Empirical Mode Decomposition with Envelope Extraction and LSTM for univariate time series forecasting

Mohamed NDIAYE
CReSTIC
Reims Champagne Ardenne University
Reims, France
mohamed.ndiaye@etudiant.univ-reims.fr

Mamadou MBOUP
CReSTIC
Reims Champagne Ardenne University
Reims, France
mamadou.mboup@univ-reims.fr

***Abstract*—This paper proposes a forecasting method for univariate time series based on a combination of Long-Short-Term-Memory (LSTM) Neural Network and Empirical Mode Decomposition (EMD) with envelope extraction. The basic steps of the method are 1) to split each Intrinsic Mode Function (IMF) into its envelope and detail, 2) to forecast separately these two components for each IMF, using LSTM and 3) to recombine the different forecasts. As already proven, the use of EMD in a forecast setting improves the accuracy. In this paper, we show that the splitting step on each IMF provides further significant improvement on the forecast. This is developed to be applied to treasury transactions forecast. The method relies on well known and easily implementable techniques and is usable with most forecasting algorithms. We experimented it on 3 non-stationary time series, showing its portability to a wide variety of univariate time series.**

***Index Terms*—Univariate time series, Forecast, Signal Decomposition, EMD, Envelope Extraction, LSTM**

I. Introduction

Time series analysis allows extracting insights for understanding the process generating the time series or to draw inference [1]. The latter task holds a strategic value in a diversity of fields. Effectively, forecasting is a broadly explored subject in the literature. In practice, most time series data are utterly complex. Complexity can be described as an intermediate condition between completely deterministic systems and completely random systems [2]. Researchers often tackle complex signal forecasting by improving the forecasting algorithm or trying to simplify the signal, whether it is through a pre-treatment to reduce noise or through decomposition under less complex sub-signals [2], [3]. Several algorithms have proven efficient in correctly forecasting real world activities generated time series. In retailing, ARIMA(p, d, q) models to manage supply chain has been explored [4] [5]. In economy and finance, neural networks among other algorithms are broadly used by researchers and corporates to forecast investment returns, economic indicators such as Unemployment, Gross Domestic Product or inflation [6].

However, data availability and practitioners knowledge to identify relevant variables are important in a corporate setting. The researches leading to this paper revolve around answering the following question: Is it possible, based solely on treasury transactions of a given company, to be able to forecast the returns on its operational activity and be able to construct efficient decision-making tools? Indeed, treasury management is a field where transactions operated by a company are processed. This department monitors the cash flows. As a result, they possess large quantities of data relevant to make forecasts on the future cash flows of a given company. Those data can be retrieved on several sources, particularly the Treasury Management System's databases. It is therefore interesting for a Treasury Management System editor to focus on how to produce these forecasts for their corporate customers. Accessing to relevant data is time-consuming and often requires continuous interactions with the treasury teams to identify exogenous features related to the company's activity. One way to lower these dependencies would be to develop a forecasting routine which rely on univariate time series of all operational transaction operated by the company (i.e. without exogenous variables). This routine should be able to produce robust and accurate forecast for a wide variety of univariate time series, as it should apply to different customers from different activities.

To make the forecasts, we considered the approach consisting of decomposing the univariate time series into lower complexity sub-signals using Empirical Mode Decomposition (EMD). The key contributions consist of computing and forecasting the envelope and detail component of each Intrinsic Mode Function (IMF) produced by the EMD. We show that this further decomposition step reduces the error metrics over the datasets and produces more accurate forecast on a one-step ahead setting.

This paper focuses on the development of a full pipeline of one-step ahead forecast using EMD and envelope Extraction coupled with Long-Short Term Neural Networks (LSTM) [7].

II. On Empirical Mode Decomposition

When it comes to decomposing a signal for forecasting purposes, lots of approach can be experimented. Early attempts looked to extract basic components such as trend

1946-0759/23/$31.00 ©2023 IEEE
DOI 10.1109/ICMLA58977.2023.00070

and seasonality [4]. Other methods, borrowed from signal decomposition techniques, such as Fourier Decomposition [8], [9] Wavelet decomposition [10] [11] and EMD [12] have produced interesting results.

A. EMD Algorithm and IMF characteristics

EMD is a well known data driven and multiscale approach to empirically analyze time series [12]. It allows mitigating the complexity of a given signal by decomposing it into several lower complexity sub-signals, known as IMF. The IMF satisfies two conditions: a) The number of extrema and the number of zero crossings must be equal or differ at most by one. 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 IMF are obtained through the following procedure:
Initialization: given the signal x_t to decompose, let $\psi_t^{(k)} = x_t$
Step 1: Identify the extrema (local maxima and minima) of the observed signal x_t
Step 2: Interpolate the local extrema using cubic spline to obtain the upper $s_t^{(+)}$ and the lower $s_t^{(-)}$ envelopes.
Step 3: Calculate the local mean value of the upper and the lower envelopes $m_t = \frac{s_t^{(+)}+s_t^{(-)}}{2}$
Step 4:Subtract the local mean m_t from the original signal such that:
$\psi_t^{(k+1)} = \psi_t^{(k)} - m_t$
Step 5: Repeat steps 1 through 4 on $\psi_t^{(k+1)}$ until it satisfies the condition of the IMF. Then, record $imf_t^{(i)} = \psi_t^{(k+1)}$
Step 6: Let $r_t^{(k)} = x_t - imf_t^{(i)}$, if the number of extrema of $r_t^{(k)}$ is greater than 3, let $i = i + 1$, $k = 0$ and restart from Step 1, stop the swifting process otherwise.
The signal x_t can be recontructed through the following formula::

$$x_t = \sum_k \psi_t^{(k)} + r_t \tag{1}$$

Where $\psi_t^{(k)}$ are the IMF and r_t is the residual of the decomposition.

The original EMD, however, can in some cases pose challenges (mode-mixing, envelope extraction [13] etc.) resulting in different attempts to solve these problems through other EMD-based signal decomposition methods such as: EEMD [14] or CEEMDAN [15] etc. For a forecasting perspective, researchers coupled different varieties of EMD algorithms with machine learning models resulting in the: EMD-FNN [16], CEEMDAN-LSTM [17], IEMD-BaNN [18]. In all cases, whether it's through improved versions of the Original EMD algorithm or more advanced forecasting algorithms, the results tend to prove that decomposing the time series with EMD and forecasting each component improves the overall accuracy of the forecasts.

B. Envelope Extraction

In practice, it appears that some of the first extracted IMFs still presents complexity and are harder to forecast compared to the last ones as shown in Figure 1.

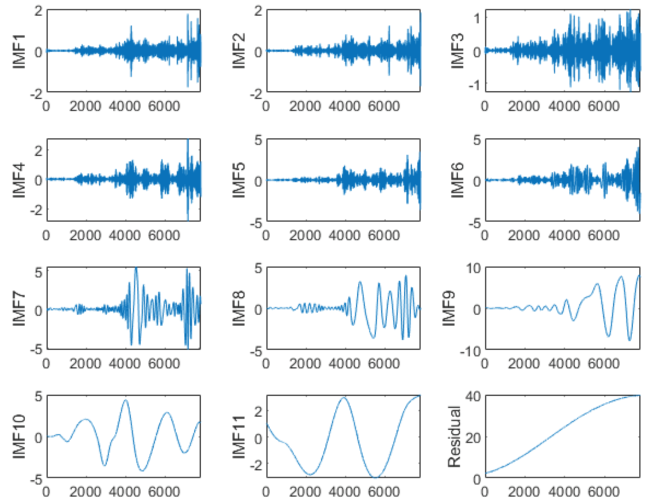

Fig. 1. Empirical Mode Decomposition of Total Energies Stock Market Price

This is why we will further decompose the IMFs.
By definition, IMFs are oscillatory functions with quasi-symmetric upper and lower envelopes [12]. Therefore, we can decompose each IMF $\psi_t^{(k)}$, $i = 1, 2, \ldots$ into two subcomponents which are its envelope $E_t^{(k)}$, obtained by cubic spline interpolation [19], and a detail component $D_t^{(k)}$, obtained through an element-wise division, such that:

$$\psi_t^{(k)} = E_t^{(k)} \cdot D_t^{(k)}, k = 0, 1, \ldots, N$$

where N is the number of IMFs.

As we can see in figure 2, the envelopes are low frequency signal which are quite easy to forecast and the details are signals exhibiting harmonic behavior and oscillating between -1 and 1, they are simpler than the IMFs and therefore easier to forecast.

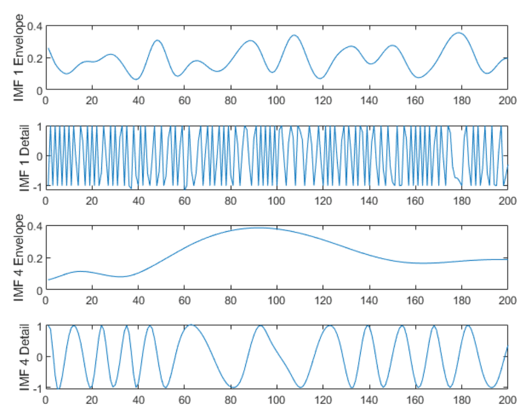

Fig. 2. Sample of Envelope and Detail on TTE IMFs

Doing so, we achieve a decomposition which produces low complexity signals that are more manageable compared to tackling directly the IMFs.

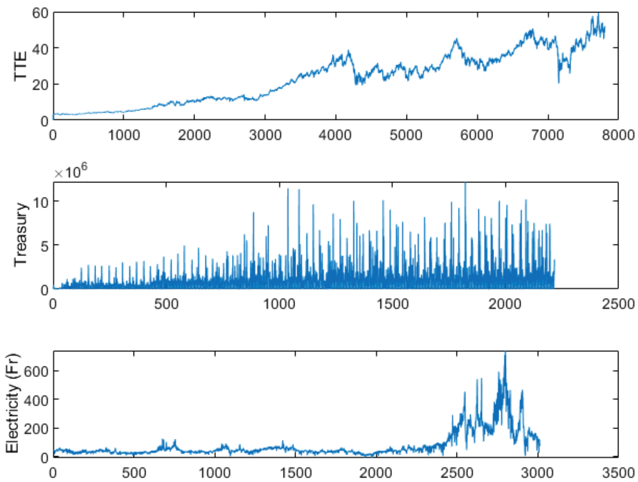

Fig. 3. Datasets

III. Data

We experimented the proposed methodology on several datasets, shown in Figure 3.

All forecasts made are one-step ahead forecasts.

A. Overview

1) Treasury Transactions: This dataset contains operational cash flows (Generated by the operational activities of the given company and not related with any financial operation such as loans or account balancing etc.) of a European company. The data are extracted from a treasury database. They represent the daily credit transactions received by the company through their operational activities. It consists of 2215 data points (around 6 years of history).
The EMD resulted into 16 IMFs and 1 residual.

2) Total Energy Stock Market: This paper also showcases the results of the forecasts on TotalEnergies (TTE) stock prices. The signal consist of 7804 data points (around 21 years of history) of the Total Energies stock market daily closing prices.
The EMD resulted into 11 IMFs and 1 residual, with an example shown in Figure 1.

3) Wholesale Day-ahead France's electricity prices: Finally, we forecast the average daily wholesale day-ahead electricity prices for France. the data set consist of 3012 data points (around 8 years of history)
The EMD resulted into 12 IMFs and 1 residual.

B. Stationarity

We ran Augmented Dickey Fuller (ADF) [20] (H_0 : Presence of a unit root thus the time series is non-stationary) and Kwiatkowski-Phillips-Schmidt-Shin(KPSS) [21] (H_0 : The time series is stationary) tests to study the stationarity of the time series. Table I summarizes the statistics of both tests throughout the used datasets. The conclusions of the test can be obtained from the p-values, a p-value < 0.05 (confidence threshold $\alpha = 5\%$) indicates the rejection of the H_0 hypothesis thus the fact that the series is stationary or non-stationary.

Test	p-value	Statistics	Stationary
Total Energy Stock Market			
ADF	0.856	-0.661	No
KPSS	0.010	19.55	No
Electricity Wholesale			
ADF	0.032	-3.035	Yes
KPSS	0.010	4.174	No
Treasury Data			
ADF	0.002	-3.812	Yes
KPSS	0.10	0.016	No

TABLE I
STATIONARITY STUDY

The TTE time series is non-stationary, the signal furthermore exhibits an upward trend (Figure 3). Model-based methods as ARIMA would not be ideal for forecasting purposes. Results for Electricity wholesale and Treasury data are however not straightforward, as the tests provide contradictory results. These are the sign of difference-stationarity [21], we confirmed this character by differencing the data to arrive to the conclusion that the data are stationary at order 1 differentiation. Both data exhibit an increase in volumes of transactions or prices over time (Figure 3). ARIMA(p, d, q) models might be suited [1], however as stated in the introduction, we chose to go for data driven models with fewer restrictions on stationarity.

C. Data transformation for deep learning

Turning a time series forecasting task into a supervised learning task is the prerequisite to apply machine learning and deep learning algorithms. Given a sequence $\{x_t\}$, framing a one-step ahead, univariate forecasting task consist of choosing a lag l, such that $[x_{t-l}, x_{t-l+1}, ..., x_t]$ is the input sequence for forecasting x_{t+1}. We can also look at it as having a fixed length window of size l sliding through the sequence and generating a sample in which the value inside the window are the features input and the first value outside the window in the order of the sequence is the target value, as shown in Figure 4.

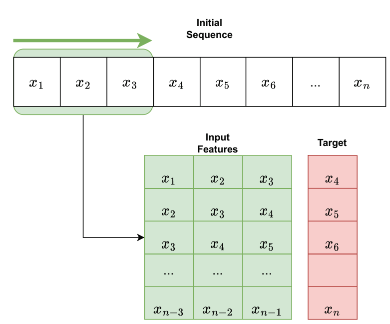

Fig. 4. Time series sequence to supervised learning framework with $l = 3$

IV. Methodology

We performed the one-step ahead forecast tasks through the following pipeline:

1) Establish the baseline results by directly forecasting the univariate time series with LSTM Neural Networks
2) Perform Empirical Mode Decomposition for the original time series, then tackle each component (i.e. IMF) of the resulting decomposition separately.
   - Produce forecast for each IMF using LSTM Neural Network. The final forecast is then obtained by recombining through the following formula:

$$\hat{y}^{t+1} = \sum_{k=1}^{n} \hat{\psi}_k^{t+1}$$

   where $\hat{y}^{t+1}$ is the one step ahead forecast of the time series, $\hat{\psi}_k^{t+1}$ is the one step ahead forecast of the k^{th} IMF
   - Decompose each IMF into an envelope E_k and a detail D_k and produce forecasts for each envelope and detail of each IMF using LSTM Neural Network. The final forecast is then obtained through the following formula:

$$\hat{y}^{t+1} = \sum_{k=1}^{n} (\widehat{E}_k^{t+1} \cdot \widehat{D}_k^{t+1})$$

   where $\hat{y}^{t+1}$ is the one step ahead forecast of the time series, $\widehat{E}_k^{t+1}$ is the one step ahead forecast of the k^{th} envelope and $\widehat{D}_k^{t+1}$ is the one step ahead forecast of the k^{th} detail.
3) We compute the relevant error metrics and selected the final architecture and model through expanding windows walk forward validation [22] through the following scheme: We expand the training set by a constant number of data points and slide a fixed size testing window to compute the error metrics until full coverage of the datasets, as shown in Figure 5.

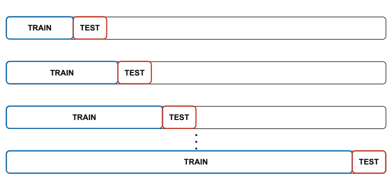

Fig. 5. Expanding Window Walk Forward

We chose the following error metrics:
   - R^2 Score:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

   This metric is a normalized Mean Squared Error computed as a percentage. It is suited to compare regression models. The higher the R2 score the better the model.
   - Mean Absolute Percentage Error (MAPE):

$$\mathsf{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{|y_i|}$$

   We considered its complement, $1 - \mathsf{MAPE}$, which is interpreted as a (percentage) metric of accuracy of forecasting rather than a (percentage) error.
   - Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

   - Rooted Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

   Across the given metrics, the y_i are the observed values on the test data set, the $\hat{y}_i$ are the forecasted values, $\bar{y}$ is the mean over the observed values in the test data set, n is the size of the test data set.

Extracting envelopes and details is not relevant for some IMFs as they already have low complexity. In the trials, we did not perform Envelope Extraction on the residual.
We used 2 layered LSTM with RelU activation ($f(x) = \max(0, x)$) over training samples lagged by 32 data points. The other parameters, such as the number of units per layer, batch sizes and callbacks were set differently from one dataset to another.
The pipeline we developed to obtain and compare the different approaches is shown in Figure 6.

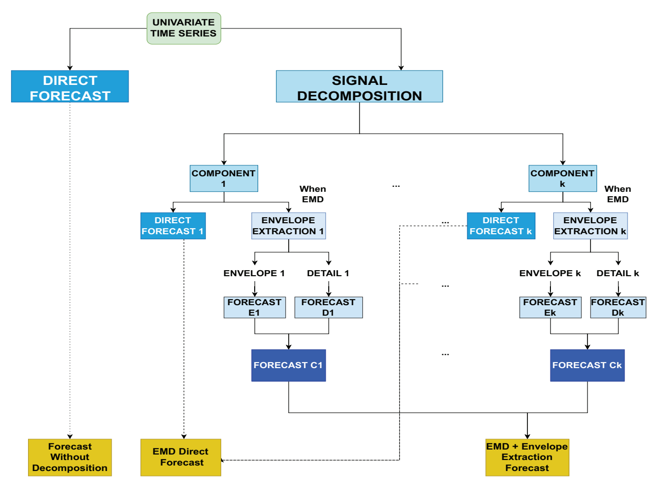

Fig. 6. Full Methodology Pipeline

V. Results and Discussion

The table II contains the result in regard to the Mean Absolute Percentage Error (MAPE) for the TTE and Electricity dataset, and the R_2 score for the Treasury transaction dataset (This dataset contains zeros which MAPE is sensitive to, explaining why this metric is excluded). As expected, in the trials, decomposing the signal through EMD improves the forecasts according to the metrics we consider. The figures 7, 8 and 9 show a clear improvement from the direct LSTM to

EMD method when performing a statistical summary of the expanding windows of the walk forward validation method. These experiments confirm results obtained by previous researches [18].

Strategy	Min	Median	Q3	Max
Total Energy Stock Market (1-MAPE)				
Direct LSTM	0.902	0.963	0.975	0.983
EMD + LSTM	0.934	0.978	0.984	0.989
Env. Ext + LSTM	0.947	0.989	0.992	0.994
Electricity Wholesale (1-MAPE)				
Direct LSTM	0.330	0.834	0.882	0.916
EMD + LSTM	0.735	0.848	0.891	0.930
Env. Ext + LSTM	0.564	0.892	0.912	0.957
Treasury Data (R_2 Score)				
Direct LSTM	0.100	0.273	0.360	0.474
EMD + LSTM	0.300	0.538	0.642	0.726
Env. Ext + LSTM	0.458	0.576	0.721	0.799

TABLE II
METHODS COMPARISON

The envelope extraction method produced better results overall in all the metrics considered throughout the different datasets (see Figures 7, 8, 9).

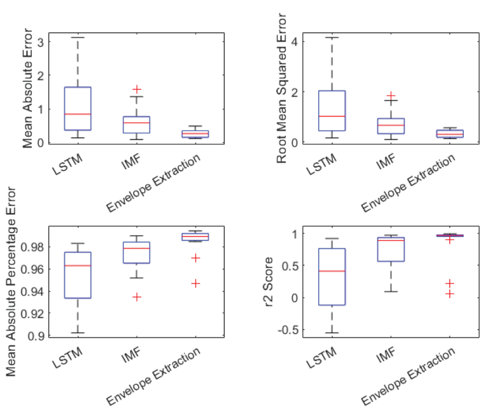

Fig. 7. TTE: LSTM vs IMF vs Envelope Extraction

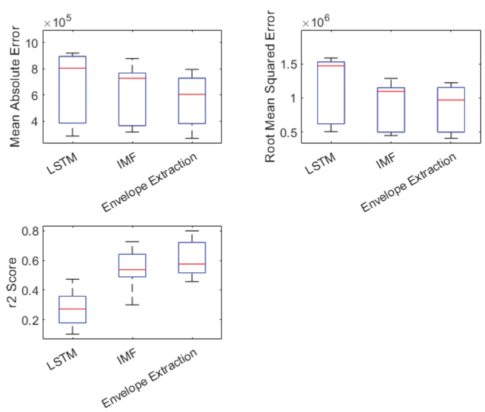

Fig. 8. Treasury Data: LSTM vs IMF vs Envelope Extraction

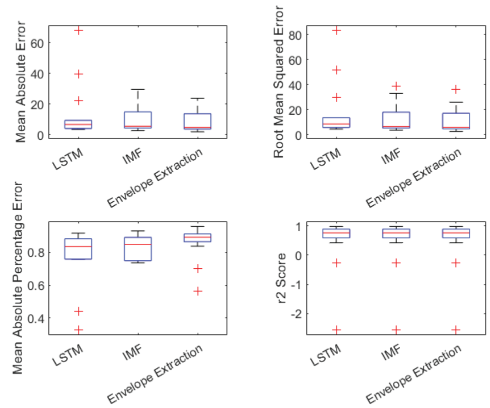

Fig. 9. Electricity Wholesale Data: LSTM vs IMF vs Envelope Extraction

Furthermore, the envelope extraction method produces forecast which are robust through time. Indeed, Figure 7 shows that error metrics throughout the forecasting windows are all above 94,7% in terms of MAPE (1-MAPE precisely) and Figure 10 shows a good fit on the test datasets, window by window.

However, in these results, we deliberately used the same forecasting algorithm for all sub-signal obtained from the decomposition. This approach, as efficient as it can be, presents the flaw of not considering sub-signals nature and choosing different more adapted algorithms for each one of them. As a result, we have leverages to obtain better results by computing different forecasting algorithms for each component of the decomposition, whether it's direct forecast on the IMF or the envelope and detail extraction. We can obtain improved results on the recombined forecast on each sub-signal by selecting

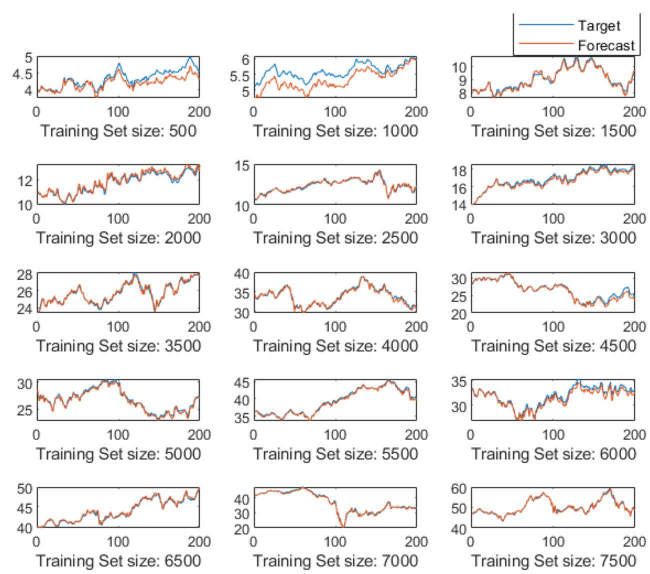

Fig. 10. TTE: Envelope Extraction Forecast walk-forward validation target vs forecast on test datasets

for each of them the most accurate forecasting algorithm, compared through the same process used above with the metrics considered.

We note that the expanding window walk-forward scheme can also provide insights about the depth of history needed to obtain accurate forecasts, as the accuracy often improves when the training window is larger.

The tradeoff with this methodology is the resources needed. The EMD and the envelope extraction produce several IMF and 2 sub-signals per IMF, which result in multiple signals to be forecasted. In addition to that, the expanding window walk forward has then to be performed on each signal, multiplying the training and testing phase. This introduces the requirement, to have efficient computation schemes as parallel computing in order for this methodology to be maintainable in a high volume data setting.

VI. Conclusion

We demonstrated in this paper that our approach of using EMD coupled with envelope extraction can significantly improve the accuracy of the forecasts. We can, furthermore, complete the forecasting pipeline by implementing forecast combination algorithms. The proposed global pipeline has the advantage of being generic enough to accommodate a wide variety of univariate time series, thus mitigating the need to find additional explanatory variables.

We can explore the following possibilities to build an improved pipeline:

- Analyzing further the information carried by each Sub-signal after the decomposition and chose the forecasting algorithms accordingly.
- Update the pipeline with novel forecasting algorithms
- Implement multistep ahead forecasting methods
- Consider multivariate settings
- Leverage on the usage of several different signal decomposition and forecasting algorithms.
- Develop advanced forecast combination techniques

Acknowledgment

We extend our heartfelt thanks to the company Diapason for providing invaluable data and domain analysis, which greatly enriched this paper.

References

[1] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, ser. Springer Texts in Statistics. Springer International Publishing, 2016. [Online]. Available: https://doi.org/10.1007/978-3-319-29854-2

[2] J. Dong, W. Dai, L. Tang, and L. Yu, "Why do emd-based methods improve prediction? a multiscale complexity perspective," *Journal of Forecasting*, vol. 38, no. 7, pp. 714–731, 2019.

[3] C. Qin, G. Shi, J. Tao, H. Yu, Y. Jin, D. Xiao, Z. Zhang, and C. Liu, "An adaptive hierarchical decomposition-based method for multi-step cutterhead torque forecast of shield machine," *Mechanical Systems and Signal Proc.*, vol. 175, p. 109148, 2022.

[4] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.

[5] G. Kenneth, "An arima supply chain model," *Management Science*, vol. 51, no. 2, pp. 305–310, 2005. [Online]. Available: http://www.jstor.org/stable/20110327

[6] H. Ghoddusi, G. G. Creamer, and N. Rafizadeh, "Machine learning in energy economics and finance: A review," *Energy Economics*, vol. 81, pp. 709–727, 2019.

[7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, p. 1735–1780, nov 1997.

[8] W. Cochran, J. Cooley, D. Favin, H. Helms, R. Kaenel, W. Lang, G. Maling, D. Nelson, C. Rader, and P. Welch, "What is the fast fourier transform?" *Proc. of the IEEE*, vol. 55, no. 10, pp. 1664–1674, 1967.

[9] D. Song, A. M. Chung Baek, and N. Kim, "Forecasting stock market indices using padding-based fourier transform denoising and time series deep learning models," *IEEE Access*, vol. 9, pp. 83 786–83 796, 2021.

[10] S. Soltani, "On the use of the wavelet decomposition for time series prediction," *Neurocomputing*, vol. 48, no. 1, pp. 267–277, 2002.

[11] F. Wang, Y. Yu, Z. Zhang, J. Li, Z. Zhen, and K. Li, "Wavelet decomposition and convolutional lstm networks based improved deep learning model for solar irradiance forecasting," *Applied Sciences*, vol. 8, no. 8, 2018.

[12] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[13] X. Hu, S. Peng, and W.-L. Hwang, "Emd revisited: A new understanding of the envelope and resolving the mode-mixing problem in am-fm signals," *IEEE Transactions on Signal Proc.*, vol. 60, no. 3, pp. 1075–1086, 2012.

[14] S. G., "A new ensemble empirical mode decomposition (eemd) denoising method for seismic signals," *Energy Procedia*, vol. 97, pp. 84–91, 2016, european Geosciences Union General Assembly 2016, EGU Division Energy, Resources the Environment (ERE).

[15] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *2011 IEEE International Conference on Acoustics, Speech and Signal Proc. (ICASSP)*, 2011, pp. 4144–4147.

[16] G. Zhenhai, Z. Weigang, L. Haiyan, and W. Jianzhou, "Multi-step forecasting for wind speed using a modified emd-based artificial neural network model," *Renewable Energy*, vol. 37, no. 1, pp. 241–249, 2012.

[17] L. Yu, Y. Yan, X. Jiali, L. Ying, and M. Feng, "Forecasting stock index price using the ceemdan-lstm model," *The North American Journal of Economics and Finance*, vol. 57, p. 101421, 2021.

[18] O. Abedinia, M. Lotfi, M. Bagheri, B. Sobhani, M. Shafie-khah, and J. P. S. Catalão, "Improved emd-based complex prediction model for wind power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 4, pp. 2790–2802, 2020.

[19] M. Martin, "Cubic spline interpolation of continuous functions," *Journal of Approximation Theory*, vol. 10, no. 2, pp. 103–111, 1974.

[20] R. Harris, "Testing for unit roots using the augmented dickey-fuller test: Some issues relating to the size, power and the lag structure of the test," *Economics Letters*, vol. 38, no. 4, pp. 381–386, 1992.

[21] D. Kwiatkowski, C. P. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *Journal of Econometrics*, vol. 54, no. 1, pp. 159–178, 1992.

[22] S. Suradhaniwar, S. Kar, S. Durbha, and A. Jagarlapudi, "Time series forecasting of univariate agrometeorological data: A comparative performance evaluation via one-step and multi-step ahead forecasting strategies," *Sensors*, vol. 21, 04 2021.