Nonnegative Matrix Factorization Based Decomposition for Time Series Modelling

Tatjana Sidekerskienė, Marcin Woźniak, Robertas Damaševičius

Nonnegative Matrix Factorization Based Decomposition for Time Series Modelling

Tatjana Sidekerskienė[1], Marcin Woźniak[3], Robertas Damaševičius[2]

[1]Department of Applied Mathematics
[2]Department of Software Engineering
Kaunas University of Technology, Kaunas, Lithuania
[3]Institute of Mathematics, Faculty of Applied Mathematics
Silesian University of Technology, Gliwice, Poland

robertas.damasevicius@ktu.lt

Abstract. We propose a novel method of time series decomposition based on the non-negative factorization of the Hankel matrix of time series and apply this method for time series modelling and prediction. An interim (surrogate) model of time series is built from the components of the time series using random cointegration, while the best cointegration is selected using a nature-inspired optimization method (Artificial Bee Colony). For modelling of cointegrated time series we use the ARX (AutoRegressive with eXogenous inputs) model. The results of modelling using the historical data (daily highest price) of S&P 500 stocks from 2009 are presented and compared against stand-alone ARX models. The results are evaluated using a variety of metrics (RMSE, MAE, MAPE, Pearson correlation, Nash-Suttcliffe efficiency coefficient, etc.) as well as illustrated graphically using Taylor and Target diagrams. The results show a 51-98% improvement of prediction accuracy (depending upon accuracy metric used). The proposed time series modelling method can be used for variety applications (time series denoising, prediction, etc.).

Keywords: time series modelling; time series decomposition; ARX; random cointegration.

1   INTRODUCTION

Modelling of time series (including the prediction of any future values) is an important problem with many areas of application, including problems in bioinformatics [1], medical diagnosis [2], air pollution forecasting [3], industrial machine condition monitoring [4], environmental modelling [5], financial investment [6], production planning [7], sales forecasting [8] and stock portfolio management [9]. In the domain of financial investment management, to invest and take proper decision, a more precise forecasting of financial environments is an important issue.

Financial time series originate from real-world systems (such as stock exchanges) and represent the outcome of complex multi-layer dynamic interactions between multiple agents [10]. Real-world systems have mostly nonlinear and non-stationary behaviour. A nonlinear time series is a signal generated by a nonlinear dynamic process, in which the output is not directly proportional to the input, i.e. even a small change in input may lead to large change in output. Moreover, real-world systems often operate under transient non-stationary conditions, characterized by time-changing statistics. These properties prevent from acquisition of an effective predictive model and its reliable application. Therefore, the researchers are motivated to explore new data modelling methods.

Time series can be represented by different patterns of behaviour. Sometimes it is useful to decompose time series into several components, each representing some pattern or class of behaviour. It is common to decompose the economical time series data into trend (long term variation), cyclical (repeated but non-periodic fluctuations), seasonal and irregular (or noise) components [11]. Predictability can be used as a criterion to decomposing a times series into deterministic and non-deterministic components (Wold decomposition [12]). STL (Seasonal Trend decomposition using Loess) [13] can be used to estimate seasonal effects on time series and to predict future seasonally adjusted values. When they have time varying properties, the Holt-Winters decomposition [14] uses exponential to derive the decomposition model, which also includes error corrections in forecasting future values of the time series. Wavelet decomposition [15] decomposes a time series using the scaling function and wavelet functions (i.e. mother wavelet). Second generation wavelets can be used to generate wavelet functions in the spatial domain and to deal with complex structure, arbitrary boundary conditions, and irregular sampling intervals of time series [16].

Empirical Mode Decomposition (EMD) [17] initially has been developed for natural and engineering sciences for analysing nonlinear and non-stationary data such as sea wave data and earthquake signals, but has been applied to financial data as well [18]. EMD decomposes any time series into a finite number of intrinsic mode functions (IMF). Since the decomposition is based on the local characteristic time scale of the data, it is applicable to non-linear and non-stationary processes, and also can extract variability on different time scales. By performing clustering [19] or remixing [20] of IMFs one can identify different structural patterns of a time series and further analyze them with respect with their orthogonality and cross-correlation properties. However, the application of EMD has many important problems such as the end effect [21] and the IMF stopping criterion [22]. Extensions of EMD such as BoostEMD [23] try to alleviate these problems and improve the characteristics of IMF.

Despite long history and the availability of many time series decomposition algorithms and models, in many cases they are unable to account for complex structure (such as multiple or fractional seasonality) of time series.

We propose a novel method of time series decomposition based on the non-negative factorization of the Hankel matrix of time series in Section 2. In Section 3 we apply the method for time series prediction using the random cointegration approach. We describe the results of experiments using the historical stock data in Section 4. Finally, conclusions and discussion of future work are presented in Section 5.

2   DECOMPOSITION METHOD

We propose a new EMD-like signal decomposition method, called Nonnegative Hankel Matrix Factorization based Decomposition (NHMFD). The task is to decompose signal $x(t)$ into long term (trend), cyclical, and irregular components:

$$x(t) = \sum_{j=1}^{k} c_j + r_T + r_R \qquad (1)$$

here c_j is a cyclical component (or Intrinsic Mode Funtion, IMF), and r_T is a trend, and r_R is the irregular (random) component.

The steps for proposed decomposition are as follows:

1. **Normalization:** the signal $x(t)$ is represented as time series X and normalized to $[0,1]$ as follows:

$$\hat{X} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \qquad (2)$$

here $X_{\min}$ and $X_{\max}$ are the smallest and largest values of the time series.

2. **Construction of Hankel matrix:** Hankel matrix H of the normalized time series $X = \{ x_1, x_2, x_3, \cdots, x_l \}$ is a square matrix:

$$H = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots \\ x_2 & x_3 & x_4 & \cdots \\ x_3 & x_4 & x_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \ h_{i,j} = x_{i+j-1} \qquad (3)$$

The size of Hankel matrix H is $n = \left\lfloor \dfrac{l+1}{2} \right\rfloor$ here l is the length of X.

3. **Nonnegative matrix factorization:** performs factorization of the form $H \approx V \times W$, here $V = \Re^{n \times k}$ and $W = \Re^{k \times n}$ are positive matrices, and k is the number of factors. Factorization is not exact: matrices V and W are chosen to minimize the residual D between H and $V \times W$:

$$D = \| H - V \times W \| \qquad (4)$$

4. **Reconstruction of Hankel matrix:** The signal is reconstructed as

$$H' = V \times W + E \qquad (5)$$

here H' is the reconstructed Hankel matrix, E is the reconstruction error.

5. **Reconstruction of signal:** The signal is reconstructed from the reconstructed Hankel matrix H' by taking the means of matrix elements along its minor (secondary) diagonals as follows:

$$X' = \{x_1', x_2', x_3', \ldots, x_l'\}, \quad x_i' = \sum_{i=j+k-1} h_{j,k}'/i \,. \tag{6}$$

The reconstruction of a signal is not exact, i.e. $X = X' + \varepsilon$, here ε is the reconstruction error. Linear regression is performed to find the fitting coefficients α and β such as to minimize error $\sum (X - \alpha - \beta X')^2$.

6. **Calculation of intrinsic mode:** the mode of a signal is calculated as follows. Let fitted reconstructed signal be: $\tilde{X} = \alpha + \beta X'$. Then the first component of the signal is defined by centering the fitted reconstructed signal as: $\hat{C}_l = \tilde{X} - \sum \tilde{X}/l$, and the decomposition residue is $\hat{R}_l = \hat{X} - \hat{C}_l$.

7. **Iterative decomposition of residue:** decomposition is continued with residue until the desired number of extracted modes is reached.

8. **Denormalization.** Extracted modes and residue are denormalized as follows:

$$C = \hat{C}(X_{max} - X_{min}), \quad R = \hat{R}(X_{max} - X_{min}) + X_{min} \tag{7}$$

9. **Extraction of trend and irregular component.** Finally, the residue is decomposed into trend and irregular (random or noise) component by computing the least-squares fit of a straight line to the residue and subtracting the resulting function from the residue as follows:

$$r_T = at + b, \, r_R = R - r_T \tag{8}$$

here a and b minimize error $\sum_t \left\{ \left[(at + b) - R(t) \right]^2 \right\}$.
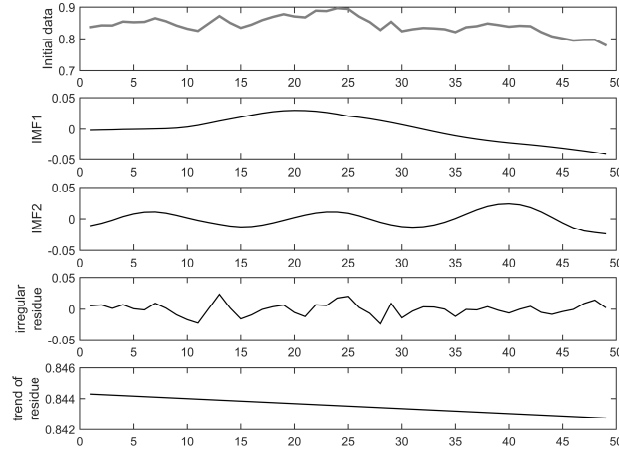
Fig. 1. Example of decomposition: sample signal, IMFs, irregular residue and trend residue

Fig. 1 shows an example of decomposition for a time series (1st subplot). The series was decomposed into two cyclical components, irregular residue and trend residue.

3 MODELLING METHODOLOGY

To derive a model of a time series, we use a combination of several methods:

1) Nonnegative Hankel Matrix Factorization based Decomposition (NHMFD) proposed in Section 2 to derive Intrinsic Mode Functions (IMFs) (see Section 2).

2) Random cointegration of IMFs. A random vector P is generated that is multiplied by the IMF component matrix C to obtain a surrogate time series S. Cointegration is defined as the existence of a stationary linear combination of nonstationary time series [24]. This indicates that there is a long run equilibrium relationship between the series. While stationarity in a strict statistical sense may not be always achieved, cointegration is considered to improve the statistical characteristics of time series [25], which is good for predictability. The cointegrated series can be considered as surrogate data series [26], which inherit the statistical properties of the original data but may have some desired properties such as predictability. The original time series can be restored as a sum of IMFs with unitary weights as given by Eq. 1. The cointegrated (surrogate) time series S can be generated by summing time series components c_j with random weights p_j as follows:

$$S = \sum_{j=1}^{k} p_j c_j = PC \tag{9}$$

3) Selection of best fitting random cointegration vector $\hat{P}$ using a nature-inspired optimization method. We use Artificial Bee Colony (ABC) [27] algorithm to find a cointegration vector that ensures the lowest RMSE of 1-step ahead prediction using the ARX model on surrogate testing data. ABC is based on a model of foraging behaviour of a honeybee colony. This model consists of three essential components: food sources, employed foragers, and unemployed foragers, and defines two leading modes of the honeybee colony behaviour: recruitment to a food source and abandonment of a source. The ABC algorithm implements a population-based search in which artificial bees change their positions aiming to discover the places of food sources with high nectar amount and finally the one with the highest nectar.

4) Modelling of cointegrated time series using ARX (autoregressive with exogenous inputs) model [28], with a surrogate time series $\hat{S} = \hat{P}C$ as an input and original time series as output. The input-output relationship of an ARX model is:

$$X(t) = \frac{B(q^{-1})}{A(q^{-1})} \hat{S}(t) + \frac{1}{A(q^{-1})} e(t) \tag{10}$$

here $\hat{S}(t)$ is model input (surrogate time series), $X(t)$ is model output (original time series), A and B are polynomials, q^{-1} is a backward shift operator, and $e(t)$ is noise.

4   EXPERIMENTS & RESULTS

We use the historical data (2009) of S&P 500 stocks (245 days, 379 stocks with non-zero value). Here we analyze the daily highest price value. The dataset is separated into 3 subsets: train (40%), test (40%) and validate (20%). Train subset is used for decomposition, test subset for finding best co-integration vector and validate subset for calculating prediction accuracy. We decompose each time series into 4 components (2 cyclical, trend and irregular residues). For the ABC algorithm, we use 100 bees (colony size), maximum number of iterations is 200. Algorithm is run 20 times and the best solution is retained.

We model the 1-day ahead value of time series and evaluate model accuracy using a variety of metrics (RMSE (Root Mean Square Error), MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), MSE (Mean Squared Error), Pearson correlation, Nash-Suttcliffe (NS) efficiency coefficient, MSPE (Mean Squared Prediction Error), RMSPE (Root Mean Square Percentage Error)) as well as illustrate graphically using Taylor diagram and target diagram.

Fig. 2 shows the mean values of RSME, MAPE, MAE and MSE. In all cases, the model of a time series derived using the proposed method is better (error is smaller).

Fig. 3 shows the mean values of Pearson correlation, NS coefficient as well as the mean rank. Mean rank was calculated by ranking the modelling results of each stock (better = 1, worse = 2) based on the RMSE value. The mean rank of the proposed method (1.21) is better (lower) than the mean rank of the ARX model (1.79). The difference is significant with $p = 3 \cdot 10^{-67}$ (using paired t-test).
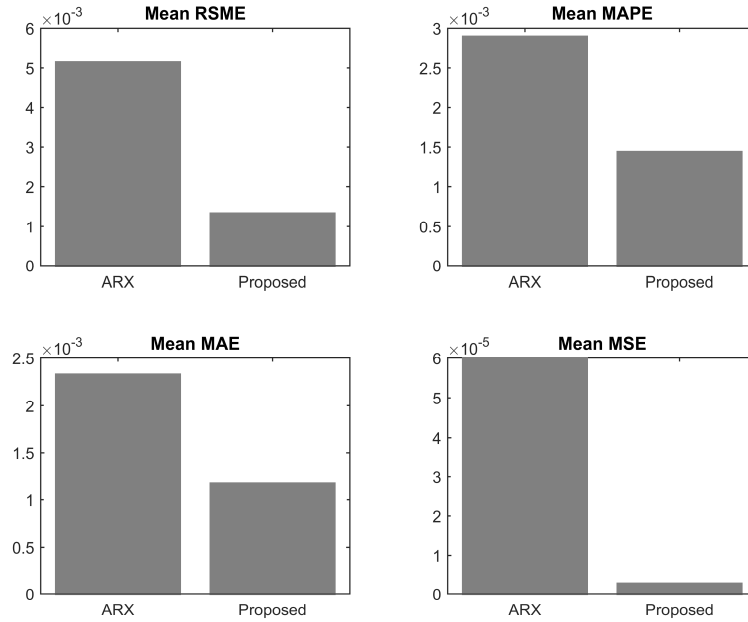

Fig. 2. Mean values of RSME, MAPE, MAE and MSE
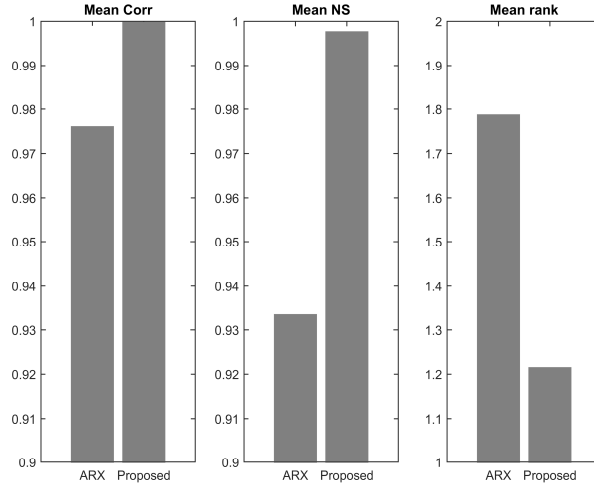
Fig. 3. Mean values of correlation, Nash-Suttcliffe (NS) efficiency coefficient, and rank

Fig. 4 presents the probability that the proposed method will score better than the ARX model on a given accuracy metric. The probability value was calculated by bootstrapping (random sampling with replacement) using 10000 samples. The probability that the proposed method yields better accuracy ranges from 0.709 (using MAE) to 0.916 (using Pearson correlation).
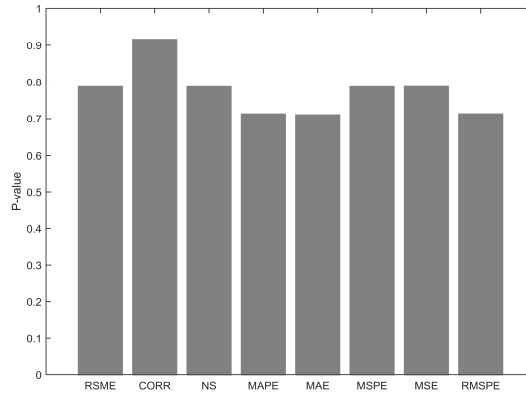
Fig. 4. Probability that the proposed method scores better than ARX

Taylor diagram (Fig. 5, left) is used to perform the comparative assessment of several different models and to quantify the degree of correspondence between the modelled and observed behaviour in terms of three statistics: Pearson correlation, RMSD (Root Mean Square Deviation), and the standard deviation [29]. The model that is located lower on the diagram is considered to represent the reality better. Note that the proposed model (*Prop*) is lower than the ARX model.

Target diagram (Fig. 5, right) [30] summarizes and visualizes information regarding the bias and the error between a model and real observations. The Y-axis corresponds to the normalized bias (Bias) and the X-axis corresponds to the normalized unbiased RMSD (uRMSD). Variables located in the upper part of the diagram $(Y > 0)$ are overestimated by the model. The standard deviation of a model is larger than the standard deviation for the observations when variables are located in the right part of the target diagram $(X > 0)$. The distance between any point and the origin is the value of the total RMSD. Note that while the proposed model performs better than the ARX model in terms of uRMSD ($5 \cdot 10^{-4}$ vs $5 \cdot 10^{-3}$) and total RSMD ($1.3 \cdot 10^{-3}$ vs $5.1 \cdot 10^{-3}$), it has a larger bias ($8.1 \cdot 10^{-3}$) than the ARX model ($1.6 \cdot 10^{-3}$).
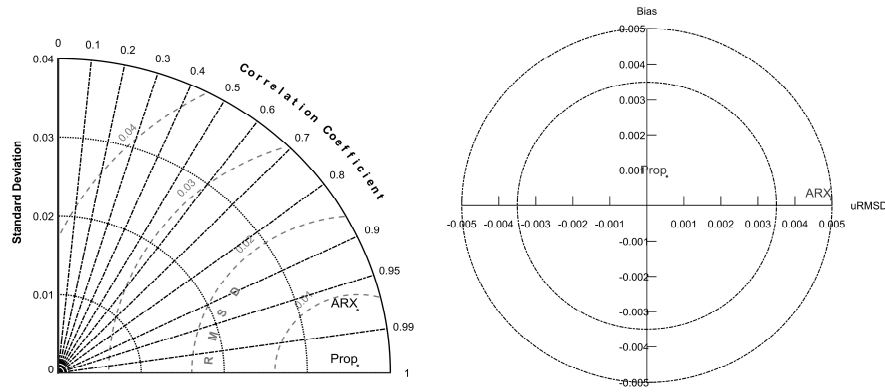
Fig. 5. Taylor diagram (left) and target diagram (right)

The results are summarized in Table 1. Mean values of accuracy metrics are given and a relative improvement of the proposed method over ARX. Overall, the proposed method has improved the prediction accuracy of 78.6 % of stocks (by RMSE).

Table 1. Summary of results

Metric	ARX	Proposed	Improvement, %
RMSE	0.0051	0.0013	74.51
RMSPE	0.0029	0.0014	51.72
MSE	$6 \cdot 10^{-5}$	$2.7 \cdot 10^{-6}$	95.43
MAE	0.0023	0.0012	47.83
MAPE	0.0029	0.0014	51.72
MASE	0.2270	0.1141	49.74
MSPE	$8.5 \cdot 10^{-5}$	$4.1 \cdot 10^{-6}$	95.19
NS	0.9335	0.9976	96.39
corr	0.976	0.9997	98.75
rank	1.786	1.214	72.77

5   CONCLUSION

We have introduced a new signal decomposition method based on the non-negative factorization of the Hankel matrix of a time series. The method offers some advantage over its conceptually similar counterpart, Empirical Mode Decomposition (EMD), by allowing to obtain the desired number of Intrinsic Mode Functions (IMF). The results of decomposition have been used to generate surrogate time series by prediction error minimization driven random cointegration of IMFs. In this paper, cointegration, to our knowledge, has been applied for the first time to IMFs, ever. The surrogate time series used as input allows for more accurate prediction of time series than predicting directly from the original time series. The results obtained using historical stock data and the ARX prediction model show a mean improvement of accuracy of 51-98% depending upon considered accuracy metric, while probability to achieve higher accuracy is 70-91% when compared to ARX prediction using the original time series only.

Future work will focus on the application of the proposed method on other datasets as well as on comparison with other time series modelling and prediction methods.

REFERENCES

1. Castellini, A., Paltrinieri, D., Manca, V.: MP-GeneticSynth: inferring biological network regulations from time series. Bioinformatics 31(5): 785-787 (2015)
2. Maharaj, E.A., Alonso, A.M.: Discriminant analysis of multivariate time series: Application to diagnosis based on ECG signals. Computational Statistics & Data Analysis 70: 67-87 (2014)
3. Nunnari, G.: Modelling air pollution time-series by using wavelet functions and genetic algorithms. Soft Comput. 8(3): 173-178 (2004)
4. Messaoud, A., Weihs, C., Hering, F.: Nonlinear Time Series Modelling: Monitoring a Drilling Process. In: From Data and Information Analysis to Knowledge Engineering, Studies in Classification, Data Analysis, and Knowledge Organization, 31, 302-309 (2006)
5. Soubeyrand, S., Morris, C.E., Bigg, E.K.: Analysis of fragmented time directionality in time series to elucidate feedbacks in climate data. Environmental Modelling and Software 61: 78-86 (2014)
6. Soto, R., Núñez, G.: Soft Modelling of Financial Time Series. Modelling and Simulation 2003: 537-542 (2003)
7. Wei, B., Pinto, H., Wang, X.: A Symbolic Tree Model for Oil and Gas Production Prediction Using Time-Series Production Data. IEEE International Conference on Data Science and Advanced Analytics (DSAA), 272-281 (2016)
8. Hülsmann, M., Borscheid, D., Friedrich, C.M., Reith, D.: General Sales Forecast Models for Automobile Markets Based on Time Series Analysis and Data Mining Techniques. Proc. of the 11th international conference on Advances in data mining: applications and theoretical aspects (ICDM), 255-269 (2011)
9. Chen, Y.-S., Cheng, C.-H., Tsai, W.-L.: Modeling fitting-function-based fuzzy time series patterns for evolving stock index forecasting. Applied Intelligence, 41(2): 327-347 (2014)
10. Cheng, C., Sa-Ngasoongsong, A., Beyca, O., Le, T., Yang, H., Kong, Z., Bukkapatnam, S.T.S.: Time series forecasting for nonlinear and non-stationary processes: A review and comparative study. IIE Transactions, 47(10), 1053–1071 (2015).

11. Dagum, E. B., Time series modeling and decomposition, Statistica, 4, 433-457 (2010)
12. Wold, H.: A Study in the Analysis of Stationary Time Series (1954)
13. Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I.: Stl: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics, 6(1):3–73 (1990)
14. Winters, P.R.: Forecasting Sales by Exponentially Weighted Moving Averages. Management Science. 6:324–342 (1960)
15. Wozniak, M., Napoli, C., Tramontana, E., Capizzi, G.: A multiscale image compressor with RBFNN and Discrete Wavelet decomposition. International Joint Conference on Neural Networks (IJCNN), 1219–1225 (2015)
16. Capizzi, G., Napoli, C., Bonanno, F.: Innovative Second-Generation Wavelets Construction With Recurrent Neural Networks for Solar Radiation Forecasting. IEEE Transactions on Neural Networks and Learning Systems 23(11): 1805-1815 (2012)
17. Huang, N. E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C.. Tung, C.C., Liu, H.H.: The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis, Proc. of the Royal Society of London A. 454, 903–995 (1998)
18. Tiwari, A.K., Dar, A. B., Bhanja, N., Gupta, R.: A Historical Analysis of the US Stock Price Index using Empirical Mode Decomposition over 1791-2015. Economics Discussion Papers, No 2016-9, Kiel Institute for the World Economy (2016)
19. Xu, M., Shang, P., Lin, A.: Cross-correlation analysis of stock markets using EMD and EEMD, Physica A: Statistical Mechanics and its Applications, 442, 82-90 (2016)
20. Damasevicius, R., Napoli, C., Sidekerskiene, T., Wozniak, M.: IMF remixing for mode demixing in EMD and application for jitter analysis. IEEE Symposium on Computers and Communication (ISCC), 50-55 (2016)
21. Deng, Y., Wang, W.: Boundary processing technique in EMD method and Hilbert transform, Chinese Science Bulletin 46 (11), 257–263 (2001)
22. Wu, Q., Riemenschneider, S. D.: Boundary extension and stop criteria for empirical mode decomposition. Advances in Adaptive Data Analysis, 2(2), 157–169 (2010)
23. Damaševičius, R., Vasiljevas, M., Martišius, I., Jusas, V., Birvinskas, D., Woźniak, M.: BoostEMD: an extension of EMD method and its application for denoising of EMG signals, Electronics and Electrical Engineering, 21(6), 57–61 (2015)
24. Engle, R.F., Granger, C.W.J. Co-integration and error correction: Representation, estimation and testing, Econometrica, 55 (2): 251–276 (1987)
25. Barghouthi, S.A., Rehman, I.U., Rawashdeh, G.: Testing the efficiency of Amman Stock Exchange by the two step regression based technique, the Johansen multivariate technique cointegration, and Granger causality. Electronic Journal of Applied Statistical Analysis, 09(03), 572-586 (2016)
26. Schreiber, T., Schmitz, A.: Surrogate time series. Journal of Physics D: Applied Physics 142, 3-4, 346-382 (2000)
27. Karaboga, D., Basturk, B.: A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. Journal of Global Optimization, 459-471 (2007)
28. Box, G., Jenkins, G.M., Reinsel, G.C.: Time Series Analysis: Forecasting and Control, 4th ed. Wiley, Chichester (2008)
29. Taylor, K.E.: Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research 106: 7183–7192 (2001)
30. Jolliff, J.K., Kindle, J.C., Shulman, I., Penta, B., Friedrichs, M.A.M., Helber, R., Arnone, R.A.: Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. Journal of Marine Systems 76: 64–82 (2009)