Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Complete ensemble empirical mode decomposition with adaptive noise integrating feedforward neural network for tourist arrival forecasting

To cite this article: S Herawati *et al* 2022 *J. Phys.: Conf. Ser.* **2193** 012049

Complete ensemble empirical mode decomposition with adaptive noise integrating feedforward neural network for tourist arrival forecasting

S Herawati[1*], YDP Negara[2], M Latif[3]

[1,2]Department of Information Systems, University of Trunojoyo Madura
[3]Department of Mechatronics Engineering, University of Trunojoyo Madura

*Corresponding author : sriherawati@trunojoyo.ac.id

Abstract. The tourism sector has an important role in helping the income of a region, especially for economic development and opportunities to expand employment. However, the trend tourist arrival to these tourist attractions has decreased since the COVID-19 pandemic. The government enforces a new normal policy to reopen tourist attractions by implementing health protocols. Local governments and tourism managers need forecasting of tourist arrivals to help plan the tourism sector in the future and anticipate an increase in tourist arrival. Most tourist arrivals are influenced by several factors, such as : seasonality, politics, disasters, crises, and other important events. One method to accommodate these factors is using Ensemble Empirical Mode Decomposition (EEMD). However, EEMD still produces a mixing mode during decomposition. Complete Ensemble Mode Decomposition with Adaptive Noise (CEEMDAN) is proposed to overcome the weaknesses of EEMD. This research integrates CEEMDAN with Feedforward Neural Network (FNN) in generating forecasts. The experiment results show that the integration of CEEMDAN and FNN can produce good forecasting accuracy.

1. Introduction

Madura is an island that has a diversity of tourism, starting from nature tourism, historical tourism, cultural tourism, and religious tourism. These tours attract tourists both domestic and foreign tourists to visit Madura. This trend of tourist arrivals has increased significantly from year to year. In 2019, the COVID-19 pandemic made a downward trend in line with government policies that temporarily closed tourist attractions. To overcome the impact of the pandemic in the tourism sector, the government seeks to provide economic stimulus and protection programs for workers in the tourism sector. The tourism sector has an important role in helping the income of a region, especially for economic growth [1][2] and opportunities for expanding employment [3].

After that, the government implemented a new normal policy to reopen tourist attractions by implementing health protocols. The issuance of health protocols is applied to the community in the context of preventing and controlling COVID-19 when visiting tourist attractions. Thus, communities around tourist attractions can remain productive and tourists feel safe during the pandemic. These efforts are made to increase the trend of tourist arrivals. The increase in tourist arrivals should be balanced with planning and preparation from local governments and tourism managers so that services for tourists can be fulfilled properly, because the tourism sector is related to other sectors, such as: hospitality [4], transportation, and travel or travel [5]. For future tourism planning, tourism managers need to forecast tourist arrivals. Forecasting is used to monitor fluctuations in tourist arrivals by taking advantage of past events. If the incident reoccurs, then anticipatory measures can be applied.

Many studies on forecasting tourist arrivals have been carried out by several previous studies. Non-linear and non-stationary factors tend to affect fluctuations in tourist arrivals, such as: seasonality, politics, disasters, crises, and other important events[6]. One method to accommodate the complexity of these factors is using Empirical Mode Decomposition (EMD). EMD decomposes the data into three components according to their characteristics, namely short-term trend fluctuations, the effects of significant event shocks, and long-term trends. To generate forecasts, EMD is integrated with Artificial Neural Networks [7]. The research resulted in a fairly good forecasting accuracy. However, EMD sometimes produces mixed modes that do not match the data pattern. To overcome its weakness, Ensemble Empirical Mode Decomposition (EEMD) with the addition of white noise is used in forecasting [8]. The research integrates EEMD with an optimized artificial neural network using a genetic algorithm. This type of artificial neural network uses Feedforward Neural Network (FNN).

However, the EEMD decomposition process still produces a mixing mode, so that the decomposition process is added with adaptive noise using the Complete Ensemble Mode Decomposition with Adaptive Noise (CEEMDAN) method. CEEMDAN reconstructs the errors and patterns of the resulting data [9][10]. Therefore, this study proposes the integration of the CEEMDAN and FNN methods to produce more accurate forecasting of tourist visits. The results of research on forecasting tourist arrivals can help local governments and tourism managers in an effort to improve tourism services.

2. Methods

This study uses the integration of the CEEMDAN and FNN methods for forecasting tourist arrivals. The trial was carried out using a variation of the learning algorithm in FNN. The process forecasting method can be seen in Figure 1. Based on Figure 1, it can be explained that the research method used is as follows:

a. Complete Ensemble Mode Decomposition with Adaptive Noise(CEEMDAN)

CEEMDAN is a decomposition method developed from EEMD with the addition of adaptive noise. CEEMDAN performs decomposition by producing IMFs and Residues. CEEMDAN reconstructs errors during the data decomposition process so that the mixing mode resulting from mixing the decomposition signal can be eliminated. This method adds adaptive noise in the time series data to compensate for the possible mixed modes. The CEEMAND algorithm can be explained as follows:

Step 1 : Ensemble number initialization(M).

Step 2 : Calculate the local EMD value of realization I to get the first residual value

$$x^i = x + \beta_0 E_1(w^{(i)}) \tag{1}$$

$$r_1 = (M(x^{(i)})) \tag{2}$$

Step 3 : For the first value (k=1) performed to calculate the first IMF

$$\tilde{d}_1 = x - r_1 \tag{3}$$

Step 4 : Estimate the second residual value as the average of the local values and define the second IMF for $r_1 + \beta_1 E_2(w^{(i)})$

$$\tilde{d}_2 = r_1 - r_2 = r_1 - (M(r_1 + \beta_1 E_2(w^{(i)}))) \tag{4}$$

Step 5 : For value k=3,..., calculate K for the k-th residual value

$$r_k = (M(r_{k-1} + \beta_{k-1} E_k(w^{(i)}))) \tag{5}$$

Step 6 : Calculate mode for k

$$\tilde{d}_k = r_{k-1} - r_k \tag{6}$$

Step 7 : Repeat step 5 for the next k

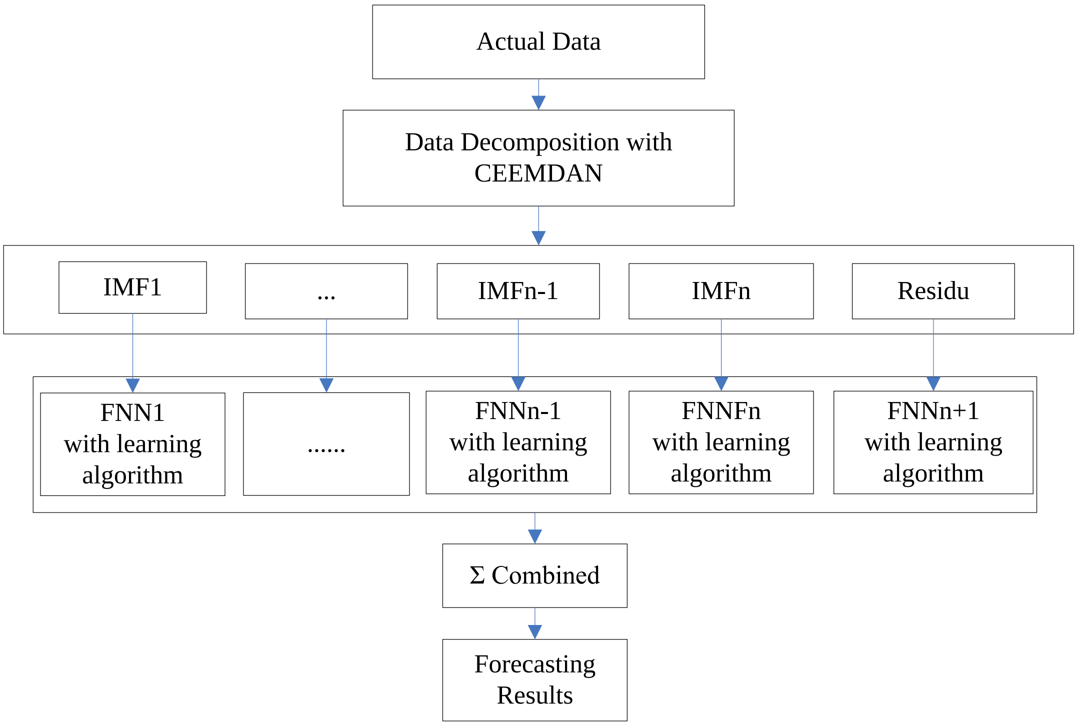

Figure 1. The process forecasting method.

b. Feedforward Neural Network (FNN)

FNN is a type of artificial neural network. Each neuron in FNN can have multiple inputs and one output. The input path on a neuron can contain raw data or data from the processing of previous neurons. Meanwhile, the output of a neuron can be a final result or an input for the next neuron. The input to the network will be processed by a function that will add up the values of all weights and biases. The result of the sum will be compared with a threshold value through the activation function of each neuron. The input value in the training of the artificial neural network is influenced by the selection of the activation function. In the network learning process, FNN has several types of learning, such as: gradient Descent Backpropagation, Resilient Backpropagation, and others. The algorithm for FNN can be explained as follows:

Step 1 : Initialization of weights and parameters
Step 2 : Each input unit X_i, i=1,….,n receive input signal x_i
Step 3 : Each hidden unit Z_j, j=1,…,p calculate value z_in_j and z_j

$$z\_in_j = v_{0j} + \sum_{i=1}^{n} x_i v_{ij} \tag{7}$$

$$z_j = f(z\_in_j) \tag{8}$$

Step 4 : Each output unit Y_k, k=1,…,m calculate value y_in_k, y_k, δ_k, Δw_{jk}, Δw_{0k}

$$y\_in_k = w_{0k} + \sum_{j=1}^{p} z_j w_{jk} \quad (9)$$

$$y_k = f(y\_in_k) \quad (10)$$

$$\delta_k = (t_k - y_k) f'(y\_in_k) \quad (11)$$

$$\Delta w_{jk} = \alpha \delta_k z_j \quad (12)$$

$$\Delta w_{0k} = \alpha \delta_k \quad (13)$$

Step 5 : Each hidden unit Z_j,j=1,…p calculate value δ_in_j, δ_j, Δv_{ij}, and Δv_{0j}

$$\delta\_in_j = \sum_{k=1}^{m} \delta_k w_{jk} \quad (14)$$

$$\delta_j = \delta\_in_j f'(z\_in_j) \quad (15)$$

$$\Delta v_{ij} = \alpha \delta_j x_i \quad (16)$$

$$\Delta v_{0j} = \alpha \delta_k \quad (17)$$

Step 6 : Each output unit Y_k, k=1,…,m update bias and weight

$$w_{0k}^{(l+1)} = w_{0k}^{l} + \Delta w_{ok} \quad (18)$$

$$w_{jk}^{(l+1)} = w_{jk}^{l} + \Delta w_{jk} \quad (19)$$

Step 7 : Each hidden unit Z_j, j=1,…,p update bias and weight

$$v_{0j}^{(l+1)} = v_{0j}^{l} + \Delta v_{0j} \quad (20)$$

$$v_{ij}^{(l+1)} = v_{ij}^{l} + \Delta v_{ij} \quad (21)$$

Step 8 : Repeat step 2 to step 7, until the test condition stops, it receaches the error level and maximum epoch.

This study evaluates the performance of forecasting results with Mean Square Error (MSE) and Correlation coefficient (R). The MSE and R equations are shown in (22) and (23). MSE is used to measure the deviation between actual data values and forecasting results. The smaller the MSE value, the closer the forecasting results and actual data. Meanwhile, the R value is used to measure the correlation between the actual data values and the forecasting results.

$$MSE = \frac{1}{n} \sum_{t=1}^{n} (x_t - y_t)^2 \quad (22)$$

$$\mathbf{R} = \frac{\sum_{i=1}^{n} x_t x y_t}{\sqrt{\sum_{i=1}^{n} x_t^2} \, x \sqrt{\sum_{i=1}^{n} y_t^2}} \quad (23)$$

Where n represents the amount of actual data, xt represents actual data, t is time, yt represents forecasted data.

3. Results And Discussions

3.1. Experiment Data
This research was conducted using actual time series data on tourist arrivals in Sumenep. The trial data was taken from data on tourist arrivals to Sumenep from January 2015 to December 2019. The data used a composition of 70% training data and 30% testing.

3.2. Research Experiment Results
The actual data of tourist arrivals is decomposed using CEEMDAN. The parameter uses an ensemble of 500 and a standard deviation of 0.2. This decomposition produces five IMFs and residues which can be seen in Figure 2. The CEEMDAN method produces a fixed amount of decomposition even though the process is repeated several times. Meanwhile, EEMD without additional adaptive noise produces an arbitrary amount of decomposition. The results of the decomposition using EEMD can be seen in Figure 3.

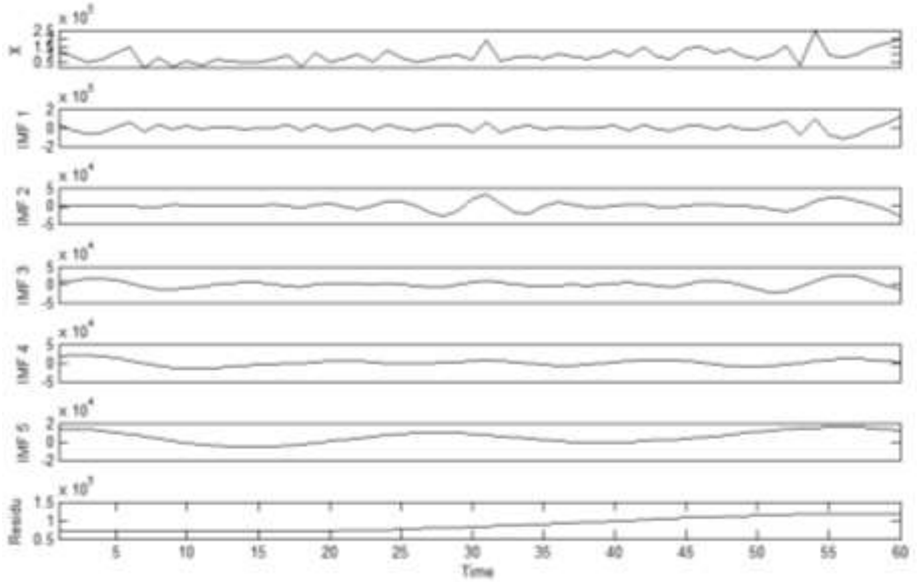

Figure 2. Decomposition with CEEMDAN

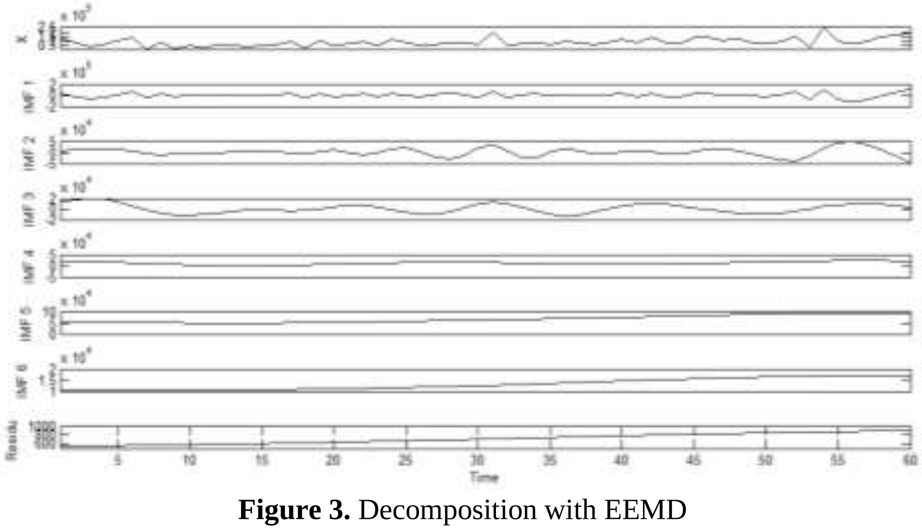

Figure 3. Decomposition with EEMD

After that, the normalization process is carried out for all IMF and the residue resulting from the decomposition. This process is carried out to match the binary sigmoid activation function in FNN. FNN is used in training and testing the decomposition data. The FNN parameter uses 10000 epochs of repetition, the learning rate is 0.1, and the error tolerance is 0.0001. The trial scenario was carried out with several learning algorithms in FNN, such as: Gradient Descent Backpropoagation (GDBP), Gradient Descent with Adaptive Learning Rate Backpropagation (GDALR), Levenberg - Marquardt Backpropagation(LMBP), Gradient Descent with Momentum Backpropagation (GDM), Gradient Descent with Momentum and Adaptive Learning Rate Backpropagation (GDMALR), Batch training with weight and bias learning rules (B), and Conjugate gradient backpropagation with Polak Ribiére(PCG). The data pattern uses 3-30-1 (three input neurons, thirty hidden neurons and one output neuron). The results of forecasting using the integration of the CEEMDAN-FNN method were compared with the integration of the EEMD-FNN method using a variation of learning algorithms. The performance of the forecasting results was measured using MSE and R which can be seen in Table 1.

Table 1. Comparison of forecasting results between CEEMDAN-FNN dan EEMD-FNN

Algorithm	CEEMDAN-FNN		EEMD-FNN	
	MSE	R	MSE	R
GDBP	0,0146	0,9682	0,0271	0,9401
PCG	0,0173	0,9622	0,0314	0,9301
B	0,0148	0,9677	0,0230	0,9494
GDMALR	0,0115	0,9751	0,0197	0,9569
LMBP	0,0128	0,9722	0,0326	0,9273
GDALR	0,0116	0,9748	0,0223	0,9511
GDM	0,0126	0,9726	0,0241	0,9469

Based on the experiments in Table 1, it can be seen that the integration of the CEEMDAN-FNN method as a whole for the learning algorithm produces the smallest MSE value and the R value is getting closer to 1. Meanwhile, the best performance uses the GDMALR algorithm with an MSE value of 0.0115 and an R value of 0, 9751. The closeness can be seen the actual data and the forecasting results in Figure 4. From the picture, it can be explained that the results of the CEEMDAN integration forecasting are getting closer to the actual data of tourist arrivals.

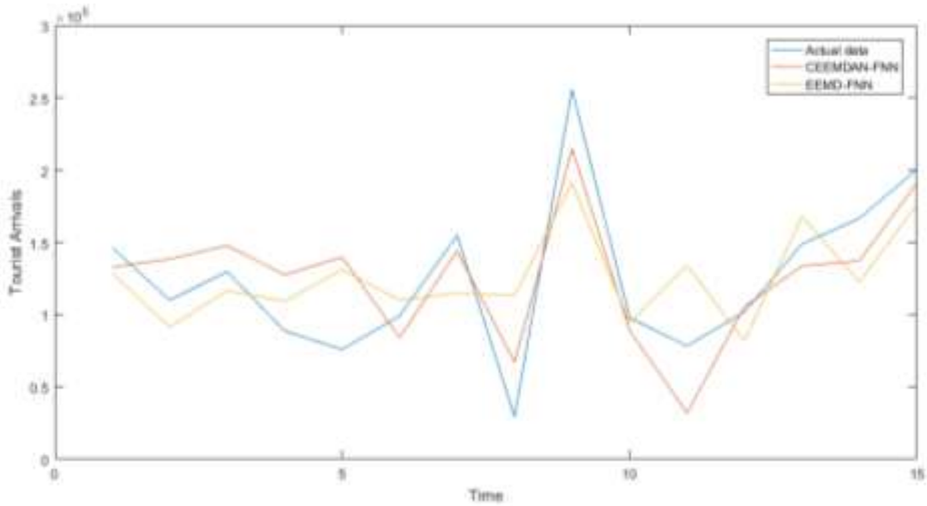

Figure 4. Compare Actual data and forecasts tourist arrivals to Sumenep

Journal of Physics: Conference Series **2193** (2022) 012049 doi:10.1088/1742-6596/2193/1/012049

4. Conclusions

This study has proposed a new model to generate accurate forecasting of tourist arrivals using the CEEMDAN and FNN methods. The trial scenario was carried out with a variation of the learning algorithm. Then, the forecasting results are compared with the integration of the EEMD and FNN methods. The best performance uses the GDMALR algorithm with an MSE value of 0.0115 and an R value of 0.9751. The trial also shows that the integration of the CEEMDAN-FNN method is better than the EEMD-FNN.

References

[1] Hassani H., Webster A., Silva E S and Heravi S 2015 Forecasting U.S. Tourist arrivals using optimal Singular Spectrum Analysis *Tourism Management 46* 322–335
[2] Xie G, Qian Y and Wang S 2020 A decomposition-ensemble approach for tourism forecasting *Annals of Tourism Research 81* 102891
[3] Pai P -F, Hung K -C and Lin K -P 2014 Tourism demand forecasting using novel hybrid system *Expert Systems with Applications* 41 3691 - 3702
[4] Guizzardi A and Stacchini A 2015 Real-time forecasting regional tourism with business sentiment surveys *Tourism Management* 47 213-223
[5] Law R, Li G, Fong D K and Han X 2019 Tourism demand Fforecasting : A deep learning approach *Annals of Tourism Research* 75 410-423
[6] Zhang H, Song H, Wen L and Liu C 2021 Forecasting tourism recovery amid COVID-19 *Annals of Tourism Research 87* 103149
[7] Shao-Jiang L, Jia-Ying C and Zhi-Xue L 2018 A EMD-BP integrated model to forecast tourist number and applied to Jiuzhaigou *Journal of Intelligent and Fuzzy Systems 34*(2) 1045–1052
[8] Herawati S and Latif M 2021 Forecasting tourist visits using data decomposition technique and learning optimization of artificial neural network *IOP Conference Series: Earth and Environmental Science 1811*(1)
[9] Colominas M A, Schlotthauer G and Torres M E 2014 Improved complete ensemble EMD: A suitable tool for biomedical signal processing *Biomedical Signal Processing and Control* 14 19–29
[10] Vanraj, Dhami S S and Pabla B S 2017 Non-contact incipient fault diagnosis method of fixed-axis gearbox based on CEEMDAN *Royal Society Open Science 4*(8)

Acknowledgments

The authors wish to thank University of Trunojoyo Madura for funding this research.