IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 13 February 2023, accepted 25 February 2023, date of publication 6 March 2023, date of current version 10 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3253279

RESEARCH ARTICLE

A New Approach to Empirical Mode Decomposition Based on Akima Spline Interpolation Technique

MUHAMMAD ALI[1], DOST MUHAMMAD KHAN[1], (Senior Member, IEEE), IMRAN SAEED[2], AND HUDA M. ALSHANBARI[3]

[1]Department of Statistics, Abdul Wali Khan University Mardan, Mardan 23200, Pakistan
[2]Centre for Soft and Living Matter, Institute of Basic Sciences, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea
[3]Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia

Corresponding authors: Huda M. Alshanbari (hmalshanbari@pnu.edu.sa) and Dost Muhammad Khan (dostmuhammad@awkum.edu.pk)

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R 299), Princess Nourah Abdulrahman University, Riyadh, Saudi Arabia.

ABSTRACT The objective of this research work is to extend the scope of the empirical mode decomposition (EMD) algorithm, as an efficient tool to decompose the nonlinear and non-stationary time series. For EMD to be widely applicable, the extension utilizes both clean and noisy data sets. When constructing upper and lower envelopes, the proposed algorithm utilizes the Akima spline interpolation technique rather than a cubic spline. The proposed EMD is called Akima-EMD, which is used to identify non-informative fluctuations in the signal, such as noise, outliers, and ultra-high frequency components, and to break down the clean and chaotic data into various components to avoid distortion. It has been shown through synthetic as well as real-world time series data analysis that the proposed method successfully extracts noise in the form of the first IMF from the data.

INDEX TERMS Akima, empirical mode decomposition (EMD), fast Fourier transform (FFT), intrinsic mode functions (IMFs), variational mode decomposition (VMD), complementary ensemble empirical mode decomposition with adaptive noise (CEEMDAN).

I. INTRODUCTION

To examine the composite signal, the usual practice is to break it down into several components ranging from simple to complicated forms. The next step is to retrieve the valuable information concealed in these subcomponents after the signal has been broken down into its many parts so that it can be used for some meaningful analysis or prediction. The advantages of breaking down a signal into distinct parts are as follows:

i. To extract various modes and complexity levels from a complex signal by using the best acceptable technique, which must be suitable for the corresponding signal.
ii. The decomposition is carried out in such a way that information, in the time domain, is transferred into the frequency domain, a different dimension, utilizing all the available information.
iii. Following the deconstruction, each subcomponent is examined so that valuable data can be extracted separately for additional examination and prediction.

A well-known and simple technique for time series analysis in a variety of disciplines, especially in mathematics and statistics is called spectral analysis. Before applying the spectral method, it requires some pre-defined parameter setting such as the collection of linearly dependent vectors, which are used to represent the data. For linear and stationary data analysis, the most common technique is the Fourier transformation [1]. However, we cannot apply the Fourier transformation in situations when the time series is nonlinear and nonstationary. To overcome these limitations researchers in the signal processing areas proposed wavelet transformation (WT) [2], which provides efficient results

The associate editor coordinating the review of this manuscript and approving it for publication was Sajid Ali.

when applied to nonlinear and nonstationary time series. However, the method of WT is based on the Fourier transform except that it uses some data transformation techniques with the help of some pre-selected functions that work well for linear time series. The use of sophisticated and cutting-edge data-driven techniques, such as fast Fourier transformation (FFT), spectral analysis, and wavelet analysis, to break down a linear and stationary signal into its subcomponents is well acknowledged [3], [4], [5], [6], [7].

However, the FFT has certain limitations, such as the requirement that the system is linear and that the data be strictly periodic or stable, without which the resulting spectrum will not make any physical sense. To solve this shortcoming of FFT a unique data-driven approach that uses the EMD to generate a collection of intrinsic mode functions (IMFs). To overcome the limitations of both the Wavelet and Fourier transform a novel method known as EMD has been proposed by [8], to analyze nonstationary, and nonlinear time series. By using the EMD technique, the input signal is divided into several simple oscillatory modes with known frequencies and a singular monotone residue (trend). The method of EMD becomes more powerful for the analysis of any given time series, especially when integrated with Hilbert spectrum analysis (HAS) to extract the instantaneous frequencies, such integration is known as the Hilbert-Huang transform (HHT) [9]. After its first development, it has been widely used in many branches of science and engineering to analyze nonstationary or nonlinear signals [10]. In most situations, this technique is used for the earth sciences [11], [12] but currently, the generality of EMD and HHT can be seen in various fields such as medicines [13], speech recognition [14], quantum systems [15], ocean engineering [16] and finance [17]. Keeping in view the versatile characteristics of EMD, it is probable that this method will find new ways to solve problems in almost every field of science that contains experimental data.

The method of EMD uses cubic spline interpolation to determine the upper, lower, and mean envelopes and, later, the IMFs. When the signal contains noise, and outliers then it is difficult to construct the mean envelopes while using the cubic spline interpolation technique. Furthermore, the cubic spline interpolation technique suffers from the problem of overshoot, which causes these non-informative oscillations to skew the results. The following advantages of the proposed EMD over the traditional EMD have been achieved by using Akima spline interpolation rather than cubic spline interpolation.

1. The proposed method known as Akima-EMD can divide such signals into suitable IMFs without the distortion caused by noise and non-informative random fluctuations such as outliers and ultra-high frequency components. Because the Akima spline interpolation technique is more resilient to noise and non-informative random fluctuations such as outliers and ultra-high frequency components.
2. As a result, it yields consistent decomposition for the entire domain, including boundary areas. Without imposing any condition on the boundary our proposed method solves the limitation of huge wiggles at both ends of the data.

This paper is further divided into the following sections. The theoretical background of the study is presented in section II. We thoroughly explained how Akima spline interpolation could be utilized inside the EMD algorithm to determine the upper, and lower envelopes, followed by the mean envelope as well as IMFs, in section III. Simulation studies are carried out in section IV, real-world application of our proposed method is presented in section V, and to end this paper, a detailed conclusion is provided in section VI.

II. THEORETICAL BACKGROUND OF THE STUDY

In this section, we will briefly describe EMD and its different extensions such as ensemble empirical mode decomposition (EEMD), statistical empirical mode decomposition (SEMD), complementary ensemble empirical mode decomposition with adaptive noise (CEEMDAN), and one of the latest signal decomposition technique similar to EMD, which is known as variational mode decomposition (VMD).

A. EMPIRICAL MODE DECOMPOSITION

Using the EMD approach, a signal is broken down into a typically small number of IMFs. The well-known HHT algorithm is used in this method to decompose the complex signal into distinct oscillatory components ranging in frequency from low to high and a single monotone residue. Two requirements must be met for a signal to be an IMF: (i) the number of zero-crossings and extrema (maxima and minima) must be equal or deviate by no more than one; and (ii) The upper and lower envelopes' means, known as the local mean, must equal zero. The EMD technique can successfully split signal $y(t)$ into numerous components. This approach is reliable, simple, and efficient, requires no major model assumptions, and is broadly applied for prediction problems in a wide range of areas [18], [19], [20].

The detailed decomposition is presented in the following flowchart.

B. ENSEMBLE EMPIRICAL MODE DECOMPOSITION (EEMD)

The fundamental drawback of the traditional EMD when breaking down complex signals into different subgroups is mode mixing. Wu and Huang [21] proposed the EEMD technique in 2009 to address this problem. The algorithm of this technique is described below.

C. STATISTICAL EMPIRICAL MODE DECOMPOSITION (SEMD)

To extract the first mode, the SEMD approach proposed by [22] substitutes smoothing for cubic spline interpolation. The smoothing method has several advantages over cubic

Algorithm 1 The Standard EMD Algorithm

1. Recognize the entire local extrema (local maxima and minima) in the signal $\{y_i(t)\}$.
2. Find out the upper $\{U(t)\}$, and lower envelope $\{L(t)\}$ in the signal $y_i(t)$.
3. To obtain the mean of both the upper and lower envelope, join all the minima and maxima using the cubic spline interpolation approach, i.e., $M(t)$:

$$Mean(t) = \frac{U(t) + L(t)}{2} \tag{1}$$

4. To produce the first component, subtract the mean envelope computed in step 3 from the original signal, i.e.

$$k_1(t) = y(t) - Mean(t) \tag{2}$$

If $k_1(t)$ meets the two conditions for the IMF as indicated above, it should be considered as the initial IMF; otherwise, steps 1 through 4 will be repeated with $k_1(t)$ treated as a new signal.

5. The first IMF determined in step 4 will be subtracted from the signal $y(t)$ to generate $r_1(t)$, i.e.

$$r_1(t) = y(t) - k_1(t) \tag{3}$$

6. The filtering procedure from step 1 is once again applied in this stage, where $r_1(t)$ is treated as a new signal. Following the last step of EMD, the overall signal trend will be a smooth monotonic residue, and the actual signal, $y(t)$, will be decomposed as follows:

$$y(t) = \sum_{i=1}^{n} k_i(t) + r_n \tag{4}$$

It is important to note that r_n is the residue and $k_1(t)$, $k_2(t) \ldots k_n(t)$ are all different IMFs with varying frequencies ranging from high to low. Where r_n is the residue and $k_1(t)$, $k_2(t) \ldots k_n(t)$ are different IMFs with different frequencies that vary from high to low.

Algorithm 2 The standard EEMD algorithm of EEMD

1. Choose the ensemble number, m, and the amplitude of white noise, n.
2. Add a white noise series $w_i(t)$ to the actual signal $y(t)$ to obtain a new series $\acute{y}_i(t)$, i.e.

$$\acute{y}_i(t) = y(t) + w_i(t) \tag{5}$$

3. Breakdown the signal $\acute{y}_i(t)$ into different IMFs and a monotone residue using EMD.
4. Repeat the above steps 2, and 3 by adding diverse white noise series respectively; and
5. Attain the (ensemble) means of the appropriate IMFs of the decompositions as the concluding result.

Algorithm 3 Standard SEMD Algorithm

1. The signal $x(t)$ can be split into K test datasets $T_1, \ldots, T_k, \ldots, T_K$ where $x(t)$ is the test signal.
2. Calculate the local average of two neighboring points for the k^{th} test dataset, and obtain $\tilde{T}_k$.
3. The SEMD algorithm with a given smoothing parameter λ is applied to the composite signal $T_1, \ldots, T_{k-1}, \tilde{T}_k, T_{k+1}, \ldots, T_k$ to decompose it into $h_{1,\lambda}$ and r_λ.
4. With a given smoothing parameter λ, apply the SEMD algorithm to decompose the composite signal $T_1, \ldots, T_{k-1}, \tilde{T}_k, T_{k+1}, \ldots, T_k$ into an $h_{1,\lambda}$ and the remaining signal r_λ.
5. To obtain the predicted values for the remaining signal, it would be necessary to evaluate it at the k^{th} step which is denoted by $r_\lambda^k(t)$.
6. Steps *(ii)* to *(iv)* should be repeated for $k = 1, \ldots, K$, and the prediction error should be defined as follows:

$$PE(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \left\{ x(t) - r_\lambda^k(t) \right\}^2 \tag{6}$$

spline interpolation, particularly when the signal has been contaminated by noise. The step-by-step algorithm of this method is outlined in the following lines.

D. COMPLEMENTARY ENSEMBLE EMPIRICAL MODE DECOMPOSITION WITH ADAPTIVE NOISE (CEEMDAN)

The mode-mixing problem in the EMD algorithm can be removed by incorporating Gaussian white noise into the signal, and this type of modification in EMD is known as EEMD. However, the EEMD technique may not remove Gaussian white noise after signal reconstruction and causes reconstruction errors. To overcome this problem, the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) was proposed by [23].

After its first development, it has been widely used in many fields such as the exploitation of marine resources to extract ship-radiated noise by integrating CEEMDAN with adaptive noise [24]. The advantage of CEEMDAN is that it eliminates the problem of mode mixing efficiently, reconstruction errors become negligible as well as the cost of calculation is significantly minimized. Let's describe the function $EMD_j(\cdot)$ that makes possible the j^{th} mode attained by EMD, and let $w_j(\cdot)$ be white noise from the standard normal distribution, the algorithm of the CEEMDAN is summarized in the following steps:

E. VARIATIONAL MODE DECOMPOSITION (VMD)

The variational mode decomposition was proposed by Dragomiretskiy and Zosso in 2014 [25]. It can decompose a signal into multiple components, and its essence and core idea is the construction and solution of variational problems.

Algorithm 4 The Standard CEEMDAN Algorithm

1. The first IMF can be obtained by decomposing the Gaussian white noise added signal $y_i(t) = y(t) + \gamma_0 w_i(t)$ (where γ_0 is a noise coefficient, $i = 1, 2, \ldots, L$) by implementing the method of EMD. The first mode is then defined as:
$$\overline{IMF}_1 = \frac{1}{L}\sum_{i=1}^{L} IMF_{i1} \tag{7}$$
2. Calculate the first residue
$$r_1(t) = y(t) - \overline{IMF}_1 \tag{8}$$
3. Decompose residue $r_1(t) + \gamma_1 EMD(w_i(t))$ to obtain the 2nd mode as:
$$\overline{IMF}_2 = \frac{1}{L}\sum_{i=1}^{L} EMD_1[r_1(t) + \gamma_1 EMD_1(w_i(t))] \tag{9}$$
4. The resulting residue can be obtained by repeating the same procedure for each IMF, mathematically;
$$R_m(t) = y(t) - \sum_{j=1}^{m} \overline{IMF}_j \tag{10}$$

The total number of IMFs is denoted with m. The IMFs collectively break down the features of the original signal at diverse timescales. The residue describes the trend of the actual series, which is flatter and diminishes the forecast error excellently.

This new variant of EMD is used in many fields such as in marine science to analyze the underwater acoustic signal and to extract more distinctive features of a ship [26]. Similarly, the method of genetic algorithm (GA) is combined with VMD to enhance the recognition performance of bearing fault signals [27], and financial time series such as predicting the direction movement of the stock prices [28], [29], carbon price prediction [30], and underwater acoustic signal denoising [30], [31]. Inspired by the EMD, the method assumes that the original signal f is composed of a quantity of so-called Intrinsic Mode Functions (IMFs) μ_k which is defined as amplitude modulation and frequency modulation (AM-FM) components.

$$f(t) = \sum_k \mu_k(t) = \sum_k A_k(t)\cos[\varphi_k(t)] \tag{11}$$

The envelope, phase, and the IMFs in (11) are represented by $A_k(t)$, $\mu_k(t)$ and $\varphi_k(t)$. It is important to mention here that every individual IMF has a center frequency and limited bandwidth. In the VMD algorithm, the key decomposition process is the constrained variational problem. The mathematical structure of this constrained variational problem can be seen in the following equation (12).

$$\min_{\{\mu_k\},\{\omega_k\}} \left\{ \sum_k \left\| \xi \left[\left(\delta(t) + \frac{j}{\pi t} \right) * \mu_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}$$
$$s.t. \sum_k \mu_k = f \tag{12}$$

where $\{\mu_k\} := \{\mu_1, \ldots, \mu_K\}$ *and* $\{\omega_k\} := \{\omega_1, \ldots, \omega_K\}$ are shorthand notations for the set of all modes and their center frequencies, respectively. Equally, $\Sigma_k := \Sigma_{k=1}^{K}$ is considered the summation of all the modes. Based on the penalty factor and the Lagrangian multiplier, we can solve the above-constrained variational problem in (12). As a result, the augmented Lagrangian can be expressed as follows:

$$\begin{aligned} L(\{\mu_k\}, \{\omega_k\}, \lambda) = \alpha \sum_{k=1}^{K} \Bigg\| \partial t \Bigg[\left(\sigma(t) + \frac{j}{\pi t} \right) \\ * \mu_t(t) \Bigg] e^{-j\omega k t} \Bigg\|_2^2 \\ + \left\| f(t) - \sum_{k=1}^{K} \mu_k(t) \right\|_2^2 \\ + \langle \lambda(t), f(t) - \sum_{k=1}^{K} \mu_k \rangle \end{aligned} \tag{13}$$

λ and α are the Lagrangian multipliers and balancing parameters respectively in (13). To obtain the saddle point, the alternating direction multiplier method (ADMM) is applied, and after that μ_k, ω_k and λ are updated periodically, in the following manner:

$$\hat{\mu}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i>k} \hat{\mu}_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \tag{14}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{\mu}_k^{n+1} \right|^2 d\omega}{\int_0^\infty \left| \hat{\mu}_k^{n+1} \right|^2 d\omega} \tag{15}$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_k \hat{\mu}_n^{n+1}(\omega) \right) \tag{16}$$

The Fourier Transform and time step in (16) are denoted by λ and τ. As long as the convergence stop condition is satisfied, the algorithm continues to run. A stop condition is defined as follows:

$$\sum_k \frac{\left\| \hat{\mu}_k^{n+1} - \hat{\mu}_k^n \right\|_2^2}{\left\| \hat{\mu}_k^n \right\|_2^2} < e \tag{17}$$

The accuracy of convergence is given by e in (17). The complete algorithm of VMD is as follows:

III. PROPOSED METHOD

Before going into details about our suggested approach, which is based on the Akima spline interpolation technique, we will demonstrate how the conventional EMD and its several types provide subpar decomposition results for a

Algorithm 5 The Standard VMD Algorithm

1. Initialize $\left\{\hat{\mu}_k^1\right\}, \left\{\omega_k^1\right\}, \left\{\hat{\lambda}_1\right\}, n = 0$
2. Repeat $n \leftarrow n + 1$ **for** $k = 1 : K$ **do**
3. Update $\hat{\mu}_k$ for all $\omega \geq 0$:

$$\hat{\mu}_k^{n+1}(\omega) \leftarrow \frac{\hat{f}(\omega) - \sum_{i<k} \hat{\mu}_i^{n+1}(\omega) - \sum_{i>k} \hat{\mu}_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha\left(\omega - \omega_k^n\right)^2} \tag{18}$$

4. Update ω_k :

$$\omega_k^{n+1} \leftarrow \frac{\int_0^\infty \omega \left|\hat{\mu}_k^{n+1}(\omega)\right|^2 d\omega}{\int_0^\infty \left|\hat{\mu}_k^{n+1}(\omega)\right|^2 d\omega} \tag{19}$$

end for

5. Dual ascent for all $\omega \geq 0$:

$$\hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}^n(\omega) + \tau\left(\hat{f}(\omega) - \sum\nolimits_k \hat{\mu}_n^{n+1}(\omega)\right) \tag{20}$$

6. Repeat steps (2) to (5) until the iteration stop condition is satisfied:

$$\sum\nolimits_k \frac{\left\|\hat{\mu}_k^{n+1} - \hat{\mu}_k^n\right\|_2^2}{\left\|\hat{\mu}_k^n\right\|_2^2} < e$$

synthetic signal. Suppose that a synthetic signal is denoted by $x(t)$, where $0 < t < 9$ with frequencies $f_1 = 6$, $f_2 = 2$, and $f_3 = 1$ then the artificial signal's mathematical structure is provided as follows:

$$x(t) = 0.5t + \sin(f_1\pi t) + \sin(f_2\pi t) + \sin(f_3\pi t) + \epsilon(t) \tag{21}$$

where $\epsilon(t)$ is Gaussian noise with a signal-to-noise ratio (SNR) of 5. Fig. 1 shows the underlying signal of (1) with Gaussian noise with an SNR level of 5.

This noisy signal is decomposed into various components known as IMFs using the methods of EMD, EEMD, SEMD, VMD, CEEMDAN, and the newly proposed approach; the decomposition outcomes are displayed in Fig. 3.

Fig. 3 (a) demonstrates that the first IMF component, which represents the noise in the signal, is not captured by the usual method of EMD. In a similar vein, the altered versions of EMD known as EEMD, SEMD, VMD, and CEEMDAN also fail to separate the IMF1-based noise from the signal. While the number of IMFs acquired by SEMD is smaller than that of EMD, EEMD, and the proposed approach, the number of IMFs obtained by EMD, EEMD, and CEEMDAN are almost equal, showing that there is only a slight difference between these decomposition methods. Furthermore, the new method of VMD that is based on the idea of EMD

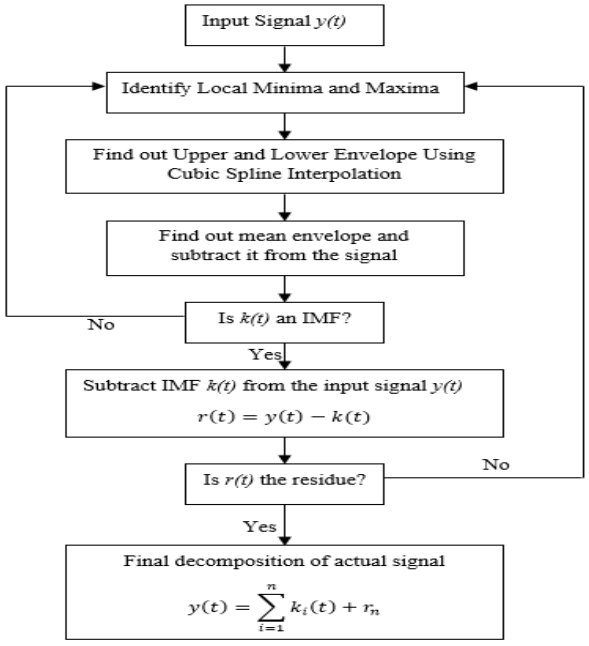

FIGURE 1. Schematic view of the EMD algorithm.

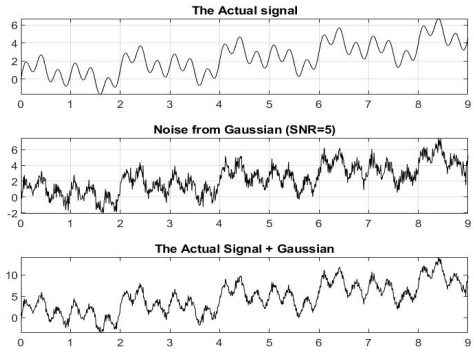

FIGURE 2. Actual signal, noise from Gaussian with SNR level of 5, and contaminated signal.

extracts the IMFs from the actual signal without producing a single monotone residue and the last IMF still consists of variations because of this structure, and algorithm design. Fig. 3(f) shows that the noise in the form of the first IMF is successfully retrieved when the signal is decomposed using the new technique. For the implementation of EMD, EEMD, SEMD, VMD, and CEEMDAN we have used three distinct libraries in RStudio, namely "Rlibeemd" [33], "EMD" [34], and "VMDecomp" [35] whereas the proposed method is implemented in MATLAB (R2021b) by switching the cubic spline interpolation method with Akima spline interpolation. Fig. 2 (a-e) demonstrates how EMD, EEMD, SEMD, VMD, and CEEMDAN fail to yield stable decomposition results from such a chaotic synthetic signal. For the implementation of EEMD, the default setting provided in the package is used with the number of siftings of 50, the ensemble size of 250, and the noise strength of 0.2, which is the standard deviation

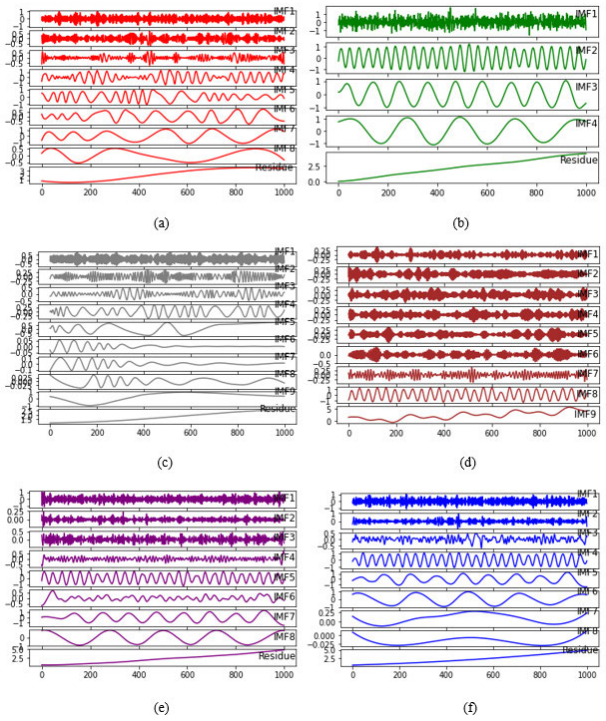

FIGURE 3. Decomposition of the noisy signal presented in Figure 2 by (a) EMD (b) SEMD (c) EEMD (d) VMD (e) CEEMDAN (f) Proposed Akima-EMD.

of the Gaussian random numbers used as additional noise. Similarly, the method of VMD is applied with the default settings. The cubic spline interpolation technique used in the building of the upper and lower envelopes through the local extrema is the primary reason for the failure of both EMD, EEMD, and CEEMDAN. The algorithm of VMD is different from EMD, EEMD, SEMD, and CEEMDAN, and therefore for any synthetic or real-world data it extracts nine IMFs with default settings. The last component known as IMF9 still has some oscillations and unlike the other variants, we are not able to obtain the monotone single residue. The lower frequency component of the signal is not effectively reflected by the mean envelope because it is vulnerable to noise. When there are heavy-tailed noises present, such as outliers, the upper or lower envelope tends toward extreme values, making it impossible for the mean envelope to accurately represent the lower frequency pattern of the signal. Additionally, SEMD, another modified version of EMD that uses smoothing rather than cubic spline interpolation, is unable to remove noise from the noisy signal. Furthermore, the first IMF is subtracted from the signal for the purpose to find out the deviations between the actual signal and the remaining one after subtracting the first IMF component from it. We used three different statistical metrics such as RMSE, MAE, and MAPE to know which method successfully extracted noise from the synthetic signal. The investigational results of this analysis are presented in the following Table 1.

TABLE 1. Statistical metrics of actual and noise-free signal.

Method	RMSE	MAE	MAPE
EMD	0.3887	0.3480	0.4717
EEMD	0.4101	0.3468	0.4527
SEMD	0.5225	0.4236	0.4976
CEEMDAN	0.3599	0.2963	2.1733
VMD	2.7399	2.3376	1.6442
Proposed Akima-EMD	0.2572	0.2204	0.4024

According to Table 1, the proposed method extracts noise from a contaminated signal more effectively because RMSE, MAE, and MAPE are the minimums among other methods, such as EMD, EEMD, SEMD, CEEMDAN, and VMD.

To solve this issue, we suggest a novel method that does not change the interpolation with smoothing as in the case of SEMD, which yields the worst decomposition results when compared to EMD, EEMD, CEEMDAN, and VMD but rather uses a different kind of interpolation approach. This interpolation technique is known as Akima spline interpolation, and therefore, we call our proposed new method Akima-EMD. Before going into the details of our proposed method we are first introducing the Akima interpolation technique in the following subsection.

A. AKIMA SPLINE INTERPOLATION TECHNIQUE

Hiroshi Akima first presented the Akima spline interpolation method in 1970 [36]. The Akima interpolation is a sub-spline interpolation that is continuously differentiable. It is constructed using fragmented third-order polynomials. Piecewise cubic Hermite interpolation finds a cubic polynomial for each interval $[x_i, x_{i+1}]$ of an input data set of nodes x and values ϑ, that not only interpolates the provided data values v_i and v_{i+1} at the interval's nodes x_i and x_{i+1}, but also has particular derivatives d_i and d_{i+1} at x_i and x_{i+1}. A key factor in the Hermite interpolation method is the selection of derivatives di. A derivative formula was suggested by the author to reduce the number of local undulations: Let $\delta_i = \frac{(v_{i+1}-v_i)}{(x_{i+1}-x_i)}$ to be the slope of the interval $[x_i, x_{i+1}]$. Akima's derivative at x_i is defined as:

$$d_i = \frac{|\delta_{i+1} - \delta_i|\,\delta_{i-1} + |\delta_{i-1} - \delta_{i-2}|\,\delta_i}{|\delta_{i+1} - \delta_i| + |\delta_{i-1} - \delta_{i-2}|} \tag{22}$$

Akima's derivative formula is altered by adjusting the weights w_1 and w_2 of the slopes δ_{i-1} and δ_i to remove overshoot and prevent edge cases where both the numerator and denominator are equal to zero.

$$d_i = \frac{w_1}{w_1 + w_2}\delta_{i-1} + \frac{w_2}{w_1 + w_2}\delta_i \tag{23}$$

where $w_1 = |\delta_{i+1} - \delta_i| + |\delta_{i+1} + \delta_i|/2$, *and* $w_2 = |\delta_{i-1} - \delta_{i-2}| + |\delta_{i-1} + \delta_{i-2}|/2$.

Observe that the five points $x_{i-2}, x_{i-1}, x_i, x_{i+1} and x_{i+2}$ are used to locally calculate Akima's derivative at x_i. It requires the slopes δ_{-1}, δ_0, and δ_n, δ_{n+1} for the endpoints x_1 and x_n. Akima recommended using quadratic extrapolation to compute these slopes as $\delta_0 = 2\delta_1 - \delta_2, \delta_{-1} = 2\delta_0 -$

δ_1 *and* $\delta_n = 2\delta_{n-1} - \delta_{n-2}$ as they are not included in the input data. Because it only employs values from nearby knot points in the formation of the upper and lower envelopes and the coefficients of the interpolation polynomial between any two-knot points, the Akima spline interpolation has an advantage over the cubic spline in the classic EMD. As a result, there are not many complicated equations to solve, and the Akima spline prevents unnatural wiggles in areas where the second derivative of the underlying curve is changing quickly. The fact that the second derivative of the Akima spline is discontinuous could be a drawback.

B. ALGORITHM OF THE PROPOSED AKIMA-EMD

The suggested novel Akima-EMD technique, which uses the Akima spline interpolation technique for both clean and noisy signals, is summarized as follows:

Algorithm 6 Algorithm of the Proposed Akima-EMD

1. By using the Akima spline interpolation approach, isolate the initial oscillatory component $k^*(t)$ from the noisy signal $y(t)$.
2. Determine the upper $\{U\ (t)\}$, and lower $\{L\ (t)\}$ envelopes of the signal $y_i(t)$ respectively.
3. To obtain the mean of the upper and lower envelopes, or $M\ (t)$, join all the minima and maxima using the Akima spline interpolation approach, i.e.
$$M\ (t) = \frac{U\ (t) + L(t)}{2} \tag{24}$$
4. The mean envelope calculated in step 3 will be subtracted from the actual signal to obtain the first component, i.e.
$$k\ (t) = y\ (t) - M\ (t) \tag{25}$$
5. Repeat steps 1-4 for the component $k\ (t)$ until a stopping criterion is satisfied, and take the resulting $k\ (t)$ as $k^*(t)$.

If the remaining signal i.e., $r\ (t) = y\ (t) - k^*\ (t)$ has still some oscillation components then it can be further decomposed with the help of a new version of EMD. constant for more than two consecutive nodes.

Any noisy signal can be decomposed with the help of the new proposed Akima EMD. The addition of $|\delta_{i+1} + \delta_i|\ /2$, *and* $|\delta_{i-1} + \delta_{i-2}|\ /2$ terms forces d_i=*0* when $\delta_i = \delta_{i+1} = 0$, *i.e.*, $d_i = 0 when v_i = v_{i+1} = v_{i+2}$, and hence it eliminates the overshoot problem when the data is

IV. SIMULATION STUDY

To assess the empirical efficiency of the proposed method, known as Akima-EMD, a simulation study is carried out to measure how effectively the first component of the IMF can extract the noise from the data. The purpose is to compare the proposed method with the benchmark EMD, as well as its different variants, such as EEMD, SEMD, CEEMDAN, and VMD. We created a noisy signal from the following model to conduct simulation studies and determine the effectiveness of our novel method:

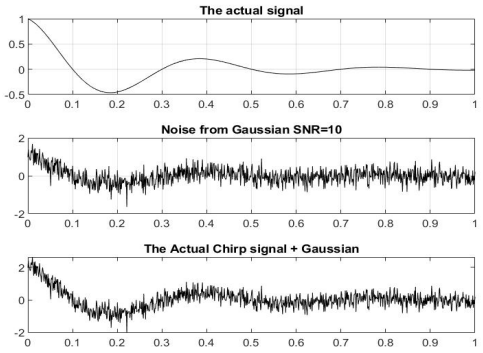

FIGURE 4. Actual chirp signal, noise from Gaussian with SNR level of 10, and contaminated chirp signal.

$$K\ (t_i) = F\ (t_i) + \mu\ (t_i) \tag{26}$$

In the above (26) $F\ (t_i)$ are a test function and $\mu\ (t_i)$ is the noise that will be generated from 1) Gaussian with different SNR levels, 2) heavy-tailed distribution such as t-distribution with varying degrees of freedom. It was observed that the final scenario depicts a situation in which noise includes outliers or is generated from a heavy-tailed distribution. The chirp signal (chirp) is the test function or signal that was used in the simulation study. Although this signal has a variety of mathematical structures, in this investigation we employed the chirp signal defined as follows:

$$F\ (t_1) = e^{(-4t)\cos(5\pi t)}\ (t \in [0, 9]) \tag{27}$$

The noise produced by various distributions contaminates the chirp signal described above; a detailed description of how to add different noises to the above chirp signal is as follows.

A. GAUSSIAN NOISE ADDED INTO CHIRP SIGNAL WITH SNR LEVEL OF TEN

In the first scenario of the simulation study, the chirp signal described in (27) is initially contaminated with Gaussian noise at an SNR level of 10. Fig. 4 displays the real signal, the noise produced by a Gaussian distribution at an SNR level of 10, and the noise-added signal.

In this scenario of simulation experiments, we replicated 500 simulated datasets from the chirp signal with various data lengths, such as 500, 1000, 5000, and 10000. To decompose the chirp test function into different IMFs, the six methods—EMD, EEMD, SEMD, CEEMDAN, VMD, and the newly proposed Akima-EMD were applied to each replicated simulated data set with a different number of data points. It is important to acquire the estimate $\hat{p}$ for the test function after decomposing the signal; $\hat{p}$ is determined by removing the first component (IMF1) of the noisy signal $K\ (t_i)$, i.e.

$$\hat{p} = K\ (t_i) - IMF1 \tag{28}$$

TABLE 2. Average accuracy metrics for 500 replicated simulated data sets with 500, 1000, 5000, and 10000 number of observations of chirp signal having Gaussian added noise with SNR level of 10.

N	Accuracy Metrics	EMD	EEMD	SEMD	Akima-EMD	CEEMDAN	VMD
500	RMSE	0.233	0.226	0.329	0.216	0.257	0.345
	MAE	0.205	0.206	0.260	0.195	0.287	0.228
	MAPE	7.816	6.116	8.826	5.170	8.391	9.454
1000	RMSE	0.234	0.226	0.329	0.209	0.265	0.354
	MAE	0.206	0.204	0.260	0.191	0.298	0.348
	MAPE	6.687	6.077	9.987	5.620	6.256	7.981
5000	RMSE	0.234	0.231	0.327	0.226	0.291	0.339
	MAE	0.206	0.209	0.259	0.201	0.219	0.231
	MAPE	6.103	6.700	9.987	5.987	6.987	6.276
10000	RMSE	0.234	0.239	0.327	0.226	0.241	0.267
	MAE	0.206	0.218	0.259	0.205	0.278	0.228
	MAPE	5.835	6.876	11.127	5.079	6.812	7.143

The proposed Akima-EMD approach is compared with other methods based on three different statistical performance metrics such as average RMSE, MAE, and MAPE. The following equations show how these various statistical metrics are mathematically structured.

$$Average\ RMSE = \frac{1}{G}\sqrt{\frac{1}{n}\sum_{j=1}^{g}\sum_{t=1}^{n}\left(K_{tj}-\hat{p}_{tj}\right)^2} \quad (29)$$

$$Average\ MAE = \frac{1}{G}\left[\frac{1}{n}\sum_{j=1}^{g}\sum_{t=1}^{n}\left|K_{tj}-\hat{p}_{tj}\right|\right] \quad (30)$$

$$Average\ MAPE = \frac{1}{G}\left[\frac{1}{n}\sum_{j=1}^{g}\sum_{t=1}^{n}\left|\frac{K_{tj}-\hat{p}_{tj}}{K_{tj}}\right|\right] \quad (31)$$

Each of these performance indicators has 500 values because there are 500 replicated simulated datasets, and hence, the average of RMSE, MAE, and MAPE is calculated. The empirical results of these performance metrics are presented in Table 2.

It is evident from the empirical results presented in Table 2 that the proposed method efficiently extracts noise from the signal. Since in the first case there are replications of 500 simulated datasets from chirp signals of different lengths such as 500, 1000, 5000, and 10000, therefore, we have 500 values of RMSE, MAE, and MAPE. Table 2 shows that for the chirp signal after adding noise from Gaussian distribution with an SNR level of 10 average of three different statistical metrics such as RMSE, MAE, and MAPE of the proposed method with different lengths of data is minimum. Hence, it can be concluded that the proposed Akima-EMD method successfully extracts noise in the form of IMF1 from the chirp signal, which is contaminated, with Gaussian noise with an SNR level of 10.

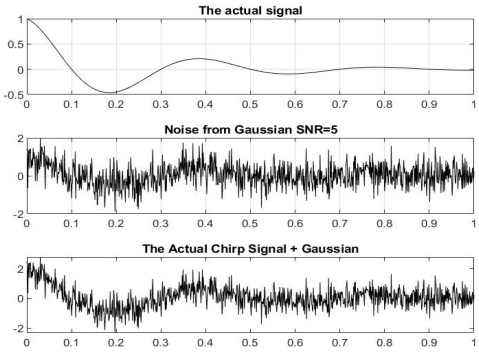

FIGURE 5. Actual chirp signal, Gaussian noise with SNR level 5, and contaminated chirp signal.

B. GAUSSIAN NOISE ADDED INTO CHIRP SIGNAL WITH SNR LEVEL OF FIVE

In the first case of the simulation, we used a chirp signal defined in (27), and Gaussian noise with an SNR level of 10 is added to make it noisy. We simulated 500 replicated datasets with a different number of observations such as 500, 1000, 5000, and 10000. In the second case, the same procedure is repeated i.e., the Gaussian noise is again added but the SNR level is changed from 10 to 5 for the purpose to contaminate the chirp signal with more noise. In Fig. 5 the actual signal, noise from Gaussian with SNR level of 5, and the noise added chirp signal are plotted. It can be seen that increasing the SNR level will make the chirp signal noisier.

The purpose of changing the SNR level is to know the efficiency of the proposed Akima-EMD for such a heavy-noisy signal. It is worth mentioning here that replacing the SNR level; from ten to five generates more noise in the signal and in such a case, the performance of the conventional EMD is affected because of the cubic spline interpolation technique that is used inside the EMD algorithm to find out the upper and lower envelope. The proposed Akima-EMD, which uses the Akima spline interpolation technique for the purpose to construct lower and upper envelopes, performs more efficiently in such a scenario and extracts the Gaussian noise in the form of the first IMF component. The recent extension of EMD known as SEMD uses smoothing rather than interpolation to find out the mean envelope is showing inefficient performance for this noisy chirp signal. Unlike the SEMD, the EEMD, and CEEMDAN techniques perform better than the classical benchmark EMD. It can be seen from Table 3 that the statistical performance measures such as the average RMSE, MAE, and MAPE are minimum for our proposed new method. These statistical metrics were obtained after subtracting the first IMF component from the signal and comparing the remaining signal with the actual noisy chirp signal.

TABLE 3. Average accuracy metrics for 500 replicated simulated data of chirp signal having Gaussian added noise with SNR level of 5.

N	Accuracy Metrics	EMD	EEMD	SEMD	Proposed Akima-EMD	CEEMDAN	VMD
500	RMSE	0.414	0.402	0.570	0.402	0.431	**0.40**
	MAE	0.364	0.367	0.454	0.364	0.391	**0.370**
	MAPE	5.848	7.656	10.098	4.730	**6.031**	**6.891**
1000	RMSE	0.417	0.426	0.564	0.406	**0.432**	**0.419**
	MAE	0.366	0.367	0.448	0.358	**0.379**	**0.407**
	MAPE	8.827	9.041	12.0987	6.076	**9.982**	**7.103**
5000	RMSE	0.417	0.436	0.568	0.403	**0.443**	**0.429**
	MAE	0.366	0.389	0.452	0.365	0.371	**0.387**
	MAPE	6.704	7.129	15.765	7.186	8.120	9.591
10000	RMSE	0.417	0.423	0.568	0.402	0.435	**0.487**
	MAE	0.366	0.370	0.453	0.365	**0.390**	**0.432**
	MAPE	7.063	8.056	20.965	5.944	**9.971**	8.032

The empirical results presented in Table 3 indicate that the average values for RMSE, MAE, and MAPE of our proposed Akima-EMD are the minimum for this chirp signal, which is again contaminated with heavy noise since we changed the SNR level from 10 to 5. By using EMD, SEMD, EEMD, CEEMDAN, VMD, and the newly proposed Akima-EMD method, we were able to decompose the noisy signal into different IMFs, and a single monotone residue, and then subtract the first IMF component from the noisy chirp signal to determine the average values of RMSE, MAE, and MAPE, as indicated in Table 3.

C. NOISE FROM HEAVY TAIL T-DISTRIBUTION WITH DIFFERENT DEGREES OF FREEDOM

In the first two cases, the Gaussian noise is added to the chirp signal with two different SNR levels i.e., ten, and five. It is evident from the empirical results presented in Tables 2, and 3 that the proposed new Akima-EMD method outperformed the other methods such as EMD, EEMD, SEMD, CEEMDAN, and VMD in terms of removing noise from the signal as the average of three statistical metrics such as RMSE, MAE, and MAPE of is minimum. However, a query has to be answered 'is it efficient in a case when noise is generated from a heavy-tailed distribution such as t-distribution'? To answer this question, noise in the chirp signal is added from a heavy-tailed t-distribution with different degrees of freedom. The shape of the t-distribution is depended upon its single degree of freedom; therefore, noise is generated from this heavy tail t-distribution with a different number of degrees of freedom. The performance of the proposed method when noise is added from heavy-tailed t-distribution can be seen in the following subsections.

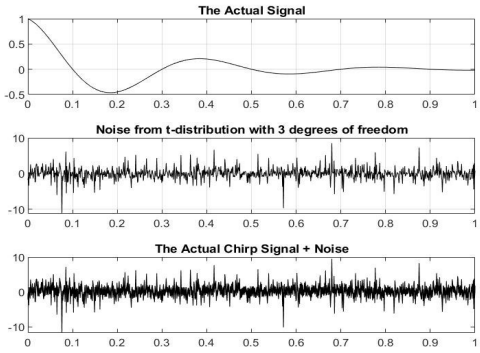

FIGURE 6. Actual, Noise, and the noisy signal after adding noise from t-distribution with three degrees of freedom.

1) NOISE FROM T-DISTRIBUTION WITH $\vartheta = 3$ (DEGREES OF FREEDOM)

As a first case, noise is added in the chirp signal from the heavy-tailed t-distribution with three degrees of freedom and decomposed the noise-added chirp signal with the help of EMD, EEMD, SEMD, CEEMDAN, VMD, and the proposed Akima-EMD. In this scenario, the proposed method performed well for each length of the data such as 500, 1000, 5000, and 10000 for 500 different replicated datasets in terms of the smallest values of average RMSE, MAE, and MAPE. The chirp signal, noise, and noise-augmented chirp signal are shown in Fig. 6.

The average statistical metrics such as RMSE, MAE, and MAPE are calculated after subtracting the first IMF component from the actual signal and comparing the noise-free remaining signal with the actual. Investigation results are presented in Table 4 which clearly shows that the proposed Akima-EMD method outperforms the other methods in terms of accuracy metrics such as average RMSE, MAE, and MAPE, in all scenarios i.e. for 500 replicated simulated datasets, having different lengths of the data such as 500, 1000, 5000, and 10000.

2) CASE-II NOISE FROM T-DISTRIBUTION WITH $\vartheta = 5$ (DEGREES OF FREEDOM)

In the preceding example, noise from the heavy-tailed t-distribution with three degrees of freedom was created and then extensively added to the chirp signal. The noise-added

TABLE 4. Average accuracy metrics for 500 replicated simulated data sets of chirp signal with noise added from heavy-tailed t-distribution with $\vartheta = 3$ (degrees of freedom).

N	*Accuracy Metrics*	*EMD*	*EEMD*	*SEMD*	*Proposed Akima-EMD*	CEEMDAN	VMD
500	RMSE	8.568	8.312	11.803	7.814	9.120	10.920
	MAE	7.394	7.600	9.119	6.997	8.413	8.001
	MAPE	5.535	8.056	11.007	4.972	5.001	7.881
1000	RMSE	8.504	7.888	11.557	6.583	7.101	8.192
	MAE	7.427	7.513	8.971	7.261	7.879	9.092
	MAPE	6.521	9.840	18.192	6.409	7.112	7.740
5000	RMSE	8.595	8.986	11.795	8.068	8.429	9.182
	MAE	7.436	7.671	9.036	7.308	7.672	8.217
	MAPE	6.163	7.091	13.030	5.460	6.429	10.340
10000	RMSE	8.622	9.071	11.770	8.133	8.562	10.281
	MAE	7.475	7.900	8.953	7.333	7.729	9.028
	MAPE	7.540	6.192	10.109	5.754	6.921	7.104

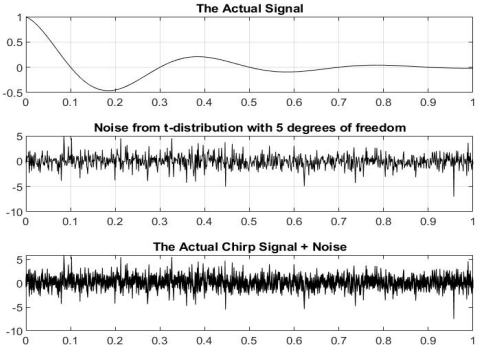

FIGURE 7. Actual, Noise, and the noisy signal after adding noise from t-distribution with five degrees of freedom.

chirp function is then decomposed with the help of EMD, EEMD, SEMD, CEEMDAN, VMD, and the new Akima-EMD method. Now, as a second case, we generated noise again from the heavy-tailed t-distribution with five degrees of freedom and added this heavy noise to the chirp signal. The actual noise-free signal, the noise, and the noisy added signal are presented in the following Fig. 7 below.

According to Fig. 7, after adding noise from a t-distribution with five degrees of freedom, the chirp signal becomes contaminated with high-amplitude noise, which is statistically known as an outlier. Now the same procedure is again repeated for the purpose to extract the noise from the chirp signal. Since the experiment is performed by simulating 500 datasets from a chirp signal with different lengths of data such as 500, 1000, 5000, and 10000. Hence, we have four different simulated cases, i.e. 500 different datasets generated from the chirp signal each of length 500 as a first case, and subsequently 1000, 5000, and 10000 lengths. Therefore, on each iteration, the chirp function is decomposed with the help of EMD, EEMD, SEMD, CEEMDAN, VMD, and the new method. The IMF1 is subtracted from the actual chirp signal, and average statistical metrics such as RMSE, MAE, and MAPE are calculated and presented in Table 5. It can be seen from the empirical results presented in Table 5 that the average values of RMSE, MAE, and MAPE of the proposed Akima-EMD method are minimum in all scenarios. In the case of such noisy chirp signals, the performance of the benchmark EMD is not satisfactory due to the use of cubic spline interpolation to determine the lower, upper, and mean envelopes. Furthermore, the performance of the newly proposed SEMD is not satisfactory for such a nonstationary and nonlinear simulated chirp signal, which seems very unusual. On the other hand, the method of EEMD, CEEMDAN, and the proposed new method produces very similar results for the average statistical measures such as RMSE, and MAE, whereas the baseline EMD performance is not very satisfactory to extract the noise from the noise-added chirp signal.

To end this case of simulation study where the noise is generated from t-distribution with five degrees of freedom and then added this noise with the chirp signal to contaminate

TABLE 5. Average accuracy metrics for 500 replicated simulated data sets of chirp signal with noise added from heavy-tailed t-distribution with $\vartheta = 5$(degrees of freedom).

N	Accuracy Metrics	EMD	EEMD	SEMD	Proposed Akima-EMD	CEEMDAN	VMD
500	RMSE	6.313	6.196	8.350	5.802	**6.102**	**7.998**
	MAE	5.515	5.630	7.583	5.269	**5.490**	**6.000**
	MAPE	10.691	6.100	25.764	4.032	**9.109**	**11.850**
1000	RMSE	6.523	6.246	10.683	6.014	**6.921**	**8.019**
	MAE	5.696	5.614	9.909	5.483	**5.810**	**7.491**
	MAPE	6.012	8.433	31.00	5.587	**6.060**	**7.102**
5000	RMSE	6.449	6.993	8.685	6.134	**6.572**	**7.199**
	MAE	5.660	5.929	9.881	5.552	**5.920**	**6.002**
	MAPE	20.153	18.001	21.086	8.345	**19.523**	**22.996**
10000	RMSE	6.417	6.351	14.673	6.145	**6.720**	**8.901**
	MAE	5.638	5.789	12.889	5.575	**5.701**	**7.000**
	MAPE	12.587	11.207	11.201	7.190	**12.031**	**13.380**

it with heavy noise. After decomposing the noise added chirp signal with EMD, EEMD, SEDM, CEEMDAN, VMD and the proposed Akima-EMD the first IMF component is subtracted from the actual chirp signal. A comparison is made between the actual chirp signal and the remaining signal (after removing the IMF1) to find out average RMSE, MAE, and MAPE. It is evident from the empirical results presented in table 5 that the proposed method outperforms the other data decomposition method in terms of minimum average values of RMSE, MAE, and MAPE, and hence to be considered as a new tool to decompose the nonlinear and nonstationary data specifically time series.

V. APPLICATION OF THE PROPOSED METHOD ON REAL DATA SETS

In section IV, the proposed Akima-EMD method was applied to the simulated datasets generated by using the chirp signal defined in (25). As a first case, noise from the Gaussian with different SNR levels i.e. five, and ten has been generated and added to the chirp signal to contaminate it with outliers. As a second case, noise from the heavy-tailed t-distribution with different degrees of freedom is also added to the chirp signal. In all of these scenarios, the suggested technique outperforms previous methods such as EMD, EEMD, SEMD, CEEMDAN, and VMD in terms of removing noise from the signal. However, in real-world scenarios, where the data is not created by any single function, nor is the noise generated by any certain distribution. Indeed, there may have a complex, nonstationary, and nonlinear dataset, or a dataset with some trend or seasonality. In general, the performance of any given approach must be consistent for both the simulated and actual data sets. The following two real-world time series datasets were used to check the performance of the proposed method.

A. CASE-I: NOISE-FREE WEATHER TIME SERIES

As a first real-world dataset, we analyzed weather time series data of Australia collected by the Australian bureau of meteorology department and available online (http://www.bom.gov.au/climate/dwo/). This daily time series dataset was gathered from several Australian weather stations and comprises 21 variables for 10 years. However, in the analysis, we only used univariate time series data, i.e., daily minimum temperature, for the Sydney weather station. It is important to mention that this dataset is not contaminated with any outlying observation, and has some seasonality. The following Fig. 8 illustrates that the Sydney daily minimum temperature weather time series data has a seasonal component and does not contain any unusual observations.

The daily minimum temperature data is decomposed into distinct IMFs and monotone residue using the EMD, EEMD, SEMD, CEEMDAN, VMD, and the newly developed Akima-EMD method.

Fig. 9 shows the results of this decomposition; it can be observed that for the noise-free real dataset, the decomposition with the proposed technique is fairly close to the EMD, EEMD, and CEEMDAN with just one more IMF, but the SEMD method provides poor decomposition results with only five IMFs. The same procedure as in the simulation studies section is repeated after decomposing the daily minimum temperature time series data into different IMFs and a monotone residue. The first IMF component is removed from the real daily minimum temperature data to generate a new series, which is then compared to the actual data to get the three important statistical measures known as RMSE, MAE, and MAPE presented in Table 6 below.

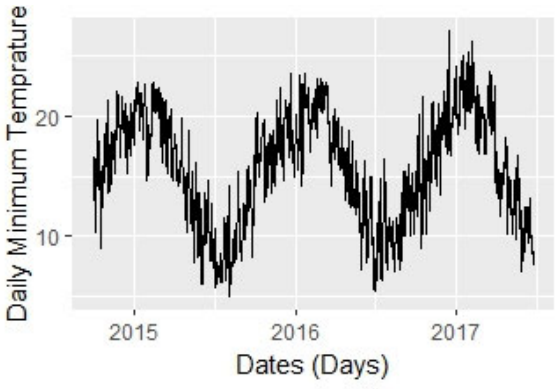

FIGURE 8. Daily minimum temperature of Australia (Sydney) for the period 9/30/2014 to 6/25/2017.

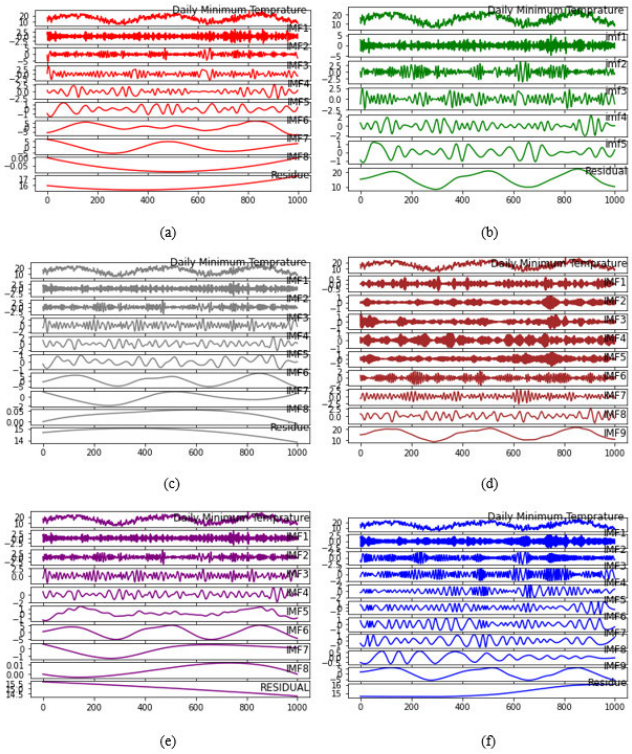

FIGURE 9. Decomposition of Australia (Sydney) daily minimum temperature time series data in the time window 9/30/2014 to 6/25/2017 by implementing a) EMD b) EEMD c) SEMD d) VMD e) CEEMDAN and f) Proposed Akima-EMD.

TABLE 6. Accuracy metrics for noise-free (not having a single outlier) temperature data.

Method	RMSE	MAE	MAPE
EMD	1.207	0.988	0.070
SEMD	1.300	1.052	0.075
EEMD	1.042	0.873	0.062
CEEMDAN	1.044	0.835	0.059
VMD	16.155	15.669	1.027
Akima-EMD	0.907	0.732	0.049

The empirical findings shown in Table 6 show that the values of RMSE, MAE, and MAPE of the suggested novel technique are minimal when compared to EMD, EEMD SEMD, CEEMDAN, and VMD. Therefore, the proposed Akima-EMD successfully removes noise in terms of IMF1 from the daily time series temperature data. Furthermore, the EEMD and CEEMDAN findings are remarkably comparable to the proposed new method, and it is considered a second choice in this comparison of how these data-driven algorithms eliminate noise from data. When the authors in [21] employed smoothing instead of interpolation to determine the lower and upper envelopes, the SEMD approach yields inferior decomposition results when compared to the baseline EMD.

B. CASE-II: WEST TEXAS INTERMEDIATE (WTI) CRUDE OIL TIME SERIES DATA

In the first real-world scenario, the daily weather time series data collected from several Australian weather stations in Australia is used. Because this time series dataset is clean of noise and contains no outliers, therefore, the decomposition results obtained by the standard EMD and the novel technique are quite comparable. In a second scenario, the performance of the proposed Akima-EMD is examined using daily crude oil price time series, i.e., from January 1, 2019, to December 31, 2022, collected from the website of yahoo finance (https://finance.yahoo.com/). This time series dataset is chosen for two reasons. To confirm that the proposed method is equally efficient on different types of datasets, the weather dataset is used in the first instance and the daily closing price of crude oil such as West Texas Intermediate (WTI) in the second case. Second, the crude oil daily closing price time series data contains an anomaly in which the daily closing price of a barrel of WTI, the US oil benchmark, dropped as low as negative $37.63 per barrel (recorded on April 19th, 2020). This means that oil producers are paying buyers to take the commodity off their hands over fears that storage capacity could run out. The reason behind this huge decrease in crude oil prices is the COVID-19 pandemic where oil demand has all but dried up as lockdowns across the world have kept people inside. Figure 10 depicts the daily closing price of WTI; it is noticeable that prices fell as low as negative $37.63.

The crude oil daily closing price time series dataset has 791 observations spanning from January 1, 2019, to December 31, 2022, including the pandemic period.

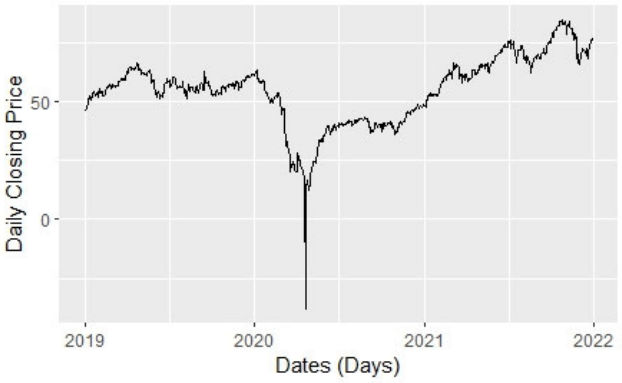

FIGURE 10. WTI daily closing prices in the time window January 01, 2019, to December 31, 2021.

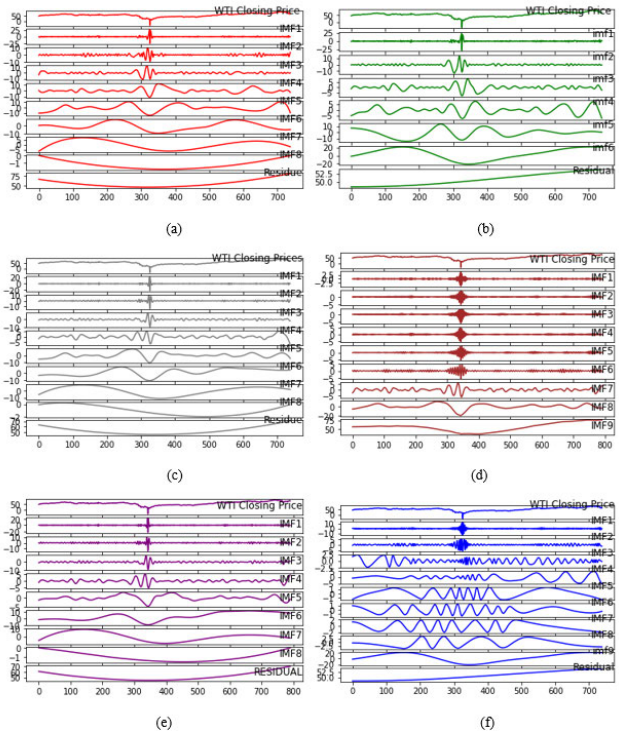

FIGURE 11. Decomposition of WTI crude oil data by implementing the method of a) EMD b) SEMD c) EEMD d) VMD e) CEEMDAN f) Proposed Akima-EMD.

This nonlinear and nonstationary daily time series data is decomposed into several components and a single monotone residue, implementing the methods of EMD, EEMD, SEMD, CEEMDAN, VMD, and the proposed Akima-EMD. Fig.11 depicts the decomposition results; it can be seen from this figure that the decomposition results of the novel approach are stable and extract noise in the form of the first IMF. Whereas the other methods such as EMD, EEMD, SEMD, CEEMDAN, and VMD fail to extract the first IMF as a noise present in the WTI crude oil data.

After decomposing the crude oil WTI daily closing price into different IMFs the next aim is to know which method

TABLE 7. Accuracy metrics for crude oil WTI daily closing price data having a single outlier.

Method	RMSE	MAE	MAPE
EMD	2.373	0.894	0.027
SEMD	3.081	1.527	0.043
EEMD	1.702	0.760	0.021
VMD	56.176	54.866	1.029
CEEMDAN	1.596	0.597	0.027
Akima-EMD	1.527	0.583	0.017

successfully extracted the noise. To verify the claim that the proposed Akima-EMD extracted the noise in the form of the first IMF component from the crude oil time series we subtracted the IMF1 from the actual time series and compared the leftover data with the actual. Three performance metrics such as RMSE, MAE, and MAPE are calculated after comparing the actual crude oil time series data with the leftover (after removing noise in terms of IMF1) and presented in Table 7.

The empirical findings shown in Table 7 show that the suggested technique successfully removes noise since the results of the three statistical metrics are the lowest when compared to the other three methods such as EMD, EEMD, SEMD, VMD, and CEEMDAN. Once again, the SEMD performs the poorest of the four approaches, demonstrating that utilizing smoothing rather than interpolation when generating the upper and lower envelopes neither improves the decomposition results nor does it address the boundary condition and wiggles at both ends of the data. In the classical EMD, the cubic spline interpolation technique is changed with the Akima spline interpolation. The advantage of this change is that it overcomes the limitation of the overshoot, boundary condition, and wiggles in the EMD. By solving the aforementioned limitations in the benchmark, EMD the proposed Akima-EMD decomposes the crude oil WTI time series data successfully and considers the first IMF component as noise.

Briefly, it is obvious that decomposing such nonstationary and nonlinear time series data with the help of the new method known as Akima-EMD effectively extracts the fluctuations, and it will become easier for financial time series experts to predict its future trajectory once they identify noise in the data.

VI. CONCLUSION

In this research article, we suggested a novel method that overcomes the limitations of EMD such as overshoot, undershoot, wide swings, and wiggles at both ends of the data due to the cubic spline interpolation technique. After developing the new method we used a synthetic signal to know its effectiveness and compare it with other decomposition methods such as EMD, EEMD, SEMD, CEEMDAN, and VMD. For the experimental purpose, a synthetic signal that is sinusoidal which is defined in (27) contaminated with noise generated from Gaussian distribution with an SNR level of 5 is used. After contaminating the signal with the noise this noisy signal is then decomposed with the newly proposed method in comparison with the other methods such as the classical EMD, SEMD, EEMD, CEEMDAN, and VMD. Three different statistical metrics such as RMSE, MAE, and MAPE are calculated after subtracting the first IMF component from the actual data and then comparing the remaining signal with the actual one. Investigation results of RMSE, MAE, and MAPE presented in Table 1 suggest that the proposed new method which is known as Akima-EMD extracts noise from the contaminated signal effectively as compared to the other methods used in the comparison. Furthermore, for the simulation study, we took the chirp signal defined in (27) and generated 500 different replicated datasets from it. In the first case, the number of observations in each of these simulated datasets was 500. This chirp signal is then contaminated by generating noise from Gaussian distribution using different SNR levels such as ten, and five. After contaminating the chirp signal with the noise then decompose this noisy signal with the newly proposed method in comparison with the other methods such as the classical EMD, SEMD, EEMD, CEEMDAN, and VMD. The same procedure was repeated by changing the number of observations from 500 to 1000, 5000, and 10000. After subtracting the first IMF component from the real data and comparing the remaining signal to the actual one, the average RMSE, MAE, and MAPE are computed. The empirical findings shown in Tables 2, and 3 suggests that the average values of the statistical metrics of the proposed novel technique are minimum in all scenarios, i.e. 500 replicated simulated datasets with 500, 1000, 5000, and 10000 lengths of data. The same chirp signal is used in the second phase of the simulation study, but the noise level is changed from Gaussian to heavy-tailed t-distribution with different degrees of freedom such as three, and five. We followed the same procedure as in the case of Gaussian to extract the heavy-tailed noise added to the chirp signal from t-distribution. The investigational findings of average RMSE, MAE, and MAPE provided in Tables 4, and 5 indicate that the suggested new approach outperforms previous methods such as EMD, EEMD, SEMD, CEEMDAN, and VMD in terms of removing noise in the form of first IMF from the noisy chirp signal. The performance of any new method involves not just simulation studies but also testing on real-world data. For this purpose, two different real-world time series datasets are used, i.e. the temperature data and the WTI crude oil prices. These two datasets meet the criteria since the latter contains an outlier and the temperature data is a clean time series without any outliers. These two real datasets are decomposed using EMD, EEMD, SEMD, CEEMDAN, VMD, and the novel Akima-EMD method. Following the decomposition of the nonstationary and nonlinear daily time series data into several IMFs and a single residue, the first IMF is removed from the real data to determine that this component contains no relevant information. According to the empirical results of performance indicators such as RMSE, MAE, and MAPE presented in Tables 6 and 7, the proposed method effectively extracts noise (useless information) from data in the form of the first IMF component. In a summary, both simulation

studies and real-world data examples indicate that the new approach, known as Akima-EMD, outperforms previous nonlinear and non-stationary data decomposition methods such as EEMD, SEMD, CEEMDAN, VMD, and the well-known benchmark EMD.

Therefore, it should be considered a new tool in the existing literature and could be used as an efficient technique for decomposing and forecasting the nonlinear and nonstationary time series.

VII. CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] A. L. B. Cauchy, *Theory of Wave Propagation on the Surface of a Heavy Fluid of Indefinite Depth*, vol. 1. Royal Academy of Sciences de France, 1827, pp. 5–318.

[2] A. Graps, "An introduction to wavelets," *IEEE Comput. Sci. Eng.*, vol. 2, no. 2, pp. 50–61, Jun. 1995.

[3] E. O. Brigham, *The Fast Fourier Transform and Its Applications*. London, U.K.: Pearson, 1988.

[4] M. B. Priestley, *Spectral Analysis and Time Series, Two-Volume Set*. New York, NY, USA: Academic, 1983.

[5] S. Mallat, *A Wavelet Tour of Signal Processing*. New York, NY, USA: Academic, 2008.

[6] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: SIAM, 1992.

[7] B. Vidakovic, *Statistical Modeling by Wavelets*. Hoboken, NJ, USA: Wiley, 2009.

[8] 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. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 454, pp. 903–995, Mar. 1971.

[9] N. E. Huang and S. S. P. Shen, "Hilbert-huang transform and its applications," in *Interdisciplinary Mathematical Sciences*. Singapore: World Scientific, 2005.

[10] A. O. Boudraa and J. C. Cexus, "EMD-based signal filtering," *IEEE Trans. Instrum. Meas.*, vol. 56, no. 6, pp. 2196–2202, Dec. 2007.

[11] N. E. Huang and Z. Wu, "A review on Hilbert–Huang transform: Method and its applications to geophysical studies," *Rev. Geophys.*, vol. 46, no. 2, 2008, pp. 1–23.

[12] M. Dätig and T. Schlurmann, "Performance and limitations of the Hilbert–Huang transformation (HHT) with an application to irregular water waves," *Ocean Eng.*, vol. 31, nos. 14–15, pp. 1783–1834, Oct. 2004.

[13] R. B. Pachori and V. Bajaj, "Analysis of normal and epileptic seizure EEG signals using empirical mode decomposition," *Comput. Methods Programs Biomed.*, vol. 104, no. 3, pp. 373–381, 2011.

[14] H. Huang and J. Pan, "Speech pitch determination based on Hilbert–Huang transform," *Signal Process.*, vol. 86, no. 4, pp. 792–803, Apr. 2006.

[15] I. O. Morales, E. Landa, P. Stránský, and A. Frank, "Improved unfolding by detrending of statistical fluctuations in quantum spectra," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 84, no. 1, Jul. 2011, Art. no. 016203.

[16] G. Li, W. Bu, and H. Yang, "Research on noise reduction method for ship radiate noise based on secondary decomposition," *Ocean Eng.*, vol. 268, Jan. 2023, Art. no. 113412.

[17] X. Zhang, K. K. Lai, and S.-Y. Wang, "A new approach for crude oil price analysis based on empirical mode decomposition," *Energy Econ.*, vol. 30, no. 3, pp. 905–918, May 2008.

[18] A. Kang, Q. Tan, X. Yuan, X. Lei, and Y. Yuan, "Short-term wind speed prediction using EEMD-LSSVM model," *Adv. Meteorol.*, vol. 2017, pp. 1–22, Jan. 2017.

[19] W.-C. Wang, K.-W. Chau, D.-M. Xu, and X.-Y. Chen, "Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition," *Water Resour. Manage.*, vol. 29, no. 8, pp. 2655–2675, 2015.

[20] L. Yu, W. Dai, and L. Tang, "A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting," *Eng. Appl. Artif. Intell.*, vol. 47, pp. 110–121, Jan. 2016.

[21] Z. H. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–14, 2009.

[22] D. Kim, K. O. Kim, and H.-S. Oh, "Extending the scope of empirical mode decomposition by smoothing," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, pp. 1–17, Dec. 2012.

[23] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 4144–4147.

[24] Y. Li, B. Tang, and S. Jiao, "Optimized ship-radiated noise feature extraction approaches based on CEEMDAN and slope entropy," *Entropy*, vol. 24, no. 9, p. 1265, Sep. 2022.

[25] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.

[26] Y. Li, B. Tang, and Y. Yi, "A novel complexity-based mode feature representation for feature extraction of ship-radiated noise using VMD and slope entropy," *Appl. Acoust.*, vol. 196, Jul. 2022, Art. no. 108899.

[27] Y. Li, B. Tang, X. Jiang, and Y. Yi, "Bearing fault feature extraction method based on GA-VMD and center frequency," *Math. Problems Eng.*, vol. 2022, pp. 1–19, Jan. 2022.

[28] R. Bisoi, P. Dash, and A. Parida, "Hybrid variational mode decomposition and evolutionary robust kernel extreme learning machine for stock price and movement prediction on daily basis," *Appl. Soft Comput.*, vol. 74, pp. 652–678, Jan. 2019.

[29] T. Liu, X. Ma, S. Li, X. Li, and C. Zhang, "A stock price prediction method based on meta-learning and variational mode decomposition," *Knowl.-Based Syst.*, vol. 252, Sep. 2022, Art. no. 109324.

[30] G. Li, C. Zheng, and H. Yang, "Carbon price combination prediction model based on improved variational mode decomposition," *Energy Rep.*, vol. 8, pp. 1644–1664, Nov. 2022.

[31] H. Yang, W.-S. Shi, and G.-H. Li, "Underwater acoustic signal denoising model based on secondary variational mode decomposition," *Defence Technol.*, Oct. 2022, doi: 10.1016/j.dt.2022.10.011.

[32] F. Liu, G. Li, and H. Yang, "A new feature extraction method of ship radiated noise based on variational mode decomposition, weighted fluctuation-based dispersion entropy and relevance vector machine," *Ocean Eng.*, vol. 266, Dec. 2022, Art. no. 113143.

[33] P. J. Luukko, J. Helske, and E. Räsänen, "Introducing libeemd: A program package for performing the ensemble empirical mode decomposition," *Comput. Statist.*, vol. 31, no. 2, pp. 545–557, 2015.

[34] D. Kim and H.-S. Oh, "EMD: A package for empirical mode decomposition and Hilbert spectrum," *R J.*, vol. 1, no. 1, p. 40, 2009.

[35] L. Mouselimis. (2022). *VMDecomp: Variational Mode Decomposition Using R. R Package Version 1.0.1*. [Online]. Available: https://CRAN.R-project.org/package=VMDecomp

[36] H. Akima, "A new method of interpolation and smooth curve fitting based on local procedures," *J. ACM*, vol. 17, no. 4, pp. 589–602, Oct. 1970.

MUHAMMAD ALI is currently pursuing the Ph.D. degree with the Department of Statistics, Faculty of Numerical Sciences, Abdul Wali Khan University Mardan, Mardan, Pakistan. His research interests include time series analysis, forecasting, machine learning, and econometrics modeling.

DOST MUHAMMAD KHAN (Senior Member, IEEE) received the Ph.D. degree in statistics from the University of Peshawar, Pakistan. He was a Visiting Fellow with the School of Statistics, University of Minnesota, USA, during the Ph.D. degree. He is currently an Associate Professor and the Head of the Department of Statistics, Abdul Wali Khan University Mardan, Mardan, Pakistan. He has published over 70 research articles in the field of robust statistics, high-dimensional data analysis, machine learning, and distribution theory. His research interests include robust regression, applied statistics, computational statistics, time series, and machine learning. He is an Associate Editor of the IEEE Access journal. He has also guest edited many special issues in many reputable international journals. He is also an active reviewer for various well-reputed international journals.

IMRAN SAEED received the Ph.D. degree from the Centre for Soft Matter Physics, Institute of Basic Sciences, Ulsan National Institute of Science and Technology, South Korea. His research interests include signal processing, condensed matter physics, and fluid dynamics.

HUDA M. ALSHANBARI received the Ph.D. degree in statistics from the University of Leeds, U.K. She is currently an Associate Professor with the Department of Mathematical Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia. Her research interests include applied statistics, computational statistics, the development of statistical methods and inference in the field of operation research, statistical bioinformatics, and machine learning.