressive integrated moving average model. In Ref. [23], the performance of a heterogeneous mixture of neural network algorithms for predicting the exchange-traded fund is investigated. A genetic algorithm is utilized to find the best mixture of neural networks, the topology of individual networks in the ensemble, and to determine the features set. In Ref. [7], the authors develop an evolving neural network (ENN) forecasting model by integrating genetic algorithms (GA) and ANN for sales forecasting in the electronics industry. Their experimental result shows that the performance of ENN is superior to various traditional statistical models and the back propagation network. In Ref. [8], the authors propose a hybrid system which combines the self-organizing map (SOM) of neural network with case-based reasoning (CBR) model for sales forecasting of newly released books. In Ref. [1], the authors present a hybrid intelligent system combining ARIMA models and neural networks for demand forecasting. Despite the popularity of the ANN related methods, the performance of all these ANN based techniques depends highly on the relationship between the patterns used in training the networks and the expected forecasting patterns. If the diversity or the inconsistency between the training patterns and the expected forecast patterns is significant, the forecast errors of the ANN technique will tend to be relatively high [15]. Moreover, the ANN related methods for sales forecasting in the item-level usually require a very long computational time which makes this method less appealing to many real-world applications in the industry such as fashion apparels because there are thousands of SKUs (products with a unique design, colour, and size). Alternatively, it is argued that sales

<sup>\*</sup> Corresponding author. Tel.: +852 27666450; fax: +852 27731432.

E-mail address: [tcjason@inet.polyu.edu.hk](mailto:tcjason@inet.polyu.edu.hk) (T.-M. Choi).

<sup>1</sup> Please refer to Ref. [16] for an interesting study on the agent-based demand forecasting in a supply chain.

amount is affected by several nonlinear exogenous variables and periodical components, which can be illustrated by different mathematical models. To improve the forecasting precision, the optimal forecasting model should make a wise use of these different data components. Thus, we can decompose the given time series into component time series with different characteristics, and based on the features of the components, we apply the appropriate type of forecasting scheme and then we convert the forecasting of the time-series at component levels back to the original time series level. One way of conducting this decomposition is via the wavelet transform.

In recent years, the wavelet transform has yielded encouraging results in multi-resolution analysis as well as many other functions [10,24] and it has also been combined with fuzzy logics for different applications [5]. The wavelet transform allows the decomposition of a signal into different levels of resolution scales, which means that we can extract the required data components. To be specific, by using discrete wavelet transformation (DWT), we can decompose the sales data time series into several scales, where both the coarse and fine parts of the data are obtained. The coarse (approximated) scales reveal the trend, while the fine (detailed) scales tend to be related to seasonal influences, and exogenous variables such as size, price, colour, climatic data, effect of media, and price changes. Afterwards, the multiple regression model and SARIMA forecast model can be adopted for forecasting in the approximated scale and the detailed scales, respectively. Finally, the inverse wavelet transform is used to generate the final sales forecasting time series. In Refs. [5] and [24] the wavelet transformation schemes are combined with AI techniques to do forecasting for the financial time series. Such hybrid schemes show significant advantages over the traditional AI models. This indicates that the wavelet can be a promising tool in the decomposition of the time series. As a remark, many researchers have studied all kinds of forecasting models while there is no well-agreed rule on specifying which one is better under which condition. In light of the complex nature of many highly volatile sales patterns, many researchers are in favour of the sophisticated models such as SARIMA and ENN, while many others are in favour of simple models such as simple linear inherited models [14,18], which are also supported by many applications in the industry. However, both groups of researchers could consider the use of a certain decomposition scheme to further enhance their original methods and hybrid models are both interesting and promising in the area of forecasting. In this paper, we concentrate on studying a hybrid model which combines the classic SARIMA method and wavelet transform (SW). We compare SW with two other models, namely, the pure SARIMA method and the Classical Seasonal Decomposition with Linear Extrapolation with Seasonal Adjustment method (CSD + LESA). The forecasting schemes are tested for their forecasting performance with artificially composed data sets (to study the impacts brought by different features of the sales pattern), real sales data sets, and publicly available data sets. We find that these methods perform differently under different conditions. For example, our experiments demonstrate that the SW method outperforms all the others when the data sets are associated with a highly volatile pattern, the CSD + LESA model performs better than the others with data sets which exhibit strong white noise, while the SARIMA model outperforms the above two when seasonality in the data set is very strong which dominates the time series, and the decomposition methods are no better than SARIMA in modelling the seasonality. After obtaining these results, we further develop a decision model (Fig. 13) which helps us select the most appropriate method for each forecasting task in order to achieve the best sales forecasting performance. The paper develops as follows. Section 2 presents multi-scale decomposition for the sales time series using the wavelet transform. Section 3 provides a detailed specification of the forecasting procedure, and the forecasting results based on various models are presented. Section 4 discusses the comparison between the methods. The concluding remarks are provided in Section 5.

## 2. Multi-scale decomposition using wavelet transform for forecasting

To illustrate the decomposition scheme using the wavelet transform, we start by using the sales time series from an apparel company.

### 2.1. Fashion sales data

In this section, the fashion sales time series of several retail stores of an apparel company have been studied, the data collected covers the sales data for various types of apparels from 1999 to 2003, and sales data points are collected for each day for every class for every store. Daily sales curve over a period from 4 July 2002 to 28 August 2002 is given in Fig. 1. The time series in Fig. 1 is presumed to have weekly seasonality as nearly all daily sales data sets do. However, as what we have observed, the weekly trend is not so clear in the curve. To be systematic, trends and seasonality of the sales data are further supported by calculating auto correlations functions (ACFs). ACF is an important tool for discerning time series patterns, ACF for a given time lag  $k$  is given by Eq. (1):

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}, \quad (1)$$

where  $Y_i$  ( $i = 1, \dots, n$ ) is a time series, where  $\bar{Y}$  is the mean value of the time series,  $k$  is the time lag. Although this kind of analysis is not a necessity for the wavelet-transform based forecasting models, it helps us understand the trends and seasonality of a given sales data set. Moreover, as we will see later, by checking ACF, we can know more about when SW outperforms CSD + LESA.

### 2.2. Multi-scale decomposition of sales time series

For a given sales time series, in general, it is difficult to describe its components very accurately. However, via a wavelet transform, it becomes easier for us to observe these factors and make the forecasting task easier. As the SARIMA model will be employed in studying the time series later, the historical sales data is expected to have at least 50 points to yield a sound analysis. In this section, the historical sales data of an apparel company in 8 weeks (56 out of 66 points is used) will be decomposed in different scales by the discrete wavelet transform.

There are two criteria for the selection of the mother wavelet [12]. Firstly, the shape and the mathematical expression of the wavelet must be selected correctly so that the physical interpretation of the wavelet coefficients is easy. Secondly, the chosen wavelet must allow a fast computation of the required wavelet coefficients. In this paper, the discrete approximation of Meyer wavelet (D-Meyer) is hence selected as it is a fast algorithm which also supports discrete transformation [4].

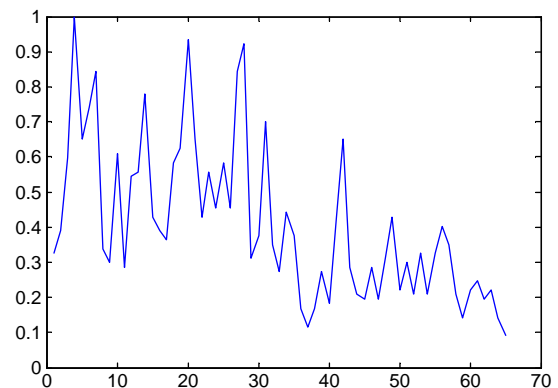


Fig. 1. Daily sale of a specific apparel product.

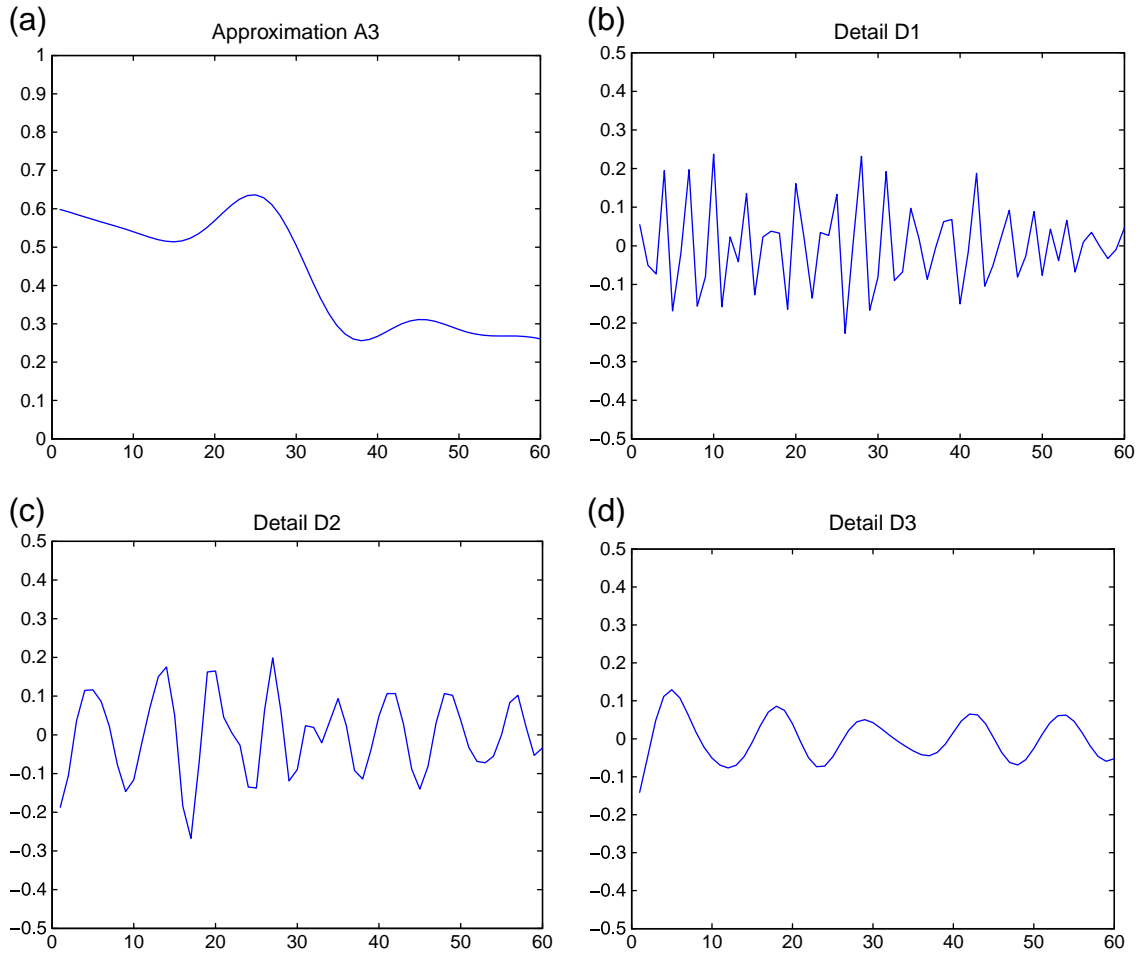


Fig. 2. D-Meyer three-level wavelet decomposition.

The results in different scales are shown in Fig. 2. Fig. 2 illustrates the three-level decomposition using D-Meyer. We can see from Fig. 2(a) that the detailed scale mainly contains the trend component, Fig. 2(b–d) represent most of the weekly periodic components and stochastic components. Time series forecasting can be produced by forecasting on the trend and seasonal components separately. We thus propose that in applying the wavelet decomposition for the sake of forecasting, it is not necessary to decompose a given time series into too many components. We will address this issue more in the next section.

### 3. Hybrid forecasting based on wavelet transform and SARIMA (SW)

Sales forecasting, especially in the industry such as fashion, is a challenging task because many endogenous as well as exogenous variables, e.g. size, price, colour, climatic data, effect of media, etc., which affect sales, are involved. In the above section, by using discrete wavelet transform (DWT), we decompose the given sales time series into several scales and the detailed and approximated components are obtained. In order to improve the forecasting precision, the optimal forecast models should be tailored for forecasting the components in different scales.

#### 3.1. Data processing in the approximated scale

The approximated scale presents mainly a trend component. As our proposed scheme mainly focuses on the sales forecasting for highly volatile patterns and this randomness makes the components in the approximate scales exhibit low seasonality. In our proposed method, we will adopt the ARIMA model for forecasting the components in the approximate scale. This approach is supported

by our preliminary experiments conducted during the development of the proposed model with the use of real sales data.

#### 3.2. Data processing in the detailed scale

The detailed scale contains mainly the seasonal and stochastic components and should be analyzed by SARIMA. SARIMA is the most popular linear model for forecasting seasonal time series. It has achieved great success in both academic research and industrial applications during the last three decades. A time series  $\{Z_t | 1, 2, \dots, k\}$  is generated by SARIMA  $(p, d, q) (P, D, Q)_s$  process of Box and Jenkins time series model [3] if

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t, \quad (2)$$

where  $p, d, q, P, D, Q$  are integers,  $s$  is the season length;

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p,$$

$$\Phi_P(B^s) = 1 - \phi_s B^s - \phi_{2s} B^{2s} - \dots - \phi_{Ps} B^{Ps},$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q, \text{ and}$$

$$\Theta_Q(B^s) = 1 - \theta_s B^s - \theta_{2s} B^{2s} - \dots - \theta_{Qs} B^{Qs},$$

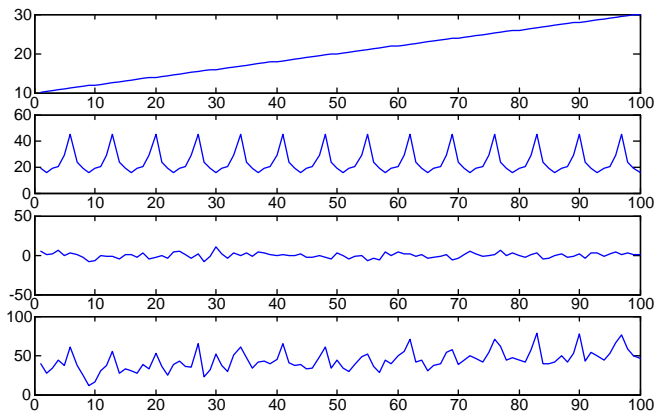


Fig. 3. The composition of the hypothetical data.

are polynomials in  $B$ , where  $B$  is the backward shift operator,  $\varepsilon_t$  is the estimated residual at time  $t$ ,  $d$  is the number of regular differences,  $D$  is the number of seasonal differences,  $Z_t$  denotes the observed value at time  $t$ ,  $t = 1, 2, \dots, k$ .

Fitting a SARIMA model to data involves the following four-step iterative cycles: (a) identify the SARIMA ( $p, d, q$ ) ( $P, D, Q$ )'s structure; (b) estimate unknown parameters; (c) perform goodness-of-fit tests on the estimated residuals; and (d) forecast future outcomes based on the known data. The fitting of SARIMA models is a challenging task (an example can be found in Ref. [19]), and it involves human judgment on the ACF charts. As we concentrate on the effect of the wavelet decomposition, to avoid the human judgment influences of SARIMA, we use SARIMA (1,1,0) for all the SARIMA models.

Every  $\varepsilon_t$  is independently and identically distributed as a normal random variable with mean 0 and constant variance  $\sigma^2$ . The roots of  $\phi_p(Z) = 0$  and  $\theta_q(Z) = 0$  all lie outside the unit circle. In addition, it is suggested by Ref. [3] that a minimum of 50 (preferably 100) observations should be used for the SARIMA model.

### 3.3. Evaluating indices

A good forecasting method needs to take into consideration the degree of accuracy. A popular measure, known as the Mean Absolute Percentage Error (MAPE), is presented as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|F_i - A_i|}{A_i}, \quad (3)$$

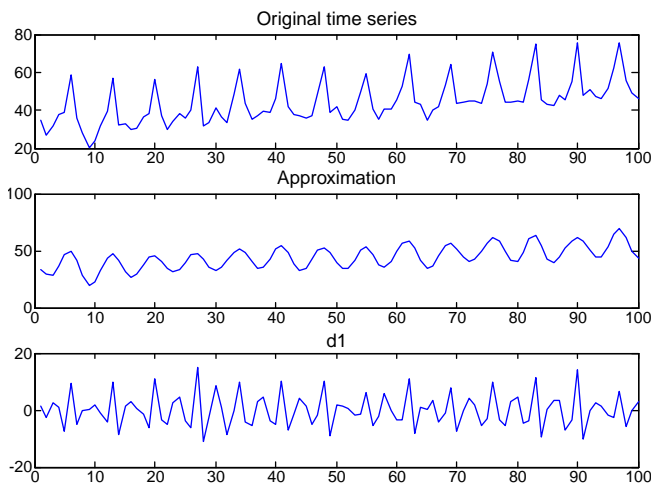
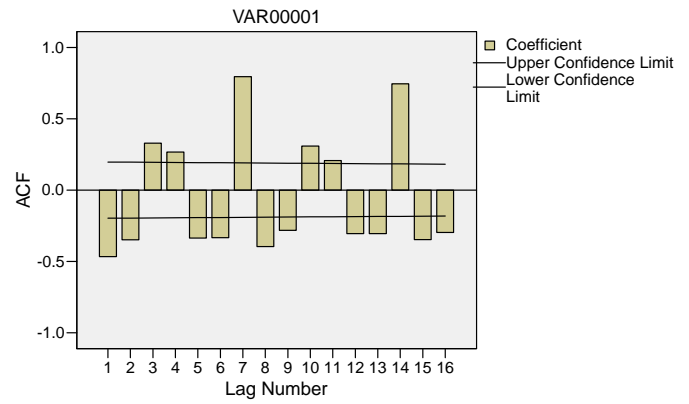


Fig. 4. The 1-level wavelet decomposition.

Fig. 5. The autocorrelation factor (ACF) of  $d_1$ .

where  $F_i$  is the forecast value for period  $i$ ,  $A_i$  is the actual value for period  $i$ ,  $n$  is the total number of periods. As MAPE is good at describing a relative error, which is especially suitable when comparing between different data sets, it is used as the performance index in this paper.

### 3.4. Wavelet-based forecasting and its parameter selection procedure

By using inverse discrete wavelet transform (IDWT), the sales forecasting results are then obtained by the reconstruction of the forecasting data in different scales. As we discussed above, when the wavelet is used in the time series decomposition, the number of decomposition levels has to be determined first. Sales time series is hypothetically composed of trend and seasonality. As a result, the major goal of decomposition is to reveal these two major factors. In wavelet decomposition, as the purpose of the decomposition procedure is to find the underlying trend and seasonality components, it is not necessary to decompose a time series into too many components. However, a simple one level decomposition often cannot identify the seasonality factors precisely. The experiments on theoretically generated data sets (as in Fig. 3) are hence used to study this feature. In Fig. 4, a 1-level wavelet decomposition is performed as the first step. The two components, high-frequency component  $d_1$  and approximation  $a$ , are retrieved. Component  $d_1$  shows the weekly seasonality when we examine its autocorrelation factor (ACF) chart in Fig. 5, since the ACF is significantly greater than the confidence limits at lag 7 and lag 14, we consider the seasonality in  $d_1$  as strong (please refer to Ref. [17] for details about the identification of seasonality in an ACF chart).

When  $d_1$  with seasonality is found from the 1-level decomposition, the approximation  $a$  is not close to the hypothetic trend, and it also has

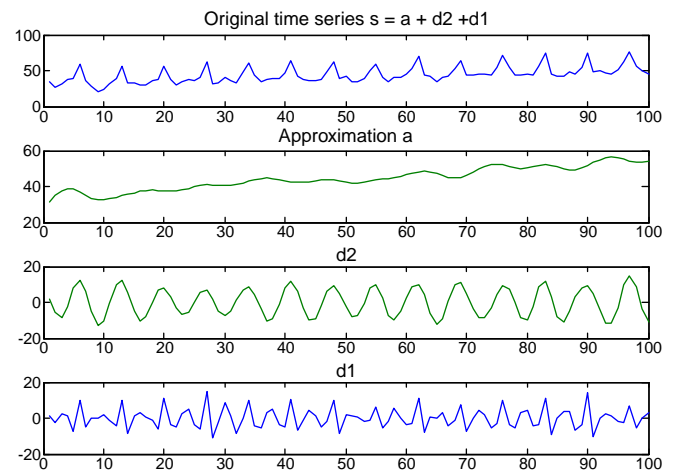
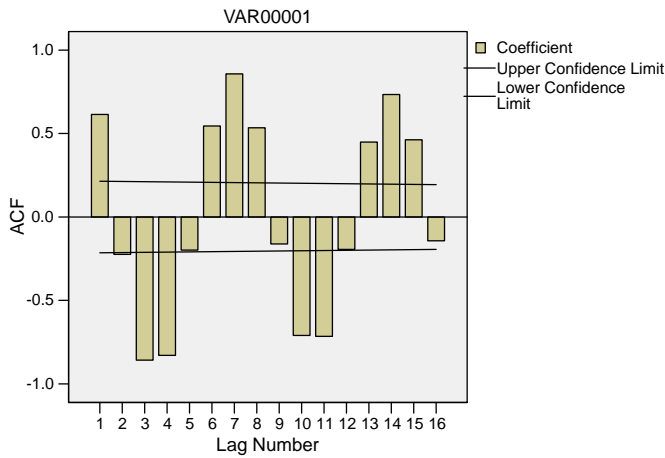
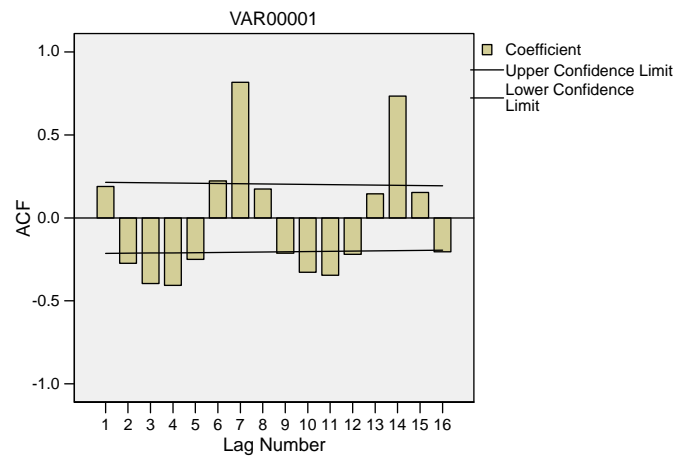


Fig. 6. The 2-level wavelet decomposition.

Fig. 7. The autocorrelation factor (ACF) of  $d_2$ .Fig. 9. The autocorrelation factor (ACF) of  $d_1 + d_2$ .

some seasonality cycle in it. As the second step, a 2-level decomposition is performed on the time series. Fig. 6 depicts the decomposition and Fig. 7 depicts the ACF of  $d_2$ . Again, component  $d_2$  has a strong weekly seasonality. As both  $d_1$  and  $d_2$  have weak seasonality, we add these two up as a single seasonal factor as shown in Fig. 8. The seasonality of the add-up factor is stronger than each single one of  $d_1$  and  $d_2$ , which can be observed in the ACF chart in Fig. 9.

In the third step, we further decompose the time series into 3 levels, as shown in Fig. 10, we can no longer find the weekly seasonality by observing the ACF in Fig. 11. Though there is a significant value at lag 7, it is not the most significant one among the nearby values and the significance does not repeat at lag 14. When comparing the 3-level wavelet decomposed time series with the original components in Fig. 12, the decomposed components are quite close to the original ones with trend, seasonal cycle and white noise. As the white noise component is not used in the forecasting, the 2-level decomposition is the right level for this situation. In practice, there are often cases where no seasonality is found during the decomposition of any level, we thus suggest using the SARIMA model for the original time series under this condition. There are also cases when a component may have no seasonality, but its fluctuation is

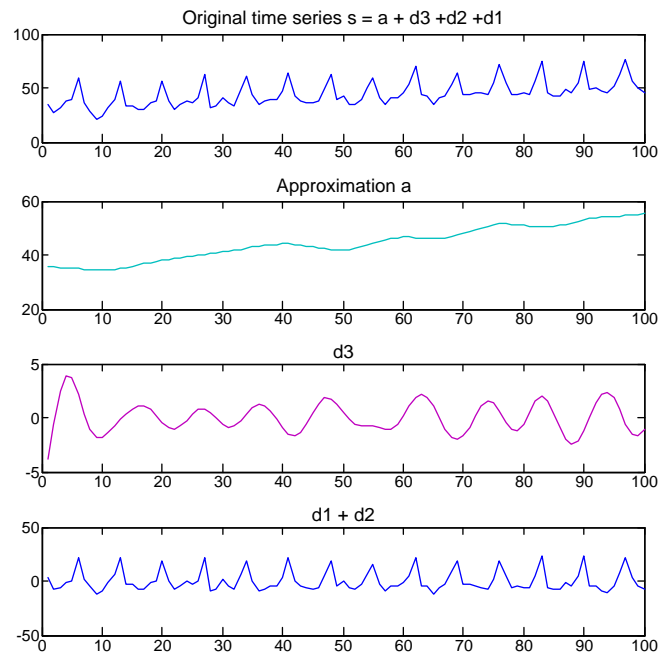
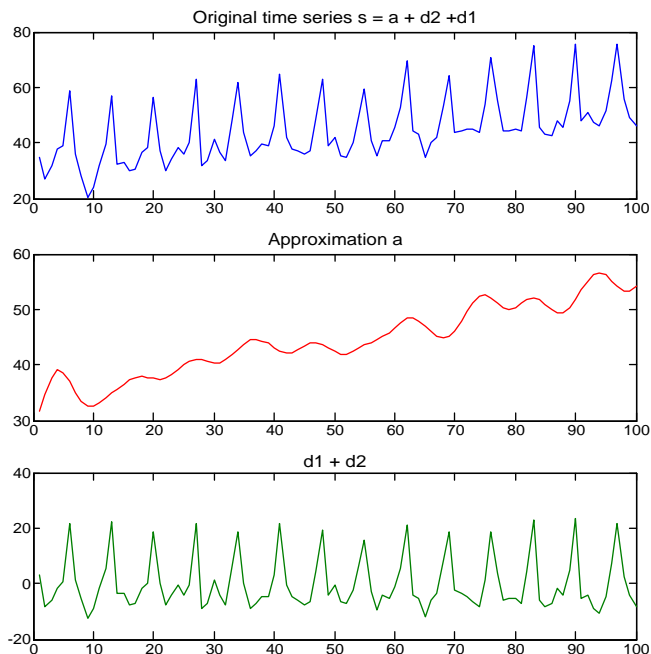
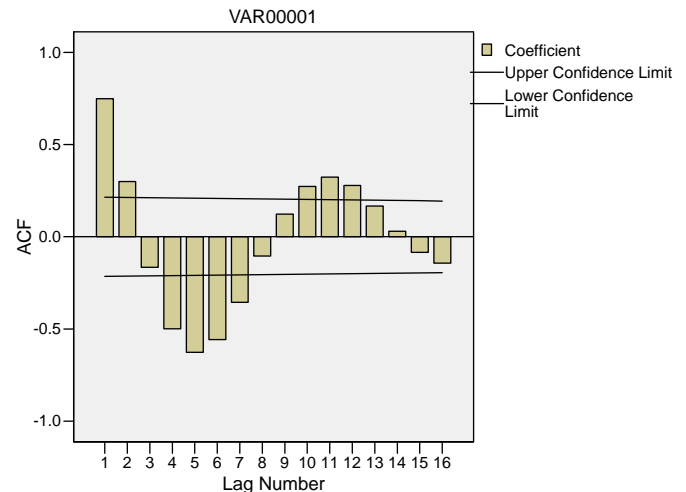
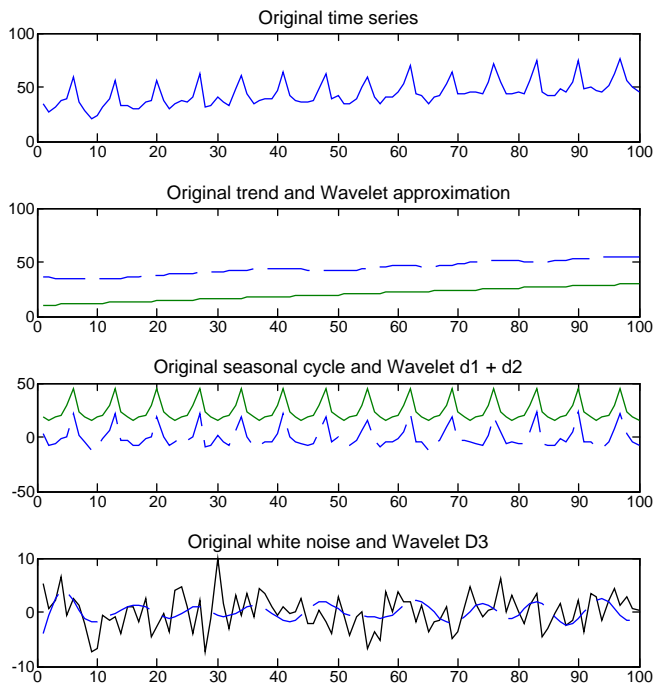


Fig. 10. The 3-level wavelet decomposition.

Fig. 8. The 2-level wavelet decomposition with combining seasonality  $d_1 + d_2$ .Fig. 11. The autocorrelation factor (ACF) of  $d_3$ .





**Fig. 12.** Comparison between the original components and the wavelet decomposed ones. (The solid lines representing the original components and the dashed lines representing the decomposed ones).

significant compared to the original time series so that it cannot be omitted in the forecasting process. Fig. 13 illustrates a complete procedure to deal with these situations.

As we will employ the SARIMA method in forecasting for the seasonal component, a component with a strong seasonality apparently benefits the forecasting. Table 1 further gives a forecasting comparison between the decompositions. This result shows that when a highly volatile pattern exhibits in the decomposed component, the forecasting result is expectedly better.

As in the CSD approach, the other decomposed components, such as  $d_3$  in Fig. 10, are considered to be white noise, since they cannot benefit the forecasting, they will not be included in estimating and reconstructing the final forecasting. Unlike the CSD approach, although components in the wavelet decomposition may have no seasonality, the amplitude of such components can be quite significant comparing to the original time series. Under this condition, omitting the components can cause inaccurate forecasting. An example of this phenomenon is given in Section 4.3, and based on this experiment, a threshold is set as follows,

$$T = | \max(d_n) - \min(d_n) | / | \max(S) - \min(S) |,$$

where  $S$  is the original time series, and  $d_n$  the decomposed series. According to the experiment in Section 4.3, if  $T > 50\%$ , the decomposition approach is not better than using the SARIMA method and hence SARIMA is suggested under this condition. Based on the previous analysis we develop an algorithm to systematically perform this forecasting procedure. The procedure is shown in Fig. 13. The algorithm starts with the decomposition level setting to be 1. The decomposition is conducted as in Fig. 5 and the  $d_1$  component's seasonality is examined by its autocorrelation factor, where  $r_k$  is defined in (1). Usually the ACF analysis involves human judgement of the ACF chart as the Box–Jenkins method suggests. To simplify this procedure, we use automatic judgement. When the ACF at the first lag of season length (such as 7 for weekly seasonality) is greater than the corresponding critical value and dominates the other nearby significant ACFs, we consider the underlying series to have strong seasonality. If it is significant but not

dominating, we further observe the ACF as the next lag of season length (14 for weekly seasonality); if that ACF is significant, we also consider the underlying series to have strong seasonality. If the seasonality of  $d_1$  is strong, the stopping criterion will be met and the algorithm stops. If the seasonality of  $d_1$  is not strong enough, the algorithm repeats with the level increased by 1 as shown in Fig. 7. Like the CSD method, we also presume a sales time series to be composed of trend, seasonality, and irregular components, and the wavelet decomposition is to reveal these underlying 3 components, especially the first 2 components for forecasting. Apparently, too many levels of the wavelet decomposition do not really help with the accuracy of forecasting because when the decomposition level is increased, the newly generated components are of minor importance. However, as we have observed in Fig. 10, combining two seasonal components with seasonality helps to increase the accuracy of forecasting, this indicates that to decompose the time series into 4 or 5 levels may be helpful. Thus we set an upper bound as 5 levels which mean that the algorithm will stop even if there is no clear seasonality found under the 5 level decomposition. When this situation happens, we recommend utilizing other forecasting methods.

#### 4. Comparisons among different methods

In this section, we study the forecasting performance of various methods, which include the pure SARIMA, SW, CSD + LESA. All methods are used in forecasting with hypothetical data and real sales data in fashion business. Some public data sets are also used in the experiments.

##### 4.1. Comparison: hypothetical data

The hypothetical data set is composed of the trend, seasonality and white noise components, which are also used in the last section, as shown in Fig. 3. When the CSD + LESA method is employed in analyzing this hypothetical data set, the time series are first decomposed into seasonal adjusted series and seasonal adjust factor. The time series are found almost perfectly decomposed into the trend and seasonal factor, as the LESA method in nature presumes a perfect seasonality which exactly matches with this hypothetical condition.

When the SW method is employed on the hypothetical data, the algorithm in Section 3 is employed to determine the levels to be decomposed, in this situation, the algorithm stops with a level of 3, Fig. 10 depicts the decomposition result. After the decomposition, forecasting is carried out on all methods, and the results are shown in Table 2.

As we can observe in Table 2, the forecasting performances of CSD and SW are quite similar, either in the overall scale or weekly scale. This result is different from Ref. [14] which claims that the simple approach LESA is better than the other sophisticated methods. These two methods are both better than the SARIMA model, which means that the decomposition benefits the forecasting in this situation.

It is natural that features of the time series greatly influence the forecasting result. These features are essentially the statistical factors of the trend, seasonality and irregular components. In this section, we carry out two experiments on the hypothetical data with different parameters, namely the hypothetical data with the white noise variance increased to 500, and the hypothetical data with the seasonal variance increased to 600. Table 3 gives the result of the comparison, in which “Column As” list the original result where the white noise variance is 10, as in Table 1, “Column Bs” list the result when the variance of the white noise is increased to 500 and “Column Cs” list the result when the variance of the seasonal cycle is increased to 600. While all methods provide worse forecasting because of the increased variance of the white noise (as seen by comparing “Column Bs” in Table 3), it is interesting to notice that the CSD + LESA method provides a little bit better forecasting than SW. However, as seen in the “Column Cs”, when the variance of the seasonal cycle increases,

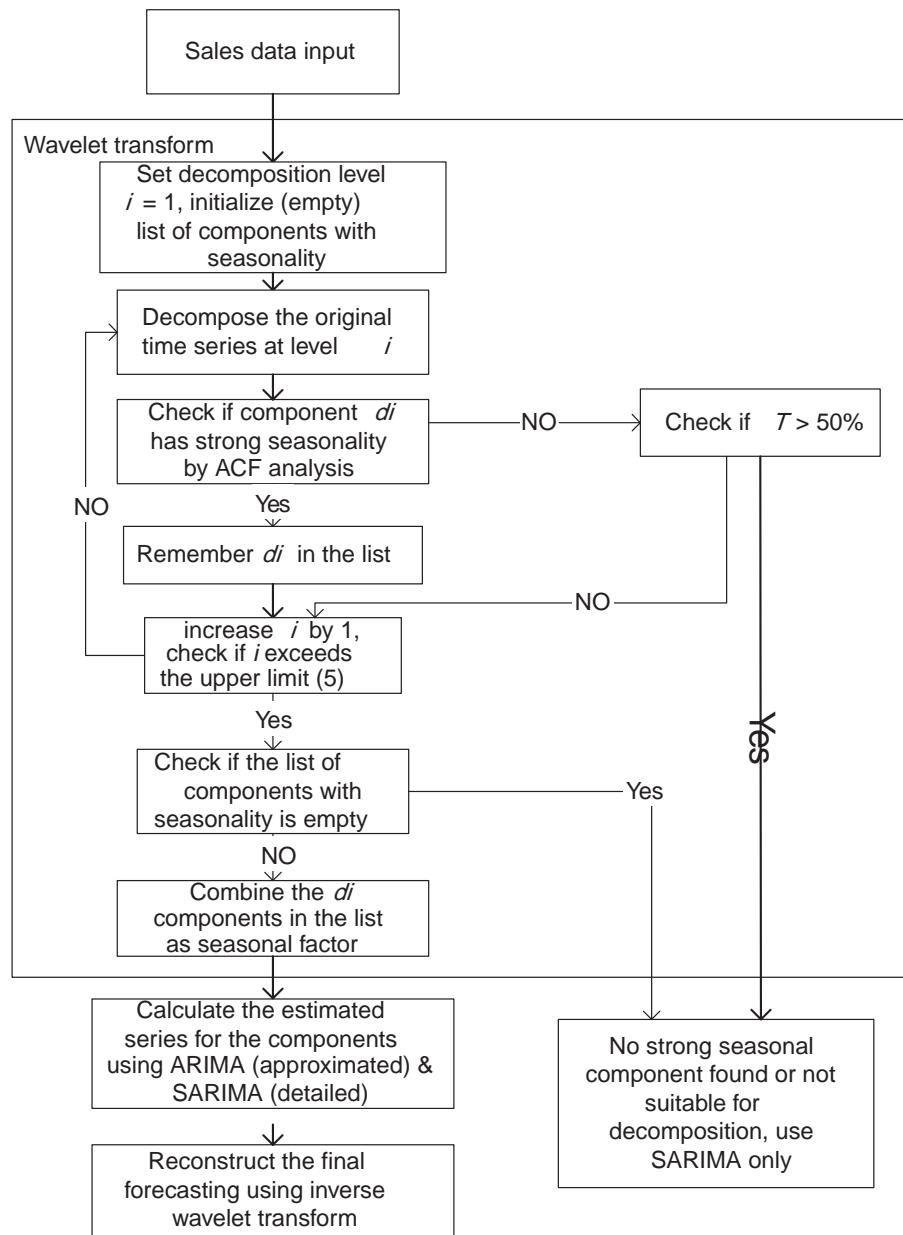


Fig. 13. The procedure of sales forecasting using the hybrid SARIMA wavelet transform method. (As a remark, this algorithm is not depicted with the standard flow chart notation).

SW performs much better than the CSD + LESA method. Table 4 summarizes the above findings.

Although the CSD + LESA method performs slightly better than SW when the white noise variance is big, it is no longer better when the

variance of the seasonal cycle increases which governs the features of the time series. The SARIMA model still produces the worst forecasting results compared to SW and CSD + LESA when the seasonal features of the time series are changed. In the following

Table 1  
Forecasting comparison of decomposition in different levels.

Week	1 level decomposition (a, d <sub>1</sub> )		2 levels decomposition (a, d <sub>2</sub> )		2 levels decomposition (a, d <sub>1</sub> + d <sub>2</sub> )		3 levels decomposition (a, d <sub>1</sub> + d <sub>2</sub> , d <sub>3</sub> )	
	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE
Week 8	0.06	8.4	0.07	15.2	0.06	9.7	0.07	13.6
Week 9	0.11	42.1	0.13	59.2	0.10	28.6	0.11	50.1
Week 10	0.07	15.3	0.08	21.1	0.06	8.3	0.09	45.3
Week 11	0.10	38.3	0.09	50.7	0.09	33.1	0.12	55.7
Week 12	0.05	10.8	0.11	52.8	0.04	6.5	0.03	2.8
Week 13	0.04	5.5	0.11	47.6	0.04	4.3	0.07	10.6
Overall MAPE	0.072	20.1	0.101	41.1	0.066	15.1	0.082	29.7

**Table 2**  
Forecasting comparison on hypothetical data.

Week	SARIMA		ARIMA_SARIMA_WAVELET (trend + season)		CSD + LESA	
	MAPE	MSE	MAPE	MSE	MAPE	MSE
Week 8	0.058	12.4	0.062	21.7	0.073	55.1
Week 9	0.120	33.1	0.060	15.4	0.052	19.6
Week 10	0.056	15.2	0.057	18.1	0.061	42.1
Week 11	0.094	56.2	0.043	7.7	0.057	10.5
Week 12	0.062	30.8	0.042	6.8	0.030	4.3
Week 13	0.033	9.6	0.043	8.9	0.041	8.5
Overall MAPE	0.071	26.2	0.051	13.1	0.052	23.4

**Table 3**  
Forecasting comparison with different parameters.

Week	SARIMA			SW			CSD + LESA		
	A	B	C	A	B	C	A	B	C
Week 8	0.058	1.11	0.068	0.062	1.18	0.067	0.073	1.11	0.099
Week 9	0.120	1.97	0.138	0.060	1.43	0.073	0.052	1.17	0.054
Week 10	0.056	0.68	0.073	0.057	0.78	0.067	0.061	0.77	0.083
Week 11	0.094	0.40	0.109	0.043	0.21	0.053	0.057	0.31	0.078
Week 12	0.062	0.46	0.066	0.042	0.44	0.045	0.030	0.33	0.059
Week 13	0.033	0.51	0.039	0.043	0.44	0.047	0.041	0.42	0.060
Overall MAPE	0.071	0.85	0.082	0.051	0.74	0.058	0.052	0.69	0.072

A: forecasting on original data (white noise's variance = 10); B: forecasting on data with the variance of white noise component increased to 500; C: forecasting on data with the variance of seasonal cycle component increased to 600.

section, we further study the forecasting performance of these methods with real sales data.

#### 4.2. Comparison: real fashion sales data

The real fashion sales data set adopted here has been discussed in Section 2. Without theoretically perfect seasonality, a seasonality factor is retrieved by the CSD method. As for the decomposition of SW, the decomposition level is increased to be 2 compared to the decomposition on the hypothetical data with perfect seasonality. The final result is presented in Table 5, which shows that the SW

method dominates the CSD + LESA method in the forecasting of real fashion sales data. While the SARIMA method produces slightly better forecasting than the CSD + LESA method, it still cannot compete with the SW method.

#### 4.3. Comparison: public available sales data

A publicly available data set, *Monthly Australian sales of sparkling wine* from Time Series Data Library <http://www-personal.buseco.monash.edu.au/~hyndman/TSDL/>, is used in this analysis. The time series is depicted in Fig. 14. The autocorrelation analysis over the time series shows a very significant ACF at lag 12, which is more than double of the critical value. This indicates a strong yearly seasonality in the monthly Australian sales of the sparkling wine time series.

Applying the method described in Section 2, two components with seasonality ( $d_1, d_2$ ) are found, while  $d_3$  has no obvious seasonality, the threshold of it is  $T = 55\%$ . So, a 3-level decomposition is performed on

**Table 4**  
Forecasting variation with different parameters.

Components changes	SW	CSD + LESA
Original	✓	O
$\sigma_{WN}$ increased	O	✓
$\sigma_{SC}$ increased	✓	O

$\sigma_{WN}$ : variance of white noise,  $\sigma_{SC}$ : variance of seasonal cycle.

✓: better in the comparison of forecasting accuracy, O: worse in the comparison of forecasting accuracy.

**Table 5**  
Forecasting comparison on real fashion sales data.

Week	SARIMA		SW (trend + seasonality)		CSD + LESA	
	MAPE	MAE	MAPE	MAE	MAPE	MAE
Week 8	0.30	0.031	0.17	0.019	0.26	0.025
Week 9	1.11	0.113	0.93	0.101	1.14	0.112
Week 10	0.32	0.033	0.26	0.033	0.39	0.040
Week 11	0.37	0.035	0.35	0.034	0.45	0.045
Week 12	0.27	0.027	0.30	0.032	0.31	0.032
Week 13	0.69	0.070	0.62	0.063	0.65	0.063
Overall MAPE	0.51	0.052	0.44	0.047	0.53	0.053

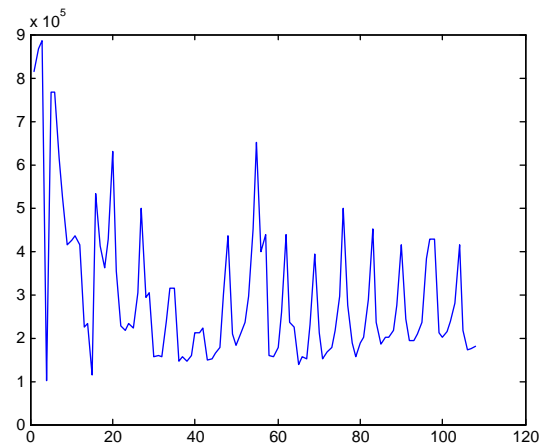


Fig. 14. Monthly Australian sales of sparkling wine.



**Table 6**  
Forecasting comparison in MAPE.

Year	SARIMA		CSD + LESA		SW ( $a, d_1 + d_2$ )	
	MAPE	MSE	MAPE	MSE	MAPE	MSE
Year 5	0.112	173,772	0.225	501,338	0.193	471,156
Year 6	0.108	177,076	0.258	632,895	0.133	268,435
Year 7	0.147	260,421	0.160	284,633	0.189	435,231
Year 8	0.194	419,619	0.452	1,158,342	0.372	828,612
Year 9	0.188	288,082	0.378	868,532	0.347	793,487
Overall MAPE	0.150	263,794	0.290	574,290	0.20	466,153

the time series. The forecasting comparison of SARIMA, SW and CSD + LESA is shown in Table 6. Again, similar to the forecasting results for the real data of fashion sales, SW performs better than the CSD + LESA method. As the simple linear extrapolation inherently presumes the unrealistic perfect linear trend in the time series, the forecasting results of CSD + LESA can be quite poor. In the comparison, the pure SARIMA method is found to be the best one among the three. The reason is that: although  $d_3$  has no seasonality, its  $T$  value is very significant ( $T > 50\%$ ) and hence it should not be omitted in the forecasting process. This implies that the proposed decomposition approach is not effective under this situation, and the pure SARIMA is the preferred method.

Another publicly available time series is ten years data of monthly production of woolen yarn in Australia (Australian Bureau of Statistics), as shown in Fig. 15.

Applying the method described in Section 3, we obtain the decomposition scheme which sets a level to be 3. The forecasting comparison is shown in Table 7. Again the SW method produces better forecasting than the CSD + LESA method in general. The CSD + LESA method produces better forecasting for years 5 and 6, but the situation changes for year 7. Although for year 8, the CSD + LESA method gives better forecasting compared to SW, the difference is actually very marginal; and for year 9, the SW method is again significantly better than the CSD + LESA method. As we can observe from the time series, the seasonal features have changed a lot between year 6 and year 7 (at month 72), and the approximation component which represents the trend in the time series also changes. This indicates that the CSD + LESA method cannot adapt to the changes of the underlying components in a time series. In this experiment, the SARIMA model produces almost the same forecasting accuracy as the SW model but does not exceed it. The Bayesian Information Criterion (BIC) [20], which is a statistical criterion for the model selection, can be used as a reference in the selection. The BIC is given by:  $BIC = n \ln(SSE/n) + k \ln(n)$ , where  $n$  is the number of

**Table 7**  
Forecasting comparison in MAPE.

Year	SARIMA			CSD + LESA			SW ( $a, d_2 + d_3$ )		
	MAPE	MSE	BIC	MAPE	MSE	BIC	MAPE	MSE	BIC
Year 5	0.050	18,543.7	595.7	0.049	20,149.1	0.054	21,521.3	617.3	
Year 6	0.048	25,165.2	616.9	0.031	14,139.1	0.046	26,611.4	631.9	
Year 7	0.050	21,061.6	604.5	0.071	33,511.7	0.055	22,412.2	620.0	
Year 8	0.227	286,908.3	785.4	0.200	250,407.4	0.203	23,1746.5	781.8	
Year 9	0.123	88,303.2	703.8	0.227	267,963.8	0.135	12,7253.8	740.3	
Overall	0.099	87,996.4		0.116	117,234.2	0.098	85,909.1		

observations,  $k$  is the number of free parameters to be estimated,  $SSE$  is the sum of squared errors from the estimated model. The BIC is an increasing function of  $SSE$  and an increasing function of  $k$ . Thus, for any two given estimated models, the model with the lower value of BIC is preferred. The BIC is also given in Table 7, where  $n$  is the number of observations and  $k$  is the number of parameters used in all SARIMA analysis. As shown in Table 7, in most cases, the BIC of the SARIMA model is lower than that of the SW model's. The implications are discussed in Section 5.

#### 4.4. Comparison between SW and other well-established and traditional methods

To study the forecasting performances of the SW method and other traditional forecasting methods, a further comparison is conducted among the SW method, CSD + LESA method, Exponential Smoothing (ES) method and Evolutionary Neural Network (ENN) method. The Winters method of ES is employed in this experiment so that the ES method can adapt to the time series with seasonality, and the ENN method which is employed here is a hybrid artificial intelligence method. The real fashion sales data which is mentioned in Section 2 is used in this comparison. The comparison of the overall forecasting accuracy in MAPE is given in Table 8, together with a comparison in the computation time of these methods.

As shown in Table 8, the ES method produces a reasonably good forecasting accuracy, although it is slightly worse than the SW method, it is better than the SARIMA and the CSD + LESA methods. The ENN method produces the best forecasting accuracy among all of these methods, but it is very time-consuming in running the evolution process. While the ES method and the SARIMA method can produce forecasting result very promptly, their accuracy is much lower than SW. In terms of computational time, for SARIMA and SW models, the computational times are similar which are both less than 5 min. For CSD + LESA and ES, the computational times are very short and are less than 1 min. For ENN, the computational time is very long. In our setting for ENN, with 14 inputs of the past two weeks sales data and 10 hidden neurons, the required computational time is 5.5 h. In fact, when the neural network is represented by a binary chromosome, the ENN with 14 inputs and 10 hidden neurons has to search in a space with a size of  $2^{14 \times 10}$ , which is obviously a huge searching space. This directly leads to a very heavy computational burden for ENN. As a consequence, from Table 8, we can observe that the SW method produces the second best result. Notice that the difference on accuracy performance between the SW and ENN methods is not big. While considering the required computational time and flexibility in terms

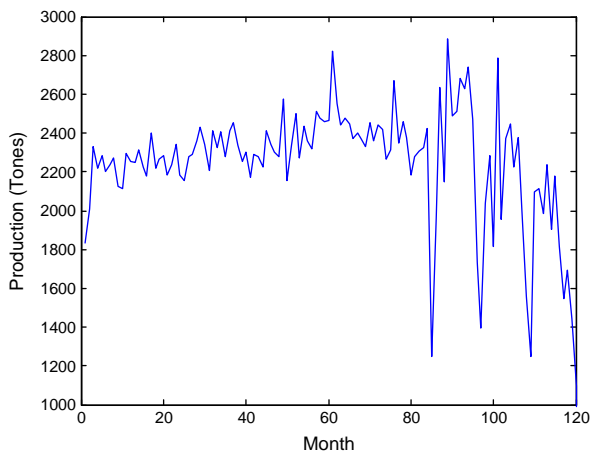


Fig. 15. Monthly production of wool yarn in Australia.

**Table 8**  
Forecasting comparison of different methods.

	SARIMA	SW	CSD + LESA	ES	ENN
MAPE	0.51	0.44	0.53	0.48	0.39
Run time	<5 min	<5 min	<1 min	<1 min	5.5 h

of applications, SW has its strength and it is actually a good choice for everyday forecasting task for highly volatile sales pattern from an industry such as fashion apparels. As a remark, when we compare SW and ES, the amount of improvement on MAPE brought by SW is 0.04 (which is equal to 9.1% of the percentage error reduction). This amount is comparable to the amount of improvement on MAPE when we compare ENN and SW (i.e., 0.05). However, ENN requires a very long computational time (5.5 h) which hinders its real world applications. Thus, even though the time for SW (<5 min) is much longer than ES (<1 min),<sup>2</sup> they both are fast enough for real-world applications. Moreover, it is well-known that ES is only good for forecasting stationary time-series [13] and its performance will generally be poor for highly volatile sales pattern whereas our study in this paper has shown that SW is much more versatile and can function well under this situation. We thus believe that SW has its strength in real world applications.

## 5. Discussions and conclusion

This paper studies three forecasting methods, the CSD + LESA, pure SARIMA, and the hybrid SARIMA with wavelet transform method (SW). The advantages and disadvantages of these methods are studied and compared. From our analysis, we find that the CSD + LESA method can produce good forecasting accuracy on data sets in which the white noise appears to dominate the other factors. It is the best one when the variance of the white noise component of the data is huge, but it becomes worse than SW when the variance of the seasonal cycle component of the data becomes significant. Further study shows that, for the real data set which exhibits significant seasonality, the SW method performs much better in terms of forecasting accuracy compared to the CSD + LESA method. For real data with relatively weak seasonality and highly variable seasonality factor, SW still outperforms the CSD + LESA method, which proposes to us that the CSD + LESA method fails to perform well for highly volatile sales pattern. As comparing to the SARIMA methods, both SW and CSD + LESA seem to perform better under most conditions, with an exception when the decomposed components have very significant amplitude which cannot be omitted in the forecasting, the SARIMA method outperforms SW and CSD + LESA. This is the case when the seasonality is the dominating component. Thus, pure SARIMA is recommended in this case. We incorporate this important finding in our proposed forecasting scheme under SW. Observing that the CSD + LESA method can only produce the best result when the white noise is huge while the SW method produces an only slightly worse result under that condition, we believe that SW does outperform the CSD + LESA method for practical applications in forecasting sales with highly patterns.

In conclusion, we have developed a novel algorithm which combines the classic SARIMA method and the wavelet transform method (called, SW) for forecasting sales time series with a highly volatile pattern. We have conducted experiments with the use of hypothetical data sets, real sales data from the fashion industry, and some public data available online. We have identified the conditions under which SW outperforms other methods. A final comparison using the real data set among various well-established methods is made and the result supports that our proposed SW method is a practical and implementable forecasting scheme for sales forecasting in an industry such as fashion.

<sup>2</sup> As a remark, the experiment on comparing computational times under SW and ES was conducted on a relatively old personal computer. If we replace it by a more powerful computer, we will find that the time for solving SW and ES will be much closer (all very short and within 1–2 min) whereas the time for running ENN is still long.

## Acknowledgements

We sincerely thank Professor Andrew Whinston (the editor), and the anonymous reviewer for their many helpful suggestions and kind advice. We would like to thank the RGC(HK) General Research Funding Scheme (grant number: PolyU 5146/06E) and the Hong Kong Polytechnic University for providing research funding to support this piece of research. Thanks are given to C.H. Chiu for many of his helpful suggestions.

## References

- [1] L. Aburto, R. Weber, Improved supply chain management based on hybrid demand forecasts, *Applied Soft Computing* 7 (2007) 136–144.
- [2] D.D. Achabal, S.H. McIntyre, S.A. Smith, K. Kalyanam, A decision support system for vendor managed inventory, *Journal of Retailing* 76 (2000) 430–454.
- [3] G.E.P. Box, G.M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, 1976.
- [4] C.S. Burrus, R.A. Gopinath, H. Guo, *Introduction to Wavelets and Wavelet Transforms: A Primer*, Prentice Hall, New Jersey, 1998.
- [5] P.C. Chang, C.Y. Fan, A hybrid system integrating a wavelet and TSK fuzzy rules for stock price forecasting, *IEEE Transactions on Systems, Man, and Cybernetics, Part C* 38 (2008) 802–815.
- [6] P.C. Chang, C.Y. Lai, A hybrid system combining self-organizing maps with case-based reasoning in wholesaler's new-release book forecasting, *Expert Systems with Applications* 29 (2005) 183–192.
- [7] P.C. Chang, Y.W. Wang, C.Y. Tsai, Evolving neural network for printed circuit board sales, *Expert Systems with Applications* 29 (2005) 83–92.
- [8] P.C. Chang, Y.W. Wang, W.N. Yang, An investigation of the hybrid forecasting models for stock price variation in Taiwan, *Journal of the Chinese Institute of Industrial Engineering* 21 (2004) 358–368.
- [9] T.M. Choi, D. Li, H. Yan, Mean-variance analysis for the newsvendor problem, *IEEE Transactions on Systems, Man, and Cybernetics, Part A* 38 (2008) 1169–1180.
- [10] K.C. Chui, *Introduction to Wavelet*, Academic Press, San Diego, CA, USA, 1992.
- [11] C. Frank, A. Garg, A. Raheja, L. Sztandera, Forecasting apparel sales using mathematical modeling, *International Journal of Clothing Science and Technology* 15 (2003) 107–125.
- [12] R. Gencay, F. Selcuk, B. Whitcher, *Wavelets and other Filtering Methods in Finance and Economics*, Academic Press, New York, 2001.
- [13] J.E. Hanke, D.W. Wichern, *Business Forecasting*, 9th ed. Prentice Hall, 2009.
- [14] C.S. Hilaris, S.K. Goudos, J.N. Sahalos, Seasonal decomposition and forecasting of telecommunication data: a comparative case study, *Technological Forecasting & Social Change* 73 (2006) 495–509.
- [15] H.S. Hippert, C.E. Pedreira, R.C. Souza, Neural networks for short-term load forecasting: a review and evaluation, *IEEE Transactions on Power Systems* 16 (2001) 44–55.
- [16] W.Y. Liang, C.C. Huang, Agent-based demand forecast in multi-echelon supply chain, *Decision Support Systems* 42 (2006) 390–407.
- [17] S.G. Makridakis, S.C. Wheelwright, V.E. McGee, *Forecasting: Methods and Applications*, 3rd ed. Wiley, New York, 1998.
- [18] F.F. Nobre, A.B.S. Monteiro, P.R. Telles, G.D. Williamson, Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology, *Statistics in Medicine* 20 (2001) 3051–3069.
- [19] K. Papagiannaki, N. Taft, Z.L. Zhang, D. Christopher, Long-term forecasting of internet backbone traffic: observations and initial models, *The 18th IEEE INFOCOM Conference (San Francisco, March 30–April 3, 2003)*, 2003.
- [20] G. Schwarz, Estimating the dimension of a model, *Annals of Statistics* 6 (1978) 461–464.
- [21] J. Taylor, Forecasting daily supermarket sales using exponentially weighted quantile regression, *European Journal of Operational Research* 178 (2007) 154–167.
- [22] S. Thomassey, A. Fiordaliso, A hybrid sales forecasting system based on clustering and decision trees, *Decision Support Systems* 42 (2006) 408–421.
- [23] M. Versace, R. Bhatt, O. Hinds, M. Shiffer, Predicting the exchange traded fund DIA with a combination of genetic algorithms and neural networks, *Expert Systems with Applications* 27 (2004) 417–425.
- [24] B.L. Zhang, R. Coggins, M.A. Jabri, D. Dersch, B. Flower, Multiresolution forecasting for futures trading using wavelet decompositions, *IEEE Transactions on Neural Networks* 12 (2001) 765–775.



**Tsan-Ming Choi** received his PhD from The Chinese University of Hong Kong. He is currently an associate professor at The Hong Kong Polytechnic University. His current research interests mainly focus on systems engineering and supply chain operations. He has published in leading academic journals such as *IEEE Transactions on Automatic Control*, *Automatica*, *Production and Operations Management*, *Decision Support Systems*, *IEEE Transactions on Systems, Man, and Cybernetics – Parts A and B*, and various other *IEEE Transactions*. He is the editor/co-editor of three research handbooks and seven special issues in academic journals. He is currently an associate editor of *IEEE Transactions on Systems, Man, and Cybernetics – Part A*.



**Yong Yu** is currently a postdoctoral research associate at the Hong Kong Polytechnic University. He received his PhD from the Hong Kong Polytechnic University and he published in journals such as *Decision Support Systems* and *International Journal of Production Economics*.



**Kin-fan Au** is associate professor in the Institute of Textiles and Clothing of The Hong Kong Polytechnic University. His research interest is in the business aspects of fashion and textiles, particularly in the area of global trading of fashion and textile products. Dr. Au has published many papers in textiles and related journals on topics of world trading, offshore production and modelling of textiles and apparel trade.