2023 2nd International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN) | 979-8-3503-4800-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSTSN57873.2023.10151498

An XgBoost based method for identifying electromechanical oscillations from ambient measurements using WAMS

Imtiaz Kumar Pradhan
Electrical Engineering Department
NIT Rourkela
Odisha, India
119ee0259@nitrkl.ac.in

Sameer Ranjan Sahu
Electrical Engineering Department
NIT Rourkela
Odisha, India
119ee0266@nitrkl.ac.in

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

***Abstract*—A novel method for online identification of modes corresponding to low frequency oscillations (LFOs) in power systems is put forth in this paper. The effect of colored noise is created by the filters employed in phasor measurement unit(PMU) has been taken into account in the ambient signal measurement. To illustrate the productivity and robustness of the suggested methodology, a comparison study of the simulation results of the proposed method with other developed techniques such as RD-ESPRIT and K-SVD has been undertaken using synthetic test signals representing ambient data. Comparison of this proposed method with other methods is also performed on two-area data simulated on Real Time Digital Simulator (RTDS) and real measurements of Western Electricity Coordinating Council (WECC).**

***Index Terms*—XgBoost, wide area measurement system (WAMS), mode identification.**

I. INTRODUCTION

Low frequency oscillations are one of the key factors affecting the stability of small signals in the power system. These oscillations are under the inter area (0.1–0.7 Hz) category and local area(0.8-2 Hz) category [1]. Due to insufficient damping, if these oscillations continue in the network, it could cause cascading network trips. Therefore, study of these low frequency oscillations to identify modes with lower damping coefficients is crucial for preserving the power system's small signal stability. The maintenance of stability is one of the primary operational concerns in today's massive, interconnected power systems. It has proven essential to use modal analysis to evaluate low-frequency oscillations or the system's small signal stability. Real-time monitoring of these low frequency modes is essential to ensuring that their damping is sufficient; if it isn't, the appropriate preventive control measures must be taken.The presence of low frequency oscillatory modes is inferred using synchronised readings from phasor measuring units (PMUs) [2]. The methods used to analyse these modes either use ambient data [3] or ringdown. When a fault or other severe disturbance occurs, methods for estimating the modes that use ringdown data, such as Prony [4], ESPRIT [5], kalman filter [6], etc., provide reliable estimations of the modes that are present in the system, but because the operating point of the power system is constantly changing,the system's current stability situation may not be provided by the estimated modes, but it can be provided by approaches that uses ambient data. Much study has been done on estimating the low frequency modes using ambient data, including approaches such as Ibrahim time domain (ITD) [7] and other methods like RDT [8], K-SVD [3], Sparsity [9], NExT–ERA [9] etc.

Due to the PMU's use of an anti-aliasing filter before sampling, the power signals collected from the PMU have an additive white noise component (showing the PMU's uncertainty) and a small component of additive coloured noise (usually ignored). Filtering away the high frequency components at the PMU output and keeping only the low-frequency components is essential for effective and quick mode identification procedures. It is therefore necessary to utilise a low pass filter, whose output is highly correlated noise (or colored noise), which is created by converting the input white noise into a filter with a cutoff frequency that is closer to the desired mode frequencies. By introducing bias into the estimation process, this colored noise considerably lowers the accuracy of the predicted modes when utilising the current techniques. Therefore, it is necessary to create a mode identification algorithm that operates well in the presence of colored noise for ambient data.

The rest of the paper is organised as follows: Section II describes the methodology of the proposed XgBoost technology,in section III the online monitoring system that uses the suggested strategy is depicted, the comparative analysis of simulation results is presented in part IV, and the conclusion is presented in section V.

II. METHODOLOGY FOR THE PROPOSED APPROACH

A. XgBoost Algorithm

XgBoost is a gradient boosting algorithm. Classification or regression predictive modelling problems can be addressed by using a class of ensemble machine learning techniques called gradient boosting [10]. Gradient boosting starts with a single leaf. The initial estimation of each sample weight is shown on

the leaf. When attempting to predict a continuous number, an average value is often the initial presumption made. Gradient boost repeats the procedure and builds a new tree based on the errors of the previous tree.

Extreme Gradient Boosting, or XgBoost, is a successful opensource use of the gradient boosting technology. The model performance and execution speed are the two main advantages of XgBoost. XgBoost predicts base learners who are universally terrible at the rest in order for bad predictions to cancel out and better ones to add up to ultimate positive predictions [11].

B. Maths behind XgBoost Algorithm

In accordance with the XgBoost algorithm's basic tenets, the regularisation term $\Omega(f_k)$ must be computed to obtain the precise value of the objective function. To begin, a decision tree must be defined. It is denoted by f_k, and its precise representation by the formula (1).

$$f_k(x) = \phi_{p(x)}, \ \phi \in R^T, \ p : R^d \longrightarrow \{1, 2, 3..., T\} \quad (1)$$

Where in this tree the number of leaf nodes present is denoted by T, the function p identifies each of the input sample belongs to which of the following leaf nodes, and ϕ is a vector that is used to store each of the leaf node's weight. Afterward, the regularisation term in the XgBoost model is represented by the formula (2).

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \phi_j^2 \quad (2)$$

From formula (2) it can be observed that the regularization term consists of γ, λ and the sum of squares of L2 norm of leaf node weight ϕ and number of leaf nodes T. In order to manage the complexity of the XgBoost model, parameters and are employed. The model will be more cautious when the value of γ is higher. The reduction of loss function used to govern node splitting is called γ. As a result, the more conservative the associated model will be, the greater the γ value is. A leaf node, for instance, won't divide if the decrease of loss function becomes less. The kth tree's objective function is represented as :-

$$obj^{(k)} \approx \sum_{i=1}^{n} [g_i \phi_{p(x_i)} + \frac{1}{2} h_i \phi_{p(x_i)}] + \gamma T + \frac{1}{2}\lambda \sum_{j=1} \phi_j^2 \quad (3)$$

$$= \sum_{j=1}^{T} [(\sum_{i \in I_j} g_i)\phi_j + \frac{1}{2}(\sum_{i \in I_j} h_i + \lambda)\phi_j^2] + \gamma T \quad (4)$$

where $I_j = \{ p(x_i) = j \}$ is a collection of tags that includes the sample tags given to the training sample's jth leaf node. For instance, $I_2 = 1, 3, 5$ can be used to show that the second leaf node has been assigned the first, third, and fifth samples for training. Formula (4) can be simplified to:

$$obj^{(k)} = \sum_{j=1}^{T} [G_j \phi_j + \frac{1}{2}(H_j + \lambda)\phi_j^2] + \gamma T \quad (5)$$

where $G_j = \sum_{i \in I_j} g_i, \ H_j = \sum_{i \in I_j} h_i$

The optimal weight ϕ_j^* of the j^{th} leaf node and the optimal value obj^* of the objective function can be stated as follows since ϕ and other terms don't relate to each other.

$$\phi_j^* = -\frac{G_j}{H_j + \lambda} \quad (6)$$

$$obj^* = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j}{H_j + \lambda} + \gamma T \quad (7)$$

$$Gain = \frac{1}{2}[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} + \frac{(G_L + G_R)^2}{H_L + H_R + \lambda}] - \gamma \quad (8)$$

It is vital to comprehend how each tree in XgBoost divides specifically after having a basic grasp of the two indexes, loss function and regularisation term, that have an impact on the model's capacity to predict outcomes and complexity. The four parts of the formula (8) are the original leaf's weight, the weight that is present on the new left leaf, the weight that is present on the new right leaf, and the presence of regularisation on the added leaf , as can be seen. Considering how the loss function and regularisation are interpreted, it is desirable not to divide the node again if the gain of the node after splitting becomes $< \gamma$, which is equivalent to not adding the branch.This is the method of pruning that is used in the construction of decision trees.

C. Working of XgBoost

1) First, an initial prediction is made, and then the residual is computed. This prediction might be anything. The average value of the variables that has to be predicted serves as the initial prediction.

$$Residuals = Observed\ Values - Predicted\ Values \quad (9)$$

2) XgBoost Tree is then built. Every tree begins with a single leaf, which receives all of the residuals. Then similarity score of the leaf is calculated.

$$Similarity\ Score = \frac{(Sum\ of\ Residuals)^2}{Number\ of\ Residuals + \lambda} \quad (10)$$

where, λ is a regularisation parameter that mitigates data overfitting and reduces the prediction's sensitivity to particular observations. The default value is considered to be 1.

3) The residuals are divided into two groups and used thresholds depending on the predictors, one should check whether the residuals can be clustered more effectively. In essence, splitting the residuals entails growing our tree by adding new branches.

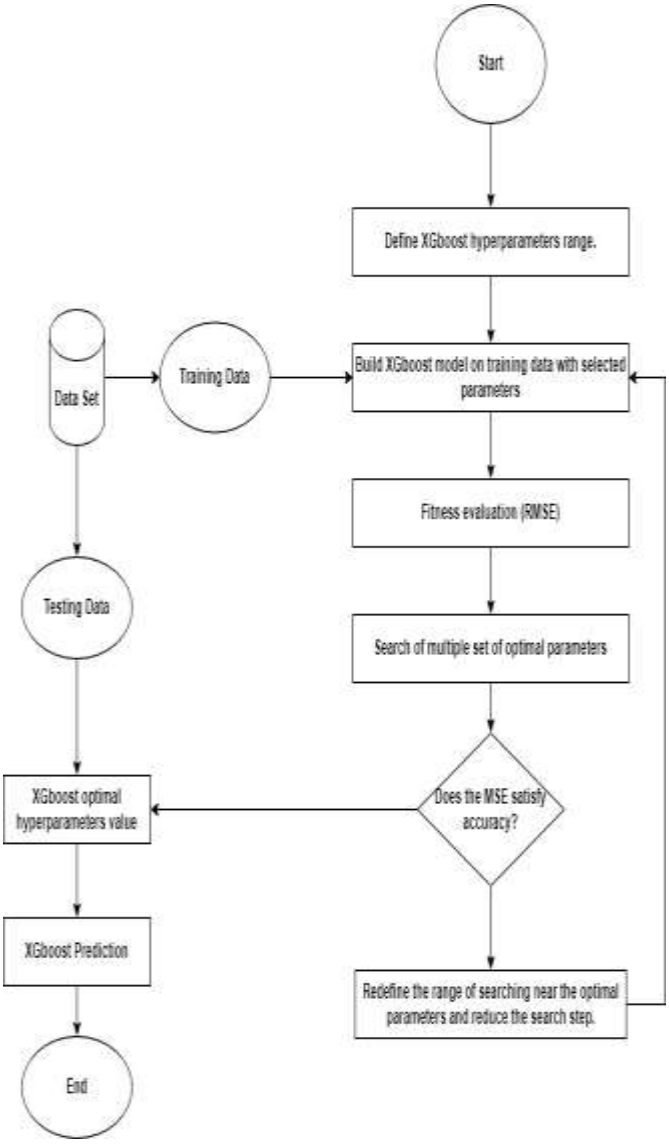

Fig. 1. A schematic of the XgBoost workflow [12]

4) The difference between how well the root and leaves cluster similar residuals then to be measured, by figuring out the Gain of dividing the residuals into two groups, this may be achieved. It is a good idea to split if the Gain is positive; otherwise, it is not.

$$Gain = Left_{similarity} + Right_{similarity} - Root_{similarity} \quad (11)$$

5) Tree will be pruned. It's another method to prevent overfitting of the data. This is done to determine if a split is valid or not by working the way up the tree starting at the bottom. γ is used to establish the validity. The split is kept if $Gain - \gamma$ is positive; otherwise, it is removed. The value of γ by default is 0.
6) One might now determine a single value in each of the leaf nodes because a leaf node cannot have more than one output. Basically, it is necessary to determine the output values of the leaves.

$$Output\ Value = \frac{Sum\ of\ Residuals}{Number\ of\ Residuals + \lambda} \quad (12)$$

The only difference between this and the procedure for calculating similarity score is that here, the residuals are not squared.

7) Last but not least, the predictions should be performed and all the predicted values are to be obtained. Using the updated predictions, the residuals are calculated. The more times it goes through this process, the smaller the residuals will become which indicates that the predicted values are approaching those of the observed values.
8) Now, the same procedure is repeated in a loop, building a new tree, making the predictions, and figuring out the residuals with each of the iteration. This process is continued until the residuals are extremely minimal or all of the algorithm's allowed iterations are finished. If the tree that is built at each iteration is denoted by T_i, where i is the current iteration, then to calculate the predictions the formula is given by:
$Initial\ Prediction + \eta(T_0 + T_1 + \ldots + T_i)$
where, η is the XgBoost Learning Rate.
9) The loop is terminated when the difference between predictions and the real values is within the permissible limit and hence the hyperparameters for the model is set.
10) Then this model is used in the test data for predictions. The associated clean signal is needed for the colored signal for the purpose of training. After the model is trained on these, prediction can be done for the clean signal for any signal corrupted with colored noise of the same type.

Thus, the output produced minimises the amount of coloured noise in the signal while preserving the important signal component.

D. Implementation of TLS-ESPRIT

The covariance matrix is divided into two orthogonal bases for TLS-ESPRIT implementation. For this, the singular value decomposition (SVD) technique is used. The signal subspace is made up of the eigenvectors connected to the predominate P eigen values. The estimated attenuation factor and frequency are derived by using the parameters of the first and second shift invariance signals.

III. Block Diagram of the Suggested Method for Locating Modes in Power Systems

The block diagram in Fig. 2 illustrates the sequential procedures for determining the online oscillatory modes of the power system. When there is a slight but sudden imbalance between generated power and needed power, the PMU can sense the the presence of oscillatory modes. The PDC receives the measured data and details about the power swing

through the communication connections. The inaccuracies in the measurements obtained from the PMUs are believed to be best approximated as additive white Gaussian noise. However, an anti-aliasing filter (with a cutoff frequency of 400–1000 Hz) and various signal processing methods that use convolution to estimate the phasors have discovered some additive colored Gaussian noise in the observations. The proposed approach starts off by using a block of N ambient data samples. The decision tree is created by passing these through our suggested XgBoost algorithm. The improved TLS-ESPRIT is then used to estimate the low frequency oscillatory mode.

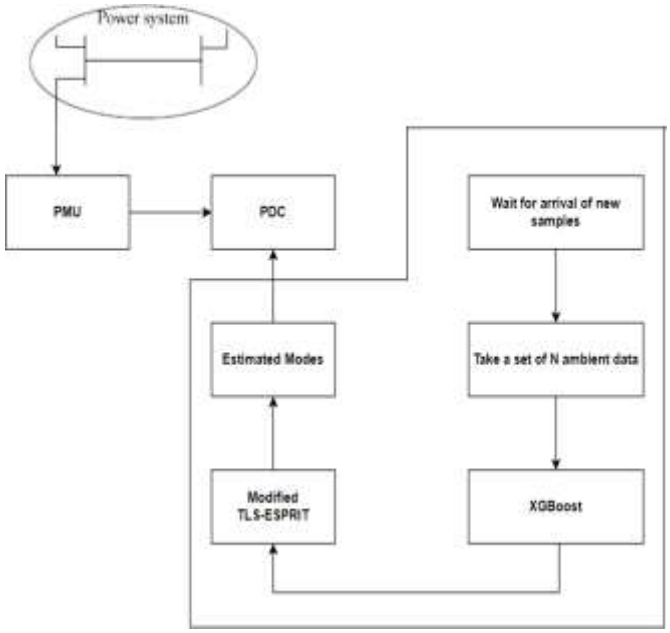

Fig. 2. Block diagram for mode identification

IV. Result Analysis

The simulation outcomes of the proposed scheme are contrasted with K-SVD and RD-ESPRIT to show its effectiveness and robustness. Utilizing a performance metric known as the total vector error (TVE) [13], the suggested method's accuracy was evaluated.

$$TVE(\%) = \frac{\sqrt{(X^e_r - X^a_r)^2 + (X^e_i - X^a_i)^2}}{\sqrt{(X^a_r)^2 + (X^a_i)^2}} \times 100 \quad (13)$$

Here X^a_i, and X^a_r represent the actual quantity's imaginary and real values,while X^e_i, X^e_r stands for the estimated parameters imaginary and real values.

A. Oscillatory mode estimation of ambient signal

In Fig. 3, damping and frequencies of -0.1 and 0.2 Hz are used to stimulate a second order system and yield test results that accurately represent the ambient signal. The output is then passed through a 5th order low pass butterworth filter to derive ambient measurements corrupted with colored noise as depicted in Fig. 4. Fig. 5 display the recovered signal using the suggested method. Table I compares the suggested XgBoost technique to RD-ESPRIT. The comparison analysis shows that the proposed strategy is superior to other ways in terms of efficiency where the results that is obtained are attenuation = -0.0992 at 0.1983Hz frequency and minimal percentage TVE = 0.8457.

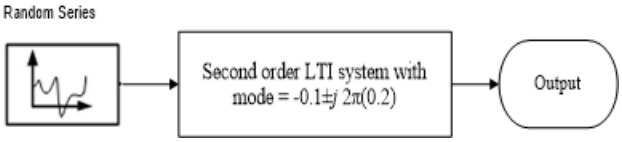

Fig. 3. SIMULINK model for generation of the ambient data

TABLE I
ESTIMATION OF THE SIMULATED SIGNAL'S MODE

Method	Attenuation	Freq	(%)TVE
RD-ESPRIT	−0.0912	0.1978	1.2937
K-SVD	−0.0953	0.1925	3.7389
XgBoost	−0.0992	0.1983	0.8457

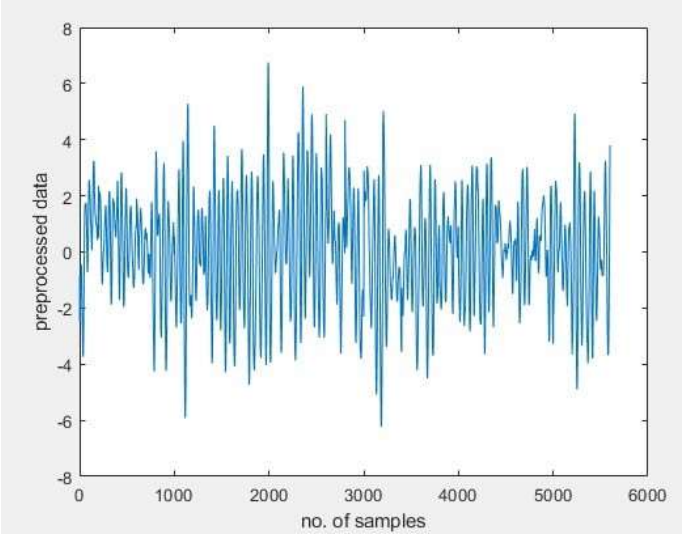

Fig. 4. Ambient signal corrupted with colored noise

B. Mode estimation for the two-area system

Real-time signals matching to interarea oscillations are produced using RTDS [13] [14]. The data generated for a 20-

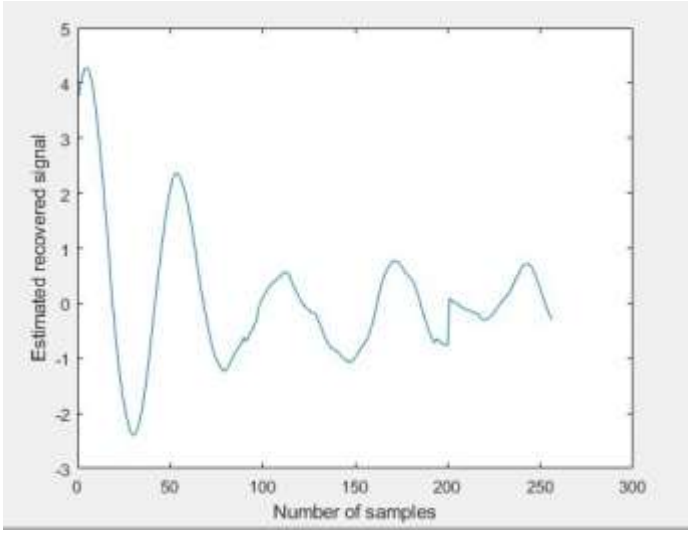

Fig. 5. Output obtained through proposed scheme

min period at a sample frequency of 50 Hz is shown in Fig. 6. The bus number 8's random load shifting in the range of 0–10 MW over a period of 1 second is what caused the low-frequency oscillations. The Table II shows that the suggested approach has a TVE of 3.4626%, which is lower than that of the other methods, and provides a close estimate of the dominant modes (i.e., damping = -0.1544 and frequency = 0.5261 Hz.)

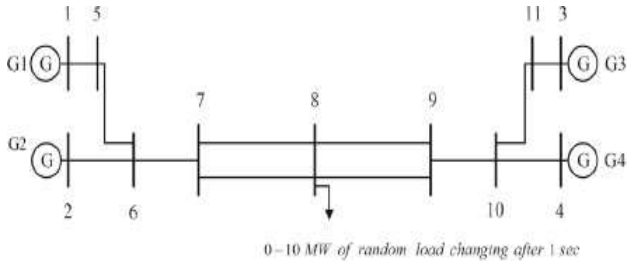

Fig. 6. Two-area system for ambient data

TABLE II
ESTIMATED MODES FOR GENERATED TWO-AREA SYSTEM

Methods	Attenuation	Freq	(%)TVE
RD-ESPRIT	−0.0933	0.5236	5.2137
K-SVD	−0.1203	0.5180	5.1750
XgBoost	−0.1544	0.5261	3.4626

C. Estimate the oscillatory modes using WECC probing data

The PMU attached to the WECC system provided the probed or monitoring data that is displayed in Fig. 7 and was acquired on September 14, 2005 [15]. It should be mentioned that the estimation data and the probing test signal data were obtained from [16], and the reported value is 8.3% of attenuation and frequency of 0.318Hz. The comparison of the simulation results is shown in Tables III and Table IV between the suggested technique and the alternative technique. The percentage damping determined using the suggested technique is found to be 8.7935% for the first window and 7.2953% for the second window. This value is extremely similar to the damping value provided in [15].The suggested technique has a minimum TVE of 0.5882% and 7.8581% for the first and second window respectively, and can estimate the mode at (0.3168 Hz and 0.2932 Hz) with a very high degree of accuracy. The results in Table III correlated noise, the proposed technique exhibits low change in the estimated modal parameters. Therefore, as shown in Table III and Table IV, the proposed method is a better option for the detection of poorly damped modes present in real-time signals when compared to RD-ESPRIT and K-SVD methods.

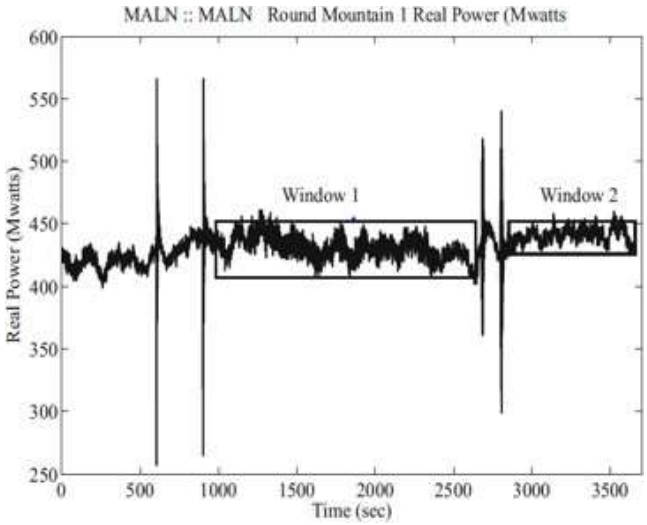

Fig. 7. WECC Probing Data

V. CONCLUSION

In this paper an ensembling boosting algorithm called XgBoost is used for the online identification of the modes corresponding to the LFOs in the power system. The presence of colored noise is less of an issue with this method, which enables a more reliable and precise estimation of the modes from ambient data. The modes obtained from this XgBoost method are closest to the true value than other methods like K-SVD, RD-ESPRIT for the synthetic test signal as evident from the minimum TVE error, thus validating its efficiency. Moreover, this scheme delivers a robust estimate of the modes for simulated two-area data in RTDS and real

TABLE III
ESTIMATED MODES FOR PRACTICAL DATA OF FIRST WINDOW

Methods	Attenuation	Freq	(%)Damping	(%)TVE
RD-ESPRIT	−0.1440	0.2990	7.6540	6.0084
K-SVD	−0.1877	0.3130	9.5044	1.8872
XgBoost	−0.1757	0.3168	8.7935	0.5882

TABLE IV
ESTIMATED MODES FOR PRACTICAL DATA OF SECOND WINDOW

Methods	Attenuation	Freq	(%)Damping	(%)TVE
RD-ESPRIT	−0.6324	0.2545	36.7760	30.2842
K-SVD	−0.2630	0.2875	14.4120	10.5947
XgBoost	−0.1347	0.2932	7.2953	7.8581

PMU measurements validated from section IV. Hence, it is reasonable to draw the conclusion that the suggested XgBoost scheme which preserves the signal characteristics is able to give precise assesment of modes, thus justifying its use for WAMS.

REFERENCES

[1] Kundur P 1994 Power system stability and control. New York, McGrawHill

[2] J. Ma, P. Zhang, H. -j. Fu, B. Bo and Z. -y. Dong, "Application of Phasor Measurement Unit on Locating Disturbance Source for Low-Frequency Oscillation," in IEEE Transactions on Smart Grid, vol. 1, no. 3, pp. 340-346, Dec. 2010, doi: 10.1109/TSG.2010.2071889.

[3] M. Sahoo and S. Rai, "An efficient K-SVD based Algorithm for detection of Oscillatory mode from ambient data for synchrophasor application," 2021 IEEE 18th India Council International Conference (INDICON), 2021, pp. 1-5, doi: 10.1109/INDICON52576.2021.9691744.

[4] S Rai, D Lalani, S K Nayak, T Jacob and P Tripathy, "Estimation of low-frequency modes in power system using robust modified Prony". IET Gener. Transm. Distrib. 2016, 10:1401–1409

[5] S. Rai, P. Tripathy, and S. Nayak, "A robust TLS-ESPIRIT method using covariance approach for identification of low-frequency oscillatory mode in power systems," in Power Systems Conference (NPSC), 2014 Eighteenth National, Dec 2014, pp. 1–6.

[6] P. Korba, M. Larsson, and C. Rehtanz, "Detection of oscillations in power systems using Kalman filtering techniques," in Proc. 2003 IEEE Conf. Control Applications, 2003., vol. 1, pp. 183–188.

[7] P Zhang, X Wang and J S Thorp "Synchronized measurement based estimation of inter-area electromechanical modes using the Ibrahim

[8] S. Rai, P. Tripathy, and S. K. Nayak, "An efficient wavelet based technique for oscillatory mode identification of ambient data via RD and TLS-ESPRIT," Natl. Power Syst. Conf. NPSC 2016, 2017

[9] S. Rai, P. Tripathy, and S. K. Nayak, "Using sparsity to estimate oscillatory mode from ambient data," Sadhana - Acad. Proc. Eng. Sci., vol. 44, no. 4, 2019, pp. 1–9.

[10] R. Ghorbani and R. Ghousi, "Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques," in IEEE Access, vol. 8, pp. 67899-67911, 2020, doi: 10.1109/ACCESS.2020.2986809.

[11] L. Zhang, Y. Ji, T. Liu and J. Li, "PM2.5 Prediction Based on XgBoost," 2020 7th International Conference on Information Science and Control Engineering (ICISCE), 2020, pp. 1011-1014, doi: 10.1109/ICISCE50968.2020.00207.

[12] Tao, Hai & Salih, Sinan & Oudah, Atheer & Abba, Sani & Ameen, Ameen & Awadh, Salih & A. Alawi, Omer & Mostafa, Reham & Udayar Pillai, Surendran & Yaseen, Zaher. (2022). Development of new computational machine learning models for longitudinal dispersion coefficient determination: case study of natural streams, United States. Environmental Science and Pollution Research. 29. 10.1007/s11356-022-18554-y.

[13] IEEE Std C37.118.1aTM 2014 IEEE Standard for synchrophasor measurements for power systems

[14] J Thambirajah, N F Thornhill and B C Pal " A multivariate approach towards inter-area oscillation damping estimation under ambient conditions via independent component analysis and random decrement" IEEE Trans. Power Syst. 2011, no.26, pp. 315–322

[15] PDCI 2005 PDCI probe testing plan. Available online at http://www.transmission.bpa.gov/business/operations/ SystemNews/

[16] WECC 2005 Report and data of WECC. Available online at ftp://ftp.bpa.gov/pub/WAMSInformation/