An Effective SOMBased Approachfor low-Quality PMU Measures toRecognizeOscillatoryModes in Power System

Varnika Reshma Department of electrical Engineering NIT RourkelaOdisha, India221ee6465@nitrkl.ac.in

*Abstract***—To sustain the reliability of the connected-grid system's performance, it's crucial to spot any low frequencyoscillations that are poorly damped. But in actual,due to the increase in the installment of Phasor measurement units (PMUs), the data collected at Phasor data concentrators (PDCs) may contain non - linear data with missing values and outliers, which results in a decay in the performance of the estimator and affects system stability. This can happen due to heavy communication network congestion, malicious attacks, malfunctioning PMUs or PDCs, among other factors.In this situation, this paper discussesthe use of Self organizing maps (SOM) technique in the beginning to give a substantial data set to handle such a problem. The second phase involves identifying the oscillatory modes using the improved totalleast square-estimation of-signal parameters through rotational invariance technique (TLS-ESPRIT) algorithm.For generated test signals with missing data and outliers at various noise levels, the suggested method has been compared against Bayesian low-rank matrix/Hankel tensor factorization (BLMF/BLTF) with k-Medoids clustering and ML-based ERA (ML-ERA).The effectiveness of the suggested algorithm is further illustratedon two area data and real probing data collected from Western Electricity Co-ordinating Council (WECC).**

Keywords—PMU, SOM, TLS-ESPRIT, Modes Estimation

I. INTRODUCTION

One of the fundamentals enabling technologies in the monitoring, control, and protection of the nextgeneration connected grids is the Wide-Area Measurement System (WAMS), which uses phasor measuring units (PMUs) [1] [2]. PMUs offer synchronized phasor measurements at multiple points in the power system, and modes estimation is possible with the help of these data. But the increase in the complexity of connected grid system anddue to the ongoing expansion of PMU deploymentregarding the design and implementation of a successful wide-area communication system has resulted in obstructionsin the communication connection, transmission delays, hardware issues, malfunctioning PMUs or PDCs, and cyberattacks [3].This expansion in installation of networked PMUs has caused low quality real-time data flow across PDC. Low quality data creates a large range of unknown probability for data loss, data quality degradation and outlies. Outliers and missing data reduce the system's observability, affectingthe system's reliability andsecurity, which reduces the accuracy of modes estimations and the identification and verification of system parameters[4]. Recovery of missing PMU measurements and the removal of outlies is now a serious and inescapable issue in power systems.

Shekha Rai Department of electrical Engineering NIT Rourkela Odisha, India rais@nitrkl.ac.in

To address this issue, researchers have developed various methods to detect corrupted fragments insynchrophasormeasurements of time-series data, such as usingKalman filter [5], deep autoencoder [6], fast-Fouriertransform (FFT) [7], the Prony methods [8] [9], the estimation of signal parameters via rotational invariance techniques (ESPRIT)[10], support vector machine (SVM) [11], and robust-principal-componentanalysis (RPCA) [12].Among these, the precision of FFTis less for data with a handful of samples.The Kalman filter has the issue of numerical instability. Prony approach fails to maintain its accuracy when the signal has a large distortion component. Beside this method researchers have also usedBLMF/BLTF with adaptive k-Medoids clustering[13] and ML-ERA [14] to reconstruct the time series data by filling of missing value and removing the outliers. Although low-rank approaches theoretically promise accurate recovery if the loss percentage is below a certain threshold, but such a bound typically undervalues the methods' capabilities and is thereforeimpractical.BLMF/BLTFfeatures a straightforward starting parameter selection, increased recovery precision, and greater flexibility but when there is continuous missing value in the data BLMF/BLTF fails to maintain its accuracy. ML-ERA estimator is capable to handle missing values and outliers of PMU observation for oscillations of low frequency in power systems., but the accuracy is very low.

This paper employs a robust self-organizing maps (SOM) technique to address both outliers and missing data in the degraded PMU signal. To anticipate missing values in time series data, the SOM method is initially implemented. Afterwards, the SOM algorithm is used to identify and eliminate outlier values from the data. SOM determines the data set's pattern [15], and using the pattern, predicts the values that are missing. Toestimate the modes, the reconstructed signal is finally run via TLS-ESPRIT. It is compared to the BLMF/BLTF approach and the ML-ERA method to see how robust the method is.

The proposed technique's approach is provided in Part 2, followed by an analysis of the results in Section 3, and finally discussion of the conclusion in Section 4.

II. METHODOLOGY OF PROPOSED SOM TECHNIQUE

A. Self Organizing Map (SOM)

A SOM is a single layer neural network with units arranged on an n-dimensional grid.There are two layers in a SOM network. The first one, known as the "Input Layer," has as many nodes as there are parameters. Each input node has a weight constant connection to the nodes in the "Output Layer". The "Output Layer" should have a small number of

nodes to ensure quick computation and a large number ofnodes to show any relationships between the parameters [16].Garcia and Gonzalez in [16] provides advice on figuring out the ideal number of neurons by using formula,

$$
N = \sqrt{M} \tag{1}
$$

Where M is the total numbers of data in a sample. After the determination of the numbers of neurons \hat{N} , rows and columns the SOM can be figured out by using formula,

$$
\frac{R}{c} = \sqrt{\frac{E_1}{E_2}}\tag{2}
$$

Where R and C are rows and columns respectively and E_1 and E_2 are the eigen values of the training dataset.

B. SOM Algorithm forReplacing Missing Values and Removal of Outliers.

SOM networks pick up on patterns and correlations in their input over time and adjust how they respond to it in the future.The algorithm starts by randomly initializing a set of neurons, each with a weight vector that represents a point in the feature space of the dataset [17] [18]. During training, the SOM adjusts these weight vectors to better represent the underlying structure of the dataset.

Steps involved in SOM technique:

Step 1:Initialize SOM

- The number of neurons isdefined andthe weight vectors for each neuron are randomlyinitialize.
- Neighborhood function and learning-rate is specified for the weight updates.
- The learning rate of the SOM trainingconsists of a variable that declines from a value of α_0 with more iterations and is expressed as

$$
\propto (t) = \alpha_0 \times e^{-t\lambda} \tag{1}
$$

Where α_0 is initial learning rate and λ is learning decay rate.

 The neighborhood function is defined using as its initial value and is equivalent to Equation.

$$
\phi(t) = e^{\frac{-D(x,y)^2}{2\sigma(t)^2}} \tag{2}
$$

Step 2:Train the SOM

- Random instances are selected from the dataset.
- Best matching unit (BMU) is calculated between the instance and all the neurons by looking for the node that isnearest to the existing input vector in respect of Euclidean measure. BMU on the map is given by the equation,

$$
D(x, y) = \sqrt{\sum_{i=1}^{k} (x_i - y_1)^2}
$$
 (3)

Where k is the dimension of vectors x, y . Step 3: Update neighbourhood distance weight

 All nearby nodes are updated based on a decreasing learning rate \propto (*t*)with 0≤∝ (*t*)≤ 1. The weight vector $w_i(t)$ is updated as

 $w_i(t + 1) = w_i(t) + H_{ci}(t) \cdot (x(t) - w_i(t))$ (4) Where $H_{ci}(t)$ is neighborhood distance weight and is expressed as,

$$
H_{ci}(t) = \alpha(t) \cdot \phi(t) \tag{5}
$$

Where \propto (t) is the learning rate and $\phi(t)$ is theneighborhood function.

 This process is repeated for all instances in the dataset, for anumber of iterations.

Step 4: Filling of missing data

- For each instance with missing values,the distances is calculated between the instance and all the neurons in the SOM.
- Node with the closest weight vector to the instance is selected.
- Values from the weight vector of the chosen node is used to fill in the instance's missing values.

Step 5: identify outlies

- For each instance in the dataset, the distances are calculated between the instance and all the neurons in the SOM.
- Instances with distances greater than a certain threshold value are identified as outliers.Threshold value can be determined using equation,

$$
TD = mean(D) + 2 \cdot std(D)
$$

Where Dis BMU distance.

If

$$
D(x) = \begin{cases} \text{not outlines, } D(x) < TD \\ \text{outliers, } D(x) \geq TD \end{cases} \tag{6}
$$

STEP 6: Evaluate the result

- The performance of the filled dataset is checked using evaluation metrics such as mean-squared error or meanabsolute error.
- If the performance is satisfactory,the filled dataset is used for further analysis.
- The performance of the outlier is checked detection using evaluation metrics F1 score.

C. SOM Flowchart Representing the Steps Involved in Algorithm

Fig. 1. Flow chart for SOM

D. TLS-ESPRIT for Modes Identification

Theexistence of missing data and outliers in the estimation of modes through TLS-ESPRIT can lead to inaccurate estimates of modes. Hence, before sending the signal to TLS-ESPRIT algorithm, the signal is treated for outliers and missing values by using SOM Algorithm. The TLS-ESPRIT [19] builds a strong auto-correlation matrix to accommodate the PMU's imprecise measurements using the whole data set acquired via the SOM technique. The realtime signal simulation is shown as

$$
s(n) = \sum_{k=1}^{m} A_k e^{-\sigma_k T_p} \cos(2\pi f_k n T_p + \theta_k)
$$
 (7)

Where, T_p sampling time-period; A_k is amplitude; f_k is frequency and θ_k is phase angle and σ_k is the damping factor.

III. BLOCK DIAGRAM FOR SOMTECHNIQUE

PMU technology transmits time-stamped data in accordance with a single clock that originates from the GPS satellite network. Data from PMU is transmitted to PDC using a wired line. Certain outliers and missing values are produced in the data during data transmission to the PDC owing to communication breakdowns, PMU issues, etc. To obtain complete data, the SOM technique is applied to the partial data. For mode estimation, the SOM algorithm's output data is sent using TLS-ESPRIT. The block diagram of the proposed mode estimation technique is presented below in Fig.2.

Fig. 2. Block Diagram for Proposed Algorithm

IV. RESULTS AND DISCUSSION

The suggested technique is implemented on the degraded test signal having local-area and inter-area modes of oscillations, on four generators two-area data and lastly, real time probing data.To validate the proposed technique, a comparative statistical analysis studyat different noise levelsis carried out with BLMF/BLTFand ML-ERAbased technique.

A. Modes Estimation Of Signal Oscillating In Inter Area Mode

Using MATLAB, a test signal simulation with 0.7 Hz frequency and -0.1damping is performed. Theoutliers are purposely inserted and some data are removedfrom the signal to corrupt the signal as seen in Fig. 3. The incomplete signal is then degraded by the addition of zero mean gaussian noise from 20 dB to 40 dB such that itdemonstrates the agility of the suggestedapproach in noisy circumstance. The modes are estimated using various methods using 10000 independent monte carlo cycles with various amounts of noise, and the results are then statistically analysed and manifested in Table I, thus showing that the proposed scheme gives model estimate close to the true value with less variance.

The reconstructed signal, produced from the modes predicted at a 40dB SNRusing various approaches and the suggested approach, is shown in Fig. 4. It is observed that the clean signal and reconstructed signal by proposed method overlap each other, thus showing the robustness of proposed technique.The distribution of the signal's attenuation factor for different approach at SNR value of 40dB for the interarea mode is shown in Fig. 5.The suggested technique can estimate more precise modes and closer values to the actual values of damping (-0.1) and frequency (0.7 Hz).The proposedtechnique yielded the mode frequency and damping of 0.7006Hzand -0.0996 respectively.

Fig. 3. Simulated distorted Signal with Inter-Area Modes of Oscillation

TABLE I. ESTIMATED MODE FREQUENCY AND DAMPING FOR TEST SIGNAL WITH INTER-AREA MODES OF OSCILLATION

Fig. 4. Reconstructed signal obtained by simulating modes obtained at 20dB SNR

While the modes found using the BLMF/BLTF and ML-ERA approaches had bigger deviations (0.7091Hz, -0.1052) and (0.7239Hz, 0.0917), respectively. Hence, the statistical information shown in Table 1 demonstrates that the suggested technique is superior to existing techniques for mode estimation.

B. Signals for Test Matching to local Area Mode

A test signal which has a true frequency value of 1.5 Hz and damping of -0.05 with missing values and outlies is passed through the proposed algorithm. By using 10000 Monte Carlo simulation cycles, the statistical analysis was performed which is shown in Table 2. In comparison to the other two approaches, the proposed methodology can measure the modes more precisely. Even though there is more noise, the suggested method produces better outcomes with less variance. The proposed technique has successfully passed the robustness check with mean damping of -0.0497 and mean frequency of 1.5028 Hz which is almost closer to the true value of both the frequency and the damping.

The attenuation factor plot for each of the proposed approach and other two approach at 40 dB SNR is shown in Fig. 6.The plot demonstrates that SOM approach exhibits very precise damping as well as frequency measurements with reduced fluctuation, demonstrating the feasibility of the suggested method.

TABLE II. ESTIMATED MENAS(μ) AND VARIANCE(S2) FOR TEST SIGNAL WITH LOCAL AREA MODE OF OSCILLATION

True Frequency= 1.5 Hz and True Damping= -0.05								
BLMF/BLTF								
	Damping		Fequency (Hz)					
SNR (dB)	Mean	Variance	Mean	Variance				
20	-0.0439 1.631×10 ⁻⁶		$1.51565.699\times10^{-8}$					
30	-0.0439 1.669×10 ⁻⁷		$1.51575.659\times10^{-9}$					
40	-0.0439 1.670 \times 10 ⁻⁸		$1.51575.606\times10^{-10}$					
ML-ERA								

Fig. 5. Inter Area Plot Representing Mean and Variance for Mode Frequency and Damping at SNR = 40 dB

Fig. 6. Local Area Plot Representing Mean and Variance for Mode Frequency and Damping at SNR = \$

C. Two Area Data Signal

TABLE III. ESPIMATION OF MODES FOR TWO AREA DATA

According to Fig. 7, mode estimate is put through, for a two-area data signal of power system [16].It consists of 2 loads, 11 buses and 4 generators in this system.Generators 1 and 2 together constitute area 1, whereas Generators 3 and 4 together form area 2.The forecasted modesof the power in line 10 to 9 which has low-frequency (LF) oscillationsare shown in Table III.The data required to estimate the mode is acquiredfrom the bus 8 load loss of 10 MW, which results in LF oscillations at bus 9 with outlier and missing value. After passing the degraded signal data through the proposed technique and other two technique, it is run through 10,000 Monte Carlo runs,to predict the mode and variance whichare displayed in Table III at $SNR = 40dB$. The comparative analysis of modes has led researchers to the conclusion that the suggested method can estimate correct modes that are true modes as mentioned in [20], proving the technique's robustness.

Fig. 7. Two Area Data Represented in Single-Line-Diagram

D. Probing Data Sourced from WECC

Fig. 8. Pobing Data Recorded by PMU Linked to WECC

The effectiveness of proposed technique is examined and compared with BLMF/BLTFand ML-ERA technique by using data from the WECC system's probing test that was conducted on September 14th, 2005 [21] as shown in Fig. 8.The data collected during the 1st and 2nd successive signal mode probing of 125 MW, correspond to window-1 and window-2 respectively.The estimated mode frequency was detected as 0.318 (Hz) with adamping of 8.3%, in [22]. Analysis of the probing data for both window 1 and window 2 was done at a SNR value of 40dB.

Table IV shows that the reliability of the suggested technique is markedly larger compared to the BLMF/BLTF and ML-ERA techniques for each search window. The proposed method's percentage damping obtained at windows1is 8.2860% and at window 2 is 8.2988 %, which is closer to the actual damping value measured in [22].

TABLE IV. ESPIMATION OF FREQUENCY AND PERCENTAGE DAMPING FOR PROBING DATA

WINDOW	MODES	BLMF/BLTF	ML-ERA	SOM
1	FREQ.(HZ)	0.3352	0.3408	0.3163
	$DAMP.$ %	9.3068%	9.4491%	8.2860%
2	FREQ.(HZ)	0.3202	0.2829	0.3190
	$DAMP.$ %	10.9578 %	12.9561 %	8.2988%

V. CONCLUSION

This paper suggests SOM algorithm to supervise the deviatedvalues and missing values in the sample data extracted from PMU. The proposed method employs the SOM algorithm to reconstruct the depleted PMU signal.The improved TLS-ESPRIT is then used to evaluate the modes of the reconstructed signal, and it has been demonstrated that it's effective in higher-noise conditions.The performance of theproposed method has been assessed using

known test signals, the 2-area system, and real PMU data obtained from the WECC system with the data sample corrupted with both an outlier and a continuous missing value in order to demonstrate its effectiveness. With respect to the simulation results shown in Part IV, it can be stated that the suggested method, as compared to BLMF/BLTF and ML-ERA, is robust towards degraded PMUmeasurement and hence is suitable for PMU-based wide area measurement system.

REFERENCES

- [1] M. Liao, D. Shi, Z. Yu, Z. Yi, Z. Wang and Y. Xiang, "An Alternating Direction Method of Multipliers Based Approach for PMU Data Recovery," in IEEE Transactions on Smart Grid, vol. 10, no. 4, pp. 4554-4565, July 2019, doi: 10.1109/TSG.2018.2864176.
- [2] Yi, Ming, Meng Wang, Evangelos Farantatos, and Tapas Barik. "Bayesian robust hankel matrix completion with uncertainty modeling for synchrophasor data recovery." ACM SIGENERGY Energy Informatics Review 2, no. 1 (2022): 1-19.
- [3] Pei, Jianhua, Ziyu Wang, Jingyu Wang, and Dongyuan Shi. "Robust fast PMU measurement recovery enhanced by randomized singular value and sequential Tucker decomposition." IET Generation, Transmission & Distribution 16, no. 16 (2022): 3267-3281.
- [4] Shi, Di, Daniel J. Tylavsky, Kristian M. Koellner, Naim Logic, and David E. Wheeler. "Transmission line parameter identification using PMU measurements." European Transactions on Electrical Power 21, no. 4 (2011): 1574-1588.
- [5] Korba, P., M. Larsson, and Ch Rehtanz. "Detection of oscillations in power systems using Kalman filtering techniques." In Proceedings of 2003 IEEE Conference on Control Applications, 2003. CCA 2003., vol. 1, pp. 183-188. IEEE, 2003.
- [6] Wang, Jingyu, Dongyuan Shi, Yinhong Li, Jinfu Chen, Hongfa Ding, and Xianzhong Duan. "Distributed framework for detecting PMU data manipulation attacks with deep autoencoders." IEEE Transactions on smart grid 10, no. 4 (2018): 4401-4410.
- [7] Girgis, Adly A., and Fredric M. Ham. "A quantitative study of pitfalls in the FFT." IEEE Transactions on Aerospace and Electronic Systems 4 (1980): 434-439.
- [8] Khodadadi Arpanahi, Moossa, Meysam Kordi, Roozbeh Torkzadeh, Hassan Haes Alhelou, and Pierluigi Siano. "An augmented prony method for power system oscillation analysis using synchrophasor data." Energies 12, no. 7 (2019): 1267.
- [9] Paternina, Mario R. Arrieta, Rajesh Kumar Tripathy, Alejandro Zamora-Mendez, and Daniel Dotta. "Identification of electromechanical oscillatory modes based on variational mode decomposition." Electric Power Systems Research 167 (2019): 71-85.
- [10] Chen, Jian, Tao Jin, Mohamed A. Mohamed, and Mengqi Wang. "An adaptive TLS-ESPRIT algorithm based on an SG filter for analysis of low frequency oscillation in wide area measurement systems." IEEE Access 7 (2019): 47644-47654.
- [11] Ozay, Mete, Inaki Esnaola, Fatos Tunay Yarman Vural, Sanjeev R. Kulkarni, and H. Vincent Poor. "Machine learning methods for attack detection in the smart grid." IEEE transactions on neural networks and learning systems 27, no. 8 (2015): 1773- 1786.
- [12] Mahapatra, Kaveri, Nilanjan Ray Chaudhuri, Rajesh G. Kavasseri, and Sukumar M. Brahma. "Online analytical characterization of outliers in synchrophasor measurements: A singular value perturbation viewpoint." IEEE Transactions on Power Systems 33, no. 4 (2017): 3863-3874.
- [13] Pei, Jianhua, Jingyu Wang, Ziyu Wang, and Dongyuan Shi. "Precise Recovery of Corrupted Synchrophasors Based on Autoregressive Bayesian Low-Rank Factorization and Adaptive K-Medoids Clustering." IEEE Transactions on Power Systems (2022).
- [14] Singh, Bikash Swarup Bidyasagar, and Shekha Rai. "An MLbased ERA Algorithm for Estimation of Modes Utilizing PMU Measurements." In 2022 3rd International Conference for Emerging Technology (INCET), pp. 1-5. IEEE, 2022.
- [15] Miljković, Dubravko. "Brief review of self-organizing maps." In 2017 40th international convention on information and

communication technology, electronics and microelectronics (MIPRO), pp. 1061-1066. IEEE, 2017.

- [16] Rustum, Rabee, and Adebayo J. Adeloye. "Replacing outliers and missing values from activated sludge datausing Kohonen selforganizing map." Journal of Environmental Engineering 133, no. 9 (2007): 909-916.
- [17] Riese, Felix M., and Sina Keller. "Introducing a framework of self-organizing maps for regression of soil moisture with hyperspectral data." In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 6151-6154. IEEE, 2018.
- [18] Riese, Felix M., Sina Keller, and Stefan Hinz. "Supervised and semi-supervised self-organizing maps for regression and classification focusing on hyperspectral data." Remote Sensing 12, no. 1 (2019): 7.
- [19] J. Chen, T. Jin, M. A. Mohamed, and M. Wang, "An Adaptive TLS-ESPRIT Algorithm Based on an S-G Filter for Analysis of Low Frequency Oscillation in Wide Area Measurement Systems," *IEEE Access*, vol. 7, pp. 47644–47654, 2019, doi: 10.1109/ACCESS.2019.2908629.
- [20] P. Kundur, Power System Stability and Control. New York: McGrawHill, 1994.
- [21] PDCI Probe Testing Plan, 2005 [Online]. Available: [http://www.transmission.bpa.gov/business/operation](http://www.transmission.bpa.gov/business/operations/System%20News/) [s/System News/](http://www.transmission.bpa.gov/business/operations/System%20News/).
- [22] Report and data of WECC. [Online]. Available: ftp://ftp.bpa.gov/pub/WAMS Information/.