THE MODEL OF ANALOG COMPLEXING ALGORITHM BASED ON EMPIRICAL MODE DECOMPOSITION METHOD

He Jirui[1] **Tian Yixiang**[2]

Abstract: Analog Complexing (AC) algorithm can be considered a sequential pattern recognition method for prediction. However, financial Time-series data are often nonlinear and non-stationary, which cause some difficulties when used AC algorithm in prediction. Aiming at this problem, in this paper, using Empirical Mode Decomposition (EMD) to handle original data, and we will obtain a series of stationary Intrinsic Mode Functions (IMF); then each IMF is predicted dynamically by AC. By the empirical studies on NYMEX Crude Oil Futures price show that AC algorithm based on EMD method have high precision in 1 step and 3 steps dynamically prediction.
Key words: Analog Complexing algorithm, Empirical Mode Decomposition, Intrinsic Mode Function, Dynamically prediction

1. INTRODUCTION

When we modeling to complex system, such as financial market and meteorological system, the relationship between system variables is complex, even there haven't fixed regularity. But the characteristic of complex system is that a grate lot of time series patterns existed in system variables, researchers can describe dynamic system by utilizing these time series patterns, because different modes represent some periodicity, some trends and relationship existed in data.

[1] Department of Management, UESTC, Chengdu, 610054.
[2] Department of Management, UESTC, Chengdu, 610054.
* Received 5 July 2007; accepted 26 August 2007

Analog Complexing(AC) algorithm[3] was developed by Lorence and was successfully applied to meteorological forecast first. It can be considered as a sequential pattern recognition method for predicting and qualitatively explaining fuzzy objects inherently. Recently, AC algorithm has been enhanced by an inductive self-organising approach and by an advanced selection procedure to make it applicable to evolutionary processes too. For example, J.-A. Muller and Guangzhong Liu (2002)[4] explained the superiority of AC algorithm, and tested 2 stocks of the German car industry (BMW, VW), the US-Dollar/DM exchange rate. Changzheng He (2004)[5] forecasted the economic developed index in ChengDu used AC algorithm, Ming Huang (2005)[6] used AC algorithm in forecasting stocks and got the turning-point of price trends accurately. AC algorithm needs neither evaluation or assumption in advance for imput variable, nor priori knowledge used to choose variable for model construction, which means that forecasting are driven by data. It is superior than other forecasting methods that make forecasting based on figure patterns.

When we use AC algorithm, the main assumptions of the method are: ①the system is described by a multidimensional process; ②many observations of the process are available (long time series); ③the multidimensional process is sufficiently representative, i.e., essential system variables are forming the data set; ④it is likely that any observed behaviour of the process will repeat similarly over time. Therefore, the prediction can be achieved by combining the known continuations of the analogous patterns to a continuation of the reference pattern. In general, AC algorithm is a three-step-procedure (see literature 3 for detail algorithm): ①generation of alternate patterns; ②transformation of analogues; ③selection of most similar patterns; when using AC algorithm for prediction purposes, a fourth step is necessary: ④combining forecasting.

Form above, the forecasting effect of AC algorithm lies on pattern similarity measures, selection, and similarity pattern combining. The new method of GMDH (Group Method of Data Handling)[3] has solved similarity pattern optimized combination, but similarity pattern measures in AC algorithm based on Eucildean Spaces, and sensitivity to noise, so how to improve method of similarity measures need to discuss deeply. Then, financial time-series data are often nonlinear and non-stationary, the periods of volatility are different, and the trend is uncertain, which all cause some problems to choose pattern's length and to measure pattern's similarity, accordingly influence the accuracy of forecasting results. For that, we introduce empirical mode decomposition method to forecast. First, the original time series are made stationary by empirical mode decomposition method; then, all the stationary data are forecasted dynamically via AC algorithm. In this paper, we used a new model to mine the history information and understand the underlying rules for improving forecasting accuracy.

2. EMD METHOD

EMD (empirical mode decomposition) method[7] and Hilbert transform, which is proposed by Huang et al. (1998), is a new method for analysing nonlinear and non-stationary data. Contrary to almost all the previous decomposing methods, EMD is empirical, intuitive, direct, and adaptive, with the a posteriori defined basis derived from the data. It has been used in many science fields successfully, such as signal handling, oceanography, meteorology. The decomposition is designed to seek the different simple intrinsic modes of oscillations in any data based on the principle of scale separation. The data, depending

---

[3] Mueller J A, Lemke F. ***Self-orgnizing data mining*** [M], Berlin Hamburg, 2000.114-130
[4] J.A.MULLER, Guangzhong LIU. **Application of Analog Complexing Algorithm to Financial Forecasting** [J]. *Or Transactions*,2002.6(3):1-16
[5] He Changzheng. *Self-Organising Data Mining and Economics Forecasting* [M]. Science Publishing House, 2005.
[6] Huang Min. *The Application of Self-Organising Data Mining in Stock Market* [D]. Chengdu: UESTC:2006. 29-38
[7] Huang N E , Shen Z. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis [J]. *Proceedings of the Royal Society London*, 1998 , A ,454: 903-995

on it complexity, may have many different coexisting modes of oscillation at the same time. Each of these oscillatory modes is represented by an Intrinsic Mode Function (IMF) with the following definitions: (a) in the whole data set, the number of extrema and the number of zero-crossings must either equal or differ at most by one, and (b) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. The initial IMF by EMD method show the high frequency component, along the decomposed process, the IMF has lower frequency and longer period, When the IMFs and the residual are summed together they form the original time series.

Pursuant to the above definition for IMF, one can implement the needed decomposition of any function, known as sifting, as follows:

1）Identify all the maxima and minima of time series $x(t)$;

2）Generate its upper and lower envelopes $(e_{\max}(t), e_{\min}(t))$, with cubic spline interpolation.

3）Calculate the point-by-point mean $m(t)$ from upper and lower envelopes: $m(t) = (e_{\min}(t) + e_{\max}(t))/2$;

4）Extract the mean from the time series and define the difference of $x(t)$ and $m(t)$ as $d(t)$: $d(t) = x(t) - m(t)$;

5）Check the properties of $d(t)$:

If it is an IMF, denote $d(t)$ as the i[th] IMF and replace $x(t)$ with the residual $r(t) = x(t) - m(t)$; the $d(t)$ is often denote as $c_i(t)$;

If it is not, replace $x(t)$ with $d(t)$;

6）Repeat steps (1)-(5) until the residual sarisfies some stopping criterion.

The original time series could denote by: $x(t) = \sum_{i=1}^{n} imf_i(t) + r(t)$.

3. MODELING PROCESS

According to the characteristics of AC algorithm and EMD method, we combine two methods to form a new prediction model, make the stationary and periodic IMF as prediction element, for solving the choice of pattern's length and measurement of pattern's similarity. Firstly, using the empirical mode decomposition the original sequence data are stationarized, and a variety of IMFs and a residual are obtained (Figure 1). We see, from figure 1, all IMFs are not only stationary and have strong periodic, so we can choose pattern's length according to the IMFs' period; secondly the IMFs and residual are forecasted by AC algorithm separately; finally, Summing up the predicted results of IMF and residual yields he predicted data of the original data. The detail algorithm flow chart showed in Figure 2.

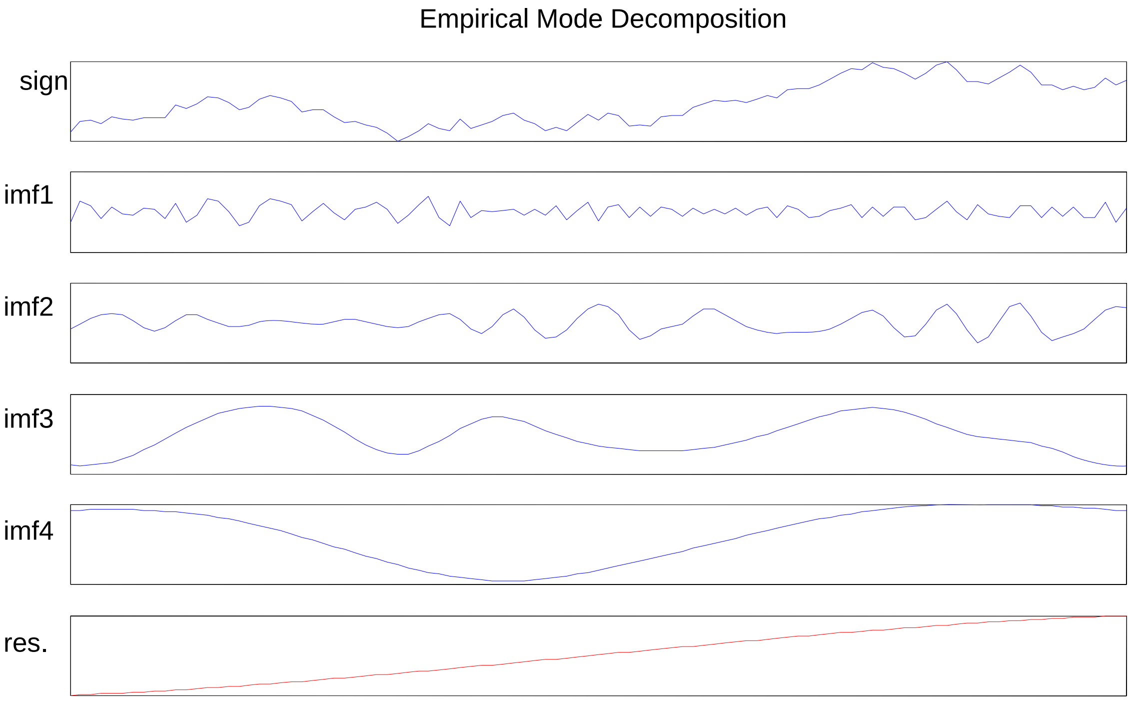

Figure 1 Empirical Mode Decomposition

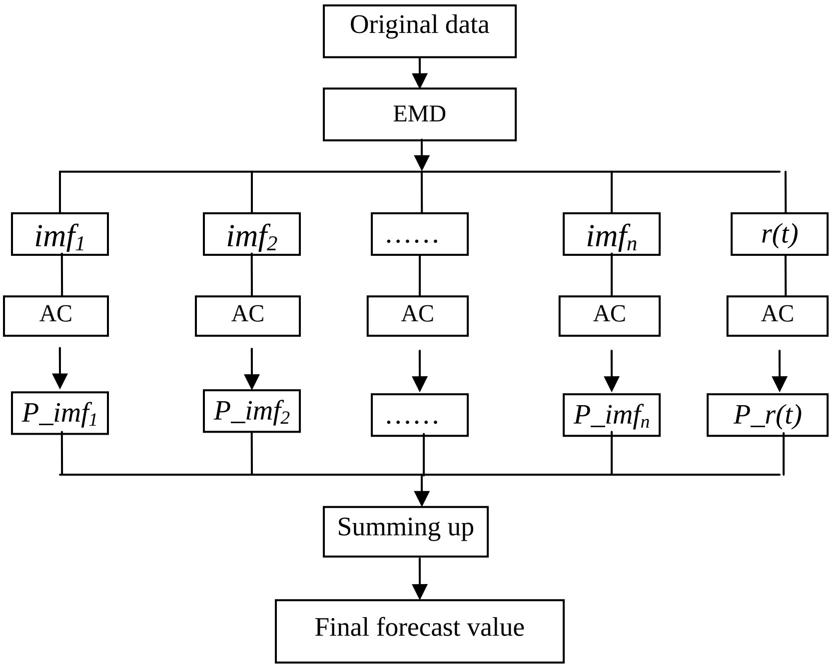

Figure 2 Model flow chart

4. EMPIRICAL ANALYSIS

4.1 Data

Our data consist of daily crude oil futures closing prices traded at the Mew York Mercantile Exchange (NYMEX) from Jan 5, 2004 to Oct 16, 2006, constituting 698 data points. The training data set included 638 data (from Jan 5, 2004 to Jul 21, 2004), the other 60 data (from Jul 24, 2006 to Oct 16, 2006) were used as a test sample. Data come from FuYuan futures software.

4.2 Forecasting

In this study, one-step-ahead forecasting policy is implemented. For example, first 638(from Jan 5, 2004 to Jul 21, 2004) training data set was decomposed by EMD, obtained 5 IMFs and a residual (figure 2). We use AC algorithm to forecast every IMF and residual, summing up these predicted value get the predicted closing price at Jul 24, 2004. Then we add the true value at Jul 24, 2004 to the training data set and re-forecast the next day. Apply the same procedure to generate 60 times one-step-ahead forecasts.

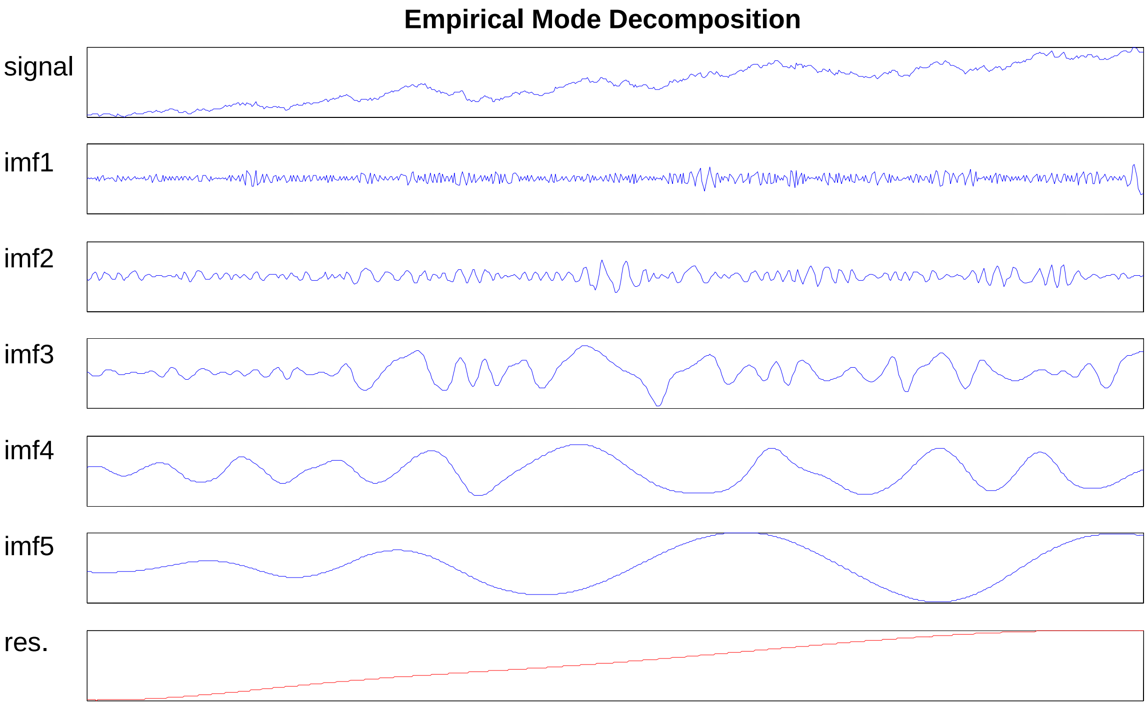

Figure 3 2004.1.5-2006.7.21data decomposed by EMD

When we use AC algorithm to forecast IMFs, the similarity between the reference pattern and a given pattern reach 0.99, even 1; but forecast the original data sequence, the similarity between the reference pattern and a given pattern only reach 0.92, even lower.

4.3 Forecasting results

To compare the forecasting results, three forecasting error functions are calculated, they are the Root

Mean Squared Error(RMSE), Mean Absolute Error(MAE) and Mean Absolute Percentage Error(MAPE).

Table 1 illustrate three error values between original AC algorithm and AC algorithm combined EMD method. Figures 4-5 shows actual and forecasting values of one-step-ahead and three-step-ahead from Jul 24, 2006 to Oct 16, 2006.

The table and figures clearly demonstrate that the model combined EMD method is able to produce lower forecasting errors by all three measures, and the forecasting value are closer to true value, suggesting that the new model has well short-term forecasting capability.

Table 1：RMSE, MAE and MAPE Value

	RMSE	MAE	MAPE
one-step-ahead	1.2416	1.0134	0.0148
three-step-ahead	1.8064	1.3717	0.0195
one-step-ahead（EMD）	1.0361	0.8278	0.0121
three-step-ahead（EMD）	1.5833	1.2471	0.0183

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}\ ,\quad MAE = \frac{1}{T}\sum_{t=1}^{T}\left|\hat{y}_t - y_t\right|,\quad MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{\hat{y}_t - y_t}{y_t}\right|$$

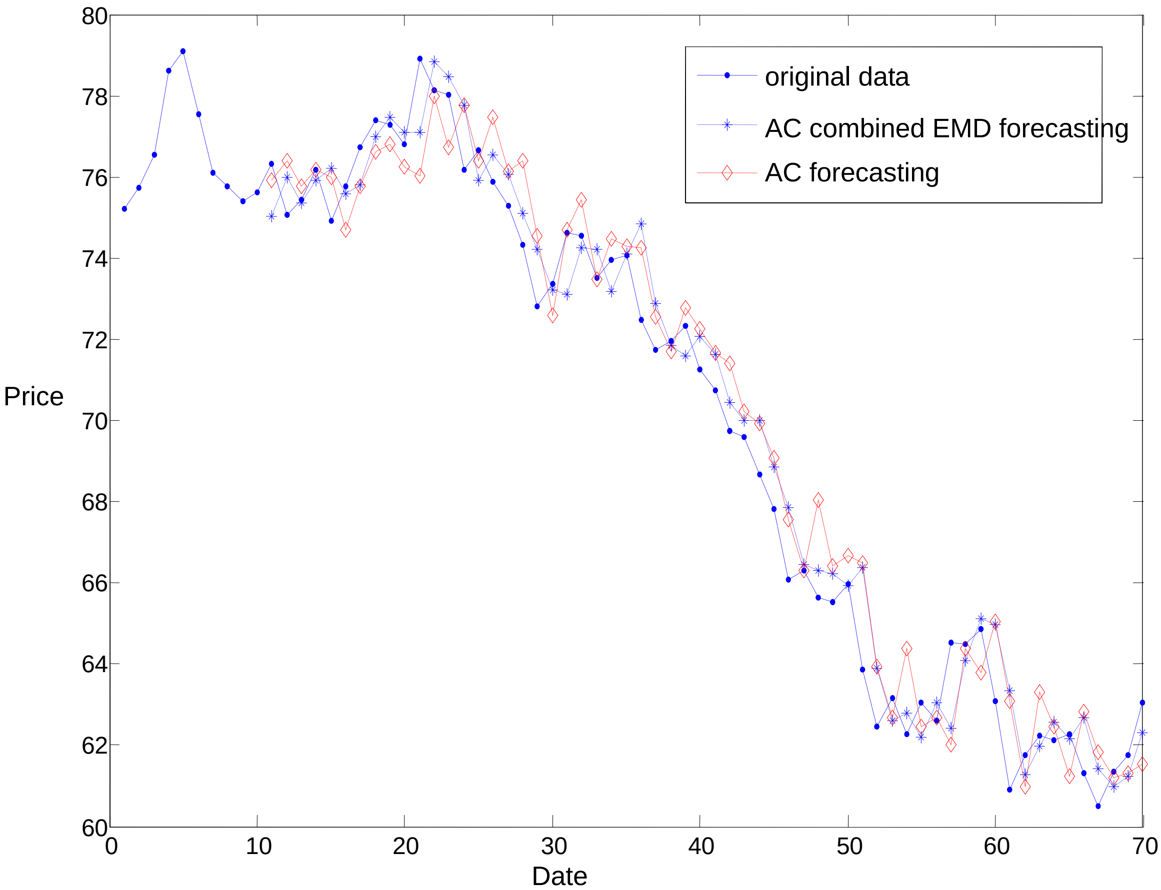

Figure 4 One-step-ahead forecasting (from Jul 24, 2006 to Oct 16, 2006)

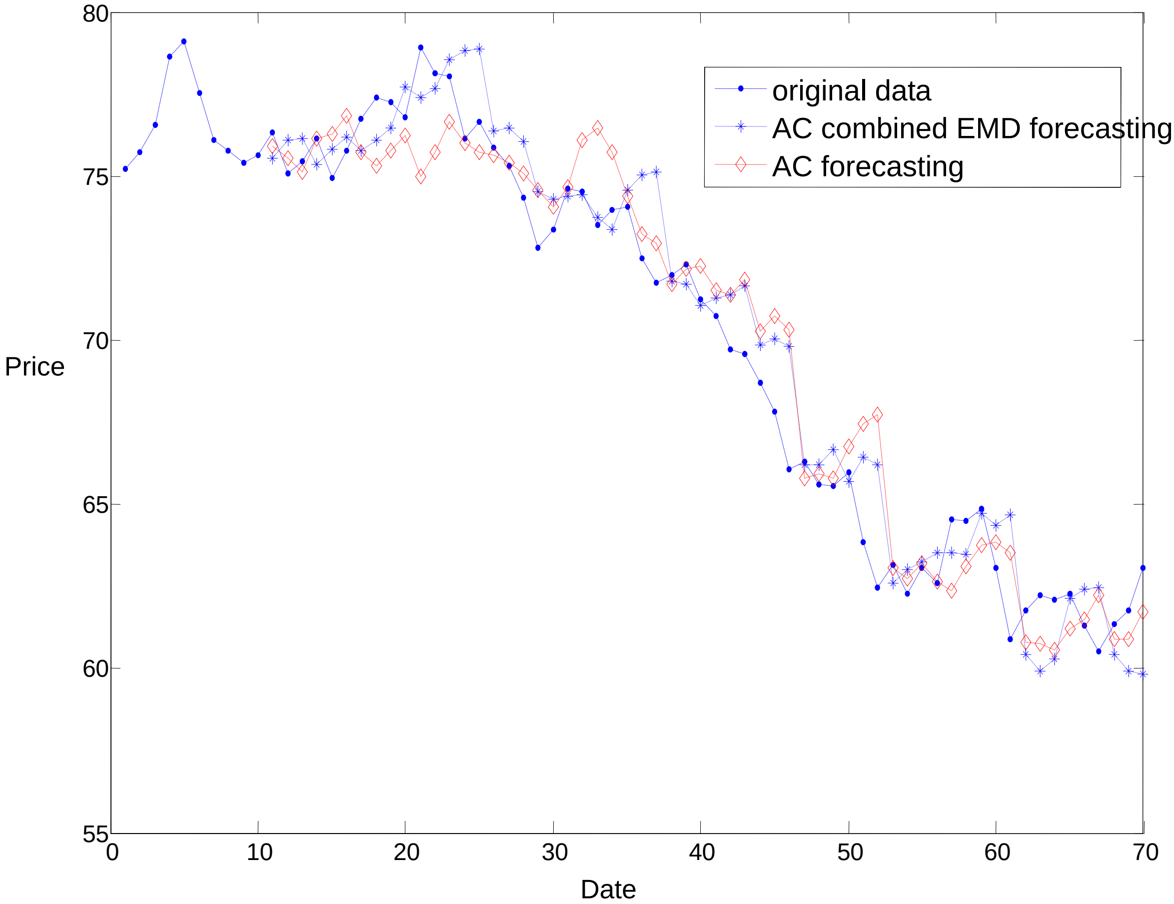

Figure 5 Three-step-ahead forecasting (from Jul 24, 2006 to Oct 16, 2006)

5. CONCLUSIONS

This paper examines the application of AC algorithm combined EMD model for time series forecasting regarding to NYMEX crude oil futures closing price. First, using the empirical mode decomposition a variety of IMFs and a residual are obtained, which were forecasted by AC algorithm, then summing up these predicted IMFs' value. Our results illustrate the new model has higher accuracy than traditional model, and it is feasible easily. In fact, the AC algorithm combined EMD model can forecast for a fairly wide time series, especially for non-stationary and nonlinear time series and complex systems, have a broad application.

REFERENCES

Arie Harel, Giora Harpaz. Forecasting futures returns in the presence of price limits [J]. *The Journal of Futures Markets*. 2005, 25(2):199-210

Jason Laws, John Thompson. The efficiency of financial futures markets Tests of prediction accuracy [J]. *European Journal of Operational Research***,** 2004, 155: 284–298

Hui Gao. Study on Correlation of Domesticand Foreign Fuel Oil Prices and Dynamic Trend Forecast Modeling [J]. *Oil Price Forum*, 2005,13(12):24-29

Ivakhnenko A.G. Sorting methods for modeling and clusterization (survey of the GMDH papers for the year 1983-1988)[J]. *Sov. J. of Automation and Information Sciences*, 1988, 21(4):1-13

Liu Huiyu, Lin Zhenshan etc. Analysis on the Fluctuation of Grain Output in China and Its Causes at Multi - time Scale Based on Empirical Mode Decomposition Method [J]. *Journal of Natural Resources*, 2005, 20(5):745-751

Madala H.R, Ivakhnenko A.G. *Inductive learning algorithms for complex systems modeling* [M]. Boca Raton,
Tokyo: CRC Press. 1994

Motnikar, B.S. Time-series forecasting by pattern imitation [J]. *OR Spektrum*, 1996, 18(1): 43-49

Perry Sandorsky. Modeling and forecasting petroleum futures volatility [J]. *Energy Economics*. 2006, 28 :267-488

Salah Abosedraa, Hamid Baghestani. On the predictive accuracy of crude oil futures prices [J]. *Energy Policy*, 2004, 2004(32):1389-1393

Saeed Moshiri, Faezeh Foroutan. Forecasting Nonlinear Crude Oil Futures Prices [J]. *The Energy Journal*. 2006; 27(4): 81~95

Wan Shiquan, Fen Guolin etc. Extracting Useful Information from The Observations for The Prediction Based on EMD Method [J]. *ACTA METEOROLOGICA SINICA*, 2005, 63(4):516-524

Xiang Xiaodong. The Application Research of Forecasting about Petroleum Futures Price Based on Wavelet Neural Network Forecasting Method[J]. *Technology Economics*, 2006, 25(6):121-124

Yu Weibin etc. The Cointegration Analysis of the Brent Futures Market[J]. *Application of Statistics and Management*, 2004, 23(5):26-33