Sādhanā (2020)45:265
https://doi.org/10.1007/s12046-020-01502-2

A robust WKNN-TLS-ESPRIT algorithm for identification of electromechanical oscillation modes utilizing WAMS

SHEKHA RAI

Department of Electrical Engineering, National Institute of Technology Rourkela, Rourkela, Orrisa, India
e-mail: rai.shekha@gmail.com; rais@nitrkl.ac.in

MS received 11 May 2020; revised 12 August 2020; accepted 16 September 2020

Abstract. This paper proposes a robust WKNN-TLS-ESPRIT algorithm that takes into account the effect of the unavailability of phasor measurement unit (PMU) data for identifying the low-frequency oscillatory modes in power systems. The main contribution of the proposed work is to create an enhanced autocorrelation matrix using a weighted *K* nearest neighbours (WKNN)-based predictive model to deal with such issues. In the present work, a Bayesian approach is utilized to determine the empirical number of neighbourhood parameters. The improved autocorrelation matrix is then used by total least square estimation of signal parameters via rotational invariance technique (TLS-ESPRIT) algorithm to provide a robust estimate of the modes. Robustness of the proposed method over the other methods is validated with a simulated test signal with missing data through Monte Carlo simulations at different SNRs. The effectiveness of the proposed approach is further verified on real data derived from PMU located in Western Electricity Coordinating Council grid.

Keywords. Modal analysis; missing data; WAMS; *K* nearest neighbours (KNN).

1. Introduction

The increasing capacity of generation through renewable sources of energy has caused uncertainty in generation and subsequently, its integration into the electric grid has posed severe operational challenges concerned with stability of large-scale power systems. Maintaining adequate damping for small-signal oscillatory modes is highly essential. Conventional techniques for identifying the dominant modes use an off-line approach that fails to observe the real-time system dynamics. In the recent times, with the advancement of wide area monitoring systems (WAMSs) that employ phasor measurement units (PMUs) and Global Positioning System (GPS) to provide highly accurate time-stamped phasors [1], the viability for on-line extraction of modal information has increased. Numerous measurement-based online estimation algorithms have been proposed like Kalman filter [2], sparsity [3], variable projection [4], Prony [5] and ESPRIT [6]. The Kalman filter uses an iterative approach and hence, has the disadvantage of being numerically unstable. A variable projection algorithm involving orthogonal projection to extract the modal parameters from the signal space is cited in [4]. Some identification methods use ESPRIT [6] to create an autocorrelation matrix from the observed data. Methods based on ESPRIT show higher noise immunity than Prony.

PMUs may have incomplete measurements due to high congestion in communication network, malfunctioning of PMUs or PDCs, malicious attack, etc. [7]. Such loss of the data makes the system unobservable and causes considerable degradation in the performance of these estimators, thereby causing a threat to the sustainability and stability of the system. Hence, design of a robust on-line mode identification scheme that can handle incomplete PMU data effectively has become highly vital. One approach is to ignore the missing values, but it may lead to loss of vital information. To overcome this problem, imputation of missing data with a value resembling the original value of the sample can be explored. Some classical techniques like mean, median and mode have been used for missing data imputation [8] but are not very effective. Recent data recovery works have suggested the use of a low-dimensional matrix [12, 13], but require appropriate approximation of the low rank. Hence this paper motivates towards the improvement of the existing modified TLS-ESPRIT [6] algorithm using an instance-based KNN learning approach to form an improved autocorrelation matrix, to mitigate the effect of missing samples in the PMU data. Section 2 introduces the methodology involved in the proposed approach. Simulation results are discussed in Section 3 and lastly conclusion is provided in Section 4.

Published online: 22 October 2020

2. Methodology of the proposed WKNN-TLS-ESPRIT scheme

2.1 *WKNN*

K nearest neighbours (KNN) algorithm [8] is a widely used non-parametric machine learning technique used for data classification in pattern recognition. The proposed work has utilized a weighted KNN (WKNN) approach to impute the missing data from PMU. The main steps involved in the proposed approach are (a) selection of *k* observations and (b) estimation of missing values, which are described here.

2.1.1 *Selection of k observations* Proper selection of optimal *k* observations is required for efficient operation of the KNN algorithm. In the proposed work, a Bayesian approach [10] is explored. The following terms are used:

$J \rightarrow$ number of partitions of the sample space
$f_j \rightarrow$ density function
$\pi_j \rightarrow$ prior probabilities of J
$\mathbf{p} \rightarrow$ vector of conditional probabilities
$\zeta_k(\mathbf{p}) \rightarrow$ prior distribution of $\mathbf{p}$ in neighbourhood around $\mathbf{x}$
$q_k \rightarrow$ number of variables in k neighbours from the j^{th} partition

The conditional distribution of $\mathbf{p}$ for fixed k and q_k is represented as

$$\psi(\mathbf{p}|k, q_k) = \xi_k(\mathbf{p})\zeta(q_k)|(\mathbf{p}, k) / \int \xi_k(\mathbf{p})\zeta(q_k)|(\mathbf{p}, k)d\mathbf{p} \tag{1}$$

where, for given $\mathbf{p}$ and k, $\zeta(q_k)|(\mathbf{p}, k)$ is the conditional distribution of q_k. Equation (1) is utilized to formulate a Bayesian strength function that is expressed as

$$S(j|k) = \int_{p_j = \max p_1, p_2, \ldots, p_J} \psi(\mathbf{p}|k, q_k)d\mathbf{p}. \tag{2}$$

If missing observations belong to a particular partition *j*, then the value of strength function must be very high. Hence, to determine the appropriate value of *k*, an accuracy index is defined as [10]

$$\gamma(k) = \sum_{j=1}^{J} \pi_j \int S(j|k) f_j(\mathbf{x})d\mathbf{x}. \tag{3}$$

The value of k for which γ is maximum is considered as the optimum value.

2.1.2 *Estimation of missing values* This is done through the columns of *K* nearest samples, which are chosen on the basis of an appropriate distance measure [9]. The steps of the algorithm [11] are as follows:

1. the data set is divided into various parts, $\mathbf{D} = \{D_1, D_2, \ldots\ldots, D_k\}$;
2. for each $x_o, x_i \in D_k$, where x_o represents incomplete case, Euclidean distance measure is used and calculated as shown in equation (4):

$$Dist(x_o, x_i) = \sqrt{\sum_{k=1}^{m} (X_{0K} - X_{ik})^2} \tag{4}$$

where $i = 1, 2, \ldots, n$; let $g_1 = DIST(x_o, x_i)$;
3. g_1 is arranged in ascending order of distance;
4. k^{th} g_1 cases are chosen; $N_k(x) = \{x_1, x_2, x_3, \ldots, x_k\}$;
5. for each neighbour $x_k \in N_k(x)$, where x_k represents complete case, $\text{DIST}(x_k, x_i)$ is computed as given in (4) and let $g_2 = DIST(x_k, x_i)$;
6. g_2 is sorted in increasing order of distance;
7. k^{th} g_2 cases are considered; $N_k(x_k) = \{x'_1, x'_2, x'_3, \ldots, x'_k\}$;
8. if $x_k \in N_k(x)$ and $x_0 \in N_k(x_k)$, $M_k(x)$ is determined by equation (5):

$$M_k(x) = x_k \in D | x_k \in N_k(x) \cap x_o \in N_k(x_k); \tag{5}$$

9. end for each;
10. x_0 is replaced by the weighted average of the most frequent values of each column in $M_k(x)$, where weight w_{ik} is

$$w_{ik} = \frac{1/dist_{ik}}{\sum_{k=1}^{K} 1/dist_{ik}}; \tag{6}$$

11. end for each;
12. final output is the complete set $\hat{\mathbf{D}} = \{\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_k\}$.

2.2 *TLS-ESPRIT*

The complete data set derived from the WKNN algorithm is used by the TLS-ESPRIT [6] for forming an robust autocorrelation matrix to deal with incomplete measurements of the PMU. The real-time signal can be represented by the mathematical model as

$$y(n) = s(n) + z(n) = \sum_{j=1}^{P} \alpha_j e^{\beta_j n} + z(n) \tag{7}$$

where α_j is the complex signal amplitude and $\beta_j = d_j \pm i\omega_j$ where d_j and ω_j are the modal parameters.

Considering a data vector of length B shown as follows:

$$\begin{bmatrix} y(n) \\ y(n+1) \\ . \\ . \\ y(n+B-1) \end{bmatrix} = \sum_{k=1}^{P} \begin{bmatrix} s_k(n) \\ s_k(n+1) \\ . \\ . \\ s_k(n+B-1) \end{bmatrix} + \begin{bmatrix} z(n) \\ z(n+1) \\ . \\ . \\ z(n+B-1) \end{bmatrix} \tag{8}$$

where the $\mathbf{s}_k(n)$ is represented as

$$\mathbf{s}_k(n) = \begin{bmatrix} s_k(n) \\ s_k(n+1) \\ . \\ . \\ s_k(n+B-1) \end{bmatrix} = \alpha_k e^{n\beta_k} \begin{bmatrix} 1 \\ e^{\beta_k} \\ . \\ . \\ e^{(B-1)\beta_k} \end{bmatrix}. \tag{9}$$

$$\therefore \mathbf{y}(n) = \mathbf{AS} + \mathbf{z}(n) \tag{10}$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & . \; . \; . & 1 \\ e^{\beta_1} & e^{\beta_2} & . \; . \; . & e^{\beta_P} \\ . & . & . \; . \; . & . \\ . & . & . \; . \; . & . \\ e^{(B-1)\beta_1} & e^{(B-1)\beta_2} & . \; . \; . & e^{(B-1)\beta_P} \end{bmatrix} \tag{11}$$

$$= [\mathbf{a}(\beta_1), \mathbf{a}(\beta_2), \ldots \mathbf{a}(\beta_P)].$$

$\mathbf{S} = [\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_P]^T$ and $\hat{a}_k = \alpha_k e^{n\beta_k}$.

The correlation matrix $\hat{\mathbf{R}}_{yy}$ estimated from the complete data set $\hat{\mathbf{D}}$ possesses rotational invariance property, which is applied by ESPRIT algorithm to estimate the modes. Here, the time shift has been modelled as a phase shift [6].

For implementing TLS-ESPRIT, the covariance matrix is separated into two orthogonal bases. Singular value decomposition (SVD) technique is utilized for this purpose. The eigenvectors associated with the dominant P eigenvalues constitute the signal subspace. First and the second shift invariance signal properties [6] are utilized to derive the estimated attenuation factor and frequency. A block diagram as shown in figure 1 narrates the several steps required for online detection of low-frequency power system mode.

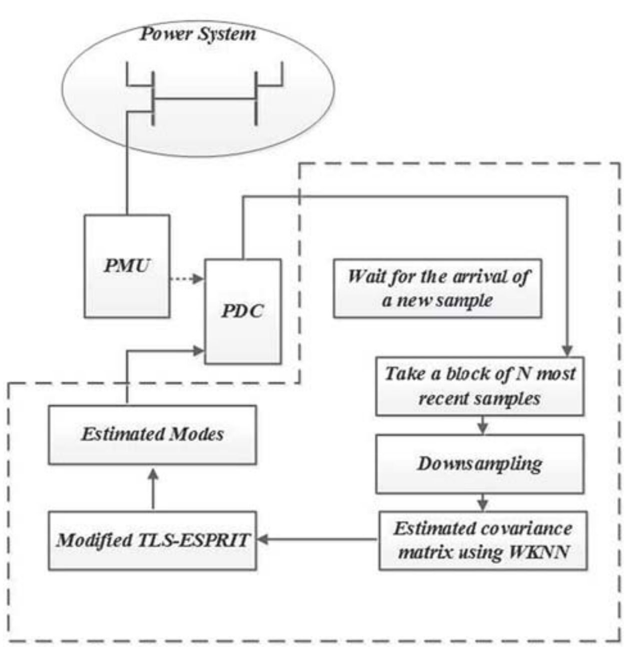

Figure 1. Block diagram of the proposed WKNN-TLS-ESPRIT scheme.

3. Results and discussion

The efficacy of the proposed WKNN-TLS-ESPRIT estimator has been elucidated using a synthetic signal resembling inter-area mode. Performance evaluation of the proposed scheme with the improved Prony, variable projection, modified TLS-ESPRIT and low rank Hankel matrix [13] method is done through Monte Carlo simulations of 10000 runs considering incomplete measurements and various noise levels. The efficiency of the proposed approach over the other mentioned methods is also assessed for real data of the WECC network.

3.1 *Mode estimation for inter-area mode*

A test data with attenuation factor = −0.0.07, frequency = 0.4 Hz corrupted by incomplete measurements is utilized for the simulation. Distribution of estimated attenuation factor and frequency for this test signal at SNR = 40 dB is shown in figure 2(a) and 2(b). The mean and variance of the evaluated modes by various estimators are listed in Table 1. It is observed from Table 1 that the proposed WKNN-TLS-ESPRIT approach excels in providing the damping estimates as compared with the other mentioned methods. As for the frequency estimated (0.3996 Hz), the proposed method outperforms the other mentioned techniques.

3.2 *Estimation of modes utilizing a practical probing data from the WECC*

Performance analysis of the proposed WKNN-TLS-ESPRIT scheme with the other schemes has been executed on a probing data [14] extracted through PMU connected in the WECC network as depicted in figure 3. The mode estimated was cited in [15] as 0.318 Hz with 8.3% damping for the North–South Swing. Each analysis window of real probing data used for the purpose of mode estimation is artificially corrupted by missing measurements. The proposed WKNN-TLS-ESPRIT approach outshines in providing reliable estimate of the damping and frequency

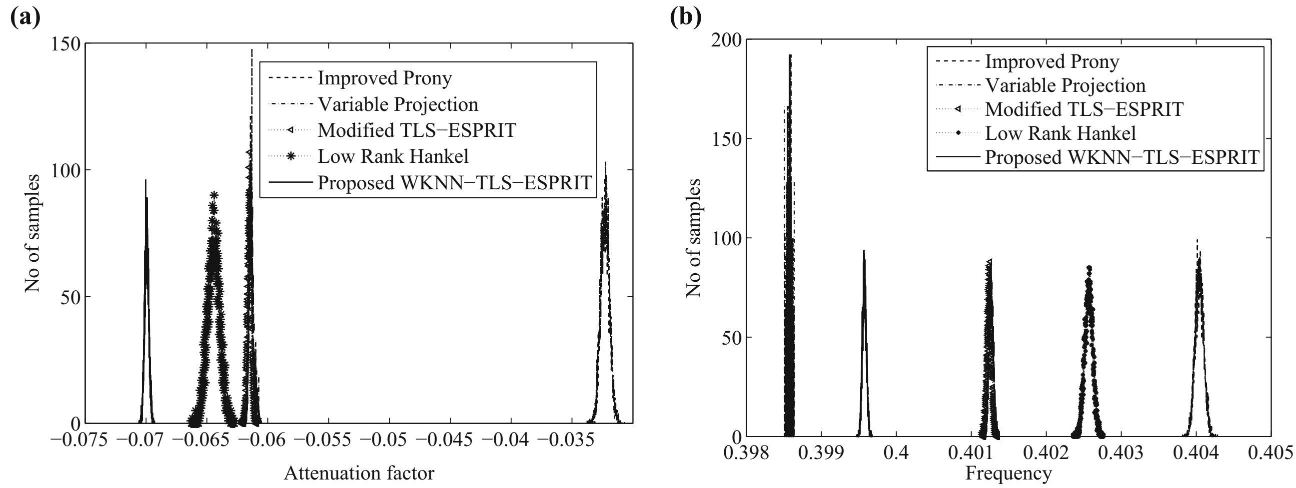

Figure 2. (**a**) Distribution of estimated attenuation factor and (**b**) distribution of estimated frequency.

Table 1. Mean and variance of the estimated mode for improved Prony, variable projection, modified TLS-ESPRIT and the proposed WKNN-TLS-ESPRIT method with incomplete data.

SNR (dB)	Attenuation factor		Frequency (Hz)	
	Mean	Variance	Mean	Variance
Improved Prony				
40	− 0.0323	1.2754×10^{-7}	0.4040	3.2705×10^{-9}
30	− 0.0322	1.2952×10^{-6}	0.4040	3.2591×10^{-8}
20	− 0.0311	1.1428×10^{-5}	0.4040	2.8650×10^{-7}
Variable projection				
40	− 0.0613	4.2706×10^{-8}	0.3986	8.3081×10^{-10}
30	− 0.0613	4.4547×10^{-7}	0.3981	7.1207×10^{-4}
20	− 0.0613	1.6011×10^{-6}	0.4023	3.0001×10^{-4}
Modified TLS-ESPRIT				
40	− 0.0615	2.6532×10^{-8}	0.4012	8.8669×10^{-10}
30	–0.0615	2.6321×10^{-7}	0.4012	8.7613×10^{-9}
20	− 0.0615	2.6548×10^{-6}	0.4012	8.8903×10^{-8}
Low rank Hankel				
40	− 0.0644	3.0373×10^{-9}	0.4026	3.0373×10^{-9}
30	− 0.0644	3.0578×10^{-8}	0.4026	3.0578×10^{-8}
20	− 0.0646	3.0334×10^{-7}	0.4026	3.0334×10^{-7}
Proposed WKNN-TLS-ESPRIT				
40	− 0.0699	2.9713×10^{-8}	0.3996	7.2943×10^{-10}
30	− 0.0699	2.9226×10^{-7}	0.3996	7.3953×10^{-9}
20	− 0.0703	2.8963×10^{-6}	0.3994	7.5621×10^{-8}

(0.3232 Hz at 7.9371%) as compared with Prony, variable projection, ESPRIT and low rank Hankel estimator in case of analysis window 1 as validated in Table 2. The proposed approach provides damping estimate (8.4109% damping) that is approximately close to the reported value [15], in contrast with the other estimators for window 2. For the third window corresponding to ambient data, the damping estimate by the proposed method is better as compared with the other methods.

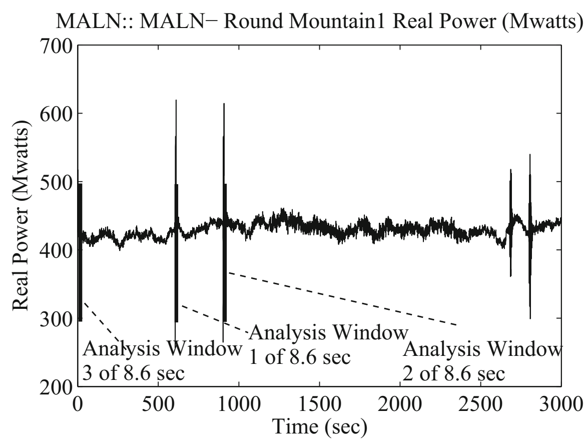

Figure 3. Probing data corresponding to flow of power.

4. Conclusion

This paper explores a WKNN-based data mining approach to build a robust autocorrelation matrix to handle missing PMU measurements. The proposed scheme selects the significant number of nearest neighbour to represent the incomplete measurements and hence designs an improved

Table 2. Estimated modes considering missing measurements in the signal.

Estimated mode for window 1	Attenuation	Freq. (Hz)	Damp. (%)
Improved Prony	– 0.1072	0.3313	5.1414
Variable projection	– 0.1017	0.3533	4.5764
Modified TLS-ESPRIT	– 0.1177	0.3233	5.7865
Low rank Hankel	– 0.1830	0.3246	8.9387
Proposed method	– 0.1617	0.3232	7.9371
Estimated mode for window 1	Attenuation	Freq. (Hz)	Damp. (%)
Improved Prony	– 0.1033	0.3204	5.1248
Variable projection	– 0.1299	0.3315	6.2262
Modified TLS-ESPRIT	– 0.1238	0.3174	6.1975
Low rank Hankel	– 0.1889	0.3115	9.6088
Proposed method	– 0.1665	0.3139	8.4109
Estimated mode for window 3	Attenuation	Freq. (Hz)	Damp. (%)
Improved Prony	0.03730	0.2917	–2.0329
Variable projection	–0.0460	0.2805	2.6117
Modified TLS-ESPRIT	–0.0496	0.2938	2.6837
Low rank Hankel	–0.0569	0.2905	3.1154
Proposed method	– 0.0601	0.2911	3.2853

autocorrelation matrix, followed by a modified TLS-ESPRIT that works fairly well for low SNR to provide a highly accurate estimate of the modes. The effectiveness of the proposed WKNN-TLS-ESPRIT scheme is demonstrated on a test signal and real probing data of the WECC system considering incomplete measurements. From the simulation results, it can be revealed that the WKNN-based enhanced autocorrelation matrix that preserves the signal characteristics is robust towards unavailable PMU measurements and therefore can be applied in wide-area monitoring systems.

Acknowledgements

This work was supported in part by a grant from World Bank for NPIU, India (TEQIP-III/CRS/1-5728870640).

List of symbols

J	Number of partitions of the sample space
f_j	Density function
π_j	Prior probabilities of J
$\mathbf{p}$	Vector of conditional probabilities
$\zeta_k(\mathbf{p})$	Prior distribution of $\mathbf{p}$ in a neighbourhood around $\mathbf{x}$
q_k	Number of variables in k neighbours from the j^{th} partition
$S(j\|k)$	Bayesian strength function
$\gamma(k)$	Accuracy index
k	Number of nearest neighbours
$\mathbf{D}$	Data set
x_o	Incomplete case
$Dist$	Euclidean distance measure
w_{ik}	Weight assigned to k^{th} neighbour
$y(n)$	Real-time signal
$s(n)$	Clean signal
$z(n)$	White Gaussian noise
α_j	Complex signal amplitude
β_j	Modal parameters
d_j	Attenuation factor ω_j Frequency
B	Length of data vector
$\hat{\mathbf{R}}_{yy}$	Estimated correlation matrix
$\hat{\mathbf{D}}$	Estimated data set
P	Number of dominant eigenvalues

References

[1] Phadke A G and Thorp J S 2008 *Synchronized phasor measurements and their applications*. New York: Springer

[2] Korba P, Larsson M and Rehtanz C 2003 Detection of oscillations in power systems using Kalman filtering techniques. In: *Proceedings of the IEEE Control Applications*, Istanbul, Turkey, pp. 183–188

[3] Rai S, Tripathy P and Nayak S K 2019 Using sparsity to estimate oscillatory mode from ambient data. *Sadhana* 44: 01–09

[4] Borden A R and Lesieutre B C 2014 Variable projection method for power system Modal Identification. *IEEE Trans. Power Syst*. 29: 2613–2620

[5] Rai S, Lalani D, Nayak S K, Jacob T and Tripathy P 2016 Estimation of low-frequency modes in power system using

robust modified Prony. *IET Gener. Transm. Distrib*. 10: 1401–1409

[6] Rai S, Tripathy P and Nayak S K 2014 A robust TLS-ESPIRIT method using covariance approach for identification of low-frequency oscillatory mode in power systems. In: *Proceedings of the IEEE National Power System Conference (NPSC 2014)*, Guwahati, India, pp. 1–6

[7] He M, Vittal V and Zhang J 2013 Online dynamic security assessment with missing pmu measurements: a data mining approach. *IEEE Trans. Power Syst*. 28: 1969–1977

[8] Malarvizhi R and Thanamani A 2012 *K*-nearest neighbor in missing data imputation. *Int. J. Eng. Res. Dev*. 5: 105–107

[9] Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D and Altman R B 2001 Missing value estimation methods for DNA microarrays. *Bioinformatics* 17: 520–525

[10] Anil K G 2006 On optimum choice of *k* in nearest neighbor classification. *Comp. Stat. Data Anal.* 50: 3113–3123

[11] Sun C M, Yao C, Shen L and Yu X 2016 Improving classification accuracy using missing data filling algorithms for the criminal dataset. *Int. J. Hybrid Inf. Technol.* 9: 367–374

[12] Gao P, Wang M, Ghiocel S G, Chow J H, Fardanesh B and Stefopoulos G 2016 Missing data recovery by exploiting low-dimensionality in power system synchrophasor measurements. *IEEE Trans. Power Syst*. 31: 1006–1013

[13] Zhang S, Hao Y, Wang M and Chow J H 2017 Multi-channel missing data recovery by exploiting the low-rank hankel structures. In: *Proceedings of the 7th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP),* Curacao, pp. 1–5

[14] PDCI Probe Testing Plan 2005 Available online at http://www.transmission.bpa.gov/business/operations/SystemNews/

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