Load Estimation of Distribution Networks Using Inverse PCA

A.A. Yaghooti [i];M. Parsa Moghaddam [ii]; M.R. Haghifam [iii]; V. Johari Majd [iv]

ABSTRACT

In this paper, an efficient method is proposed for load estimation of distribution networks with limited real-time data based on principal component analysis. The principal components of load variables are first detected off-line from historical data. Next, the interrelations between the load variables are developed using a new concept that we call Inverse PCA (IPCA) method; this leads to some proper models devised so that the distribution network becomes observable. Finally, through incorporation of a certain finite set of real-time data measurements, the loads of all network nodes in real-time are estimated with a desired degree of accuracy via the IPCA method. A case study on a real network is considered in the paper to highlight better the merit of the proposed method. The experimental results easily show that the proposed method outperforms the previously cited techniques.

KEYWORDS

Data mining, principal component analysis, load estimation, state estimation, distribution network

1. INTRODUCTION

The problem of load determination in different nodes of distribution networks can be considered as a special case of state estimation [1]. By state estimation, the operating point of the system can be determined in order to shift the network to an optimum operating condition. Due to daily expansion of distribution networks along with power system restructuring, which results in further sophistication of the system, optimum planning and operation of distribution networks has become a challenging issue. In addition, incorporation of new technologies such as distributed generation (DG) and sensitivity of end-use to power quality has made the problem of load estimation and network monitoring very crucial.

Due to the vast number of network nodes, the number of variables in the distribution network increases remarkably. Therefore, a large number of real-time data is needed for a precise state and/or load estimation of the network. However, due to technical and economical restrictions, the availability of such a large number of real-time data is almost impossible. On the other hand, incorporation of new technologies in distribution networks such as network automation, automatic meter reading (AMR), distribution network mechanization, and distribution management system (DMS), has facilitated the application of advanced sciences and technologies to find optimal operation solution for the distribution networks. The above facilities have also provided a suitable environment for adoption and realization of more precise load estimation techniques in distribution networks.

In most of the methods developed so far, not enough real-time data is used for the solution of load and/or state estimation. These methods range from the incorporation of pseudo data to heuristic methods including hybrid particle swarm optimization, or simplification of network topology and simultaneous use of weighted least square (WLS) approach, briefly discussed below.

The approaches presented in [2],[3], use customer's monthly peak load and transformers peak load data for the extraction of load curve. This curve is then used as pseudo data for load estimation. The above methods are suitable for peak load estimation and are not useful for load

i. A. A. Yaghooti is a Ph.D. student in E.E. Department of Engineering, Tarbiat Modares University, Tehran, Iran (e-mail: a-yaghoti@Trec.co.ir).
ii . M. P. Moghaddam is an Associate Professor in E.E Department of Engineering, Tarbiat Modares University, Tehran, Iran (e-mail: parsa@modares.ac.ir). Corresponding author.
iii .M. R. Haghifam is a Professor in E.E Department of Engineering, Tarbiat Modares University, Tehran, Iran (e-mail: haghifam@modares.ac.ir).
iv . Vahid Johari Majd is an Associate Professor in E.E. Department of Engineering, Tarbiat Modares University, Tehran, Iran (e-mail: majd@modares.ac.ir).

estimation.

A method for load curve determination using pseudo real-time data, such as AMR data, is addressed in [4]. The resulting load curve is then used for state estimation. By merely using these data, only some network variables can be estimated; however, due to the AMR restrictions, determination of all of the system variables is not possible.

In [5] and [6], by assuming the availability of pseudo data, all nodes' loads are first determined, and then, the states of the network are estimated. The approach uses both the historical data of energy billing and historical load curve data of different load points in the network. In [7], a probabilistic model for estimating the loads of different points (nodes) of the network is developed based on the historical data resulting in better performance compared with the previously developed methods. A fuzzy logic and neural network based approach for active load estimation of distribution network is also proposed in [8], where, the inputs of the neural network are taken as: classified load curves, monthly energy consumption together with real-time measurements.

In some other studies, the network topology is simplified according to the position of the metering equipments. Then, a sensitivity analysis is employed on historical load data to identify the relationship between the load changes in different sections of the network. These results are then compared with the historical data of the measurement equipments in each section; accordingly, the load estimation problem is then solved using a limited set of real-time data [9],[10]. The method can be improved by establishing relations between the load changes of different points and the real-time measurement equipment in the network, and employing data analysis techniques.

Some researchers have developed state estimation methods based on WLS or heuristic optimization techniques such as hybrid particle swarm optimization (HPSO) assuming that good enough data is available [11],[12].

The problem of optimum allocation of measurement equipments in distribution networks, which has a key role in load estimation problem, has also been addressed by some researches [13],[14]. The approaches proposed in these references are mainly based on determining some correlation between the load changes in different nodes of the network using historical data.

Incorporation of data mining based techniques [15],[16] for the classification of the collected data and exploration of their hidden interrelations are very useful for improvement of load estimation in distribution networks with limited data. The Principal Component Analysis (PCA) method is one of the widely-used data analysis techniques in data mining studies. This method explores the relations and extracts features of different variables in huge sets of data in order to reduce the dimensionality of the problem. Identification of governing patterns between the variables results in more precise solutions while less data have been used [17],[19]. Comparison of six different short-term load forecasting (STLF) methods in [20] shows that using PCA can considerably improve the performance of these methods.

In this paper, we use PCA for load estimation of distribution networks with limited real-time data. The governing patterns between the load variables of the network are first detected off-line from historical data. Next, the interrelations between the load variables are developed using a new concept that we call Inverse PCA (IPCA) method; this leads to some proper models devised so that the load distribution network becomes observable. Finally, through incorporation of a certain finite set of real-time data measurements, the loads of all network nodes in real-time are estimated with a desired degree of accuracy via the IPCA method.

We verify our method in a real network, which is a part of Tehran's distribution network and compare the results with those of other existing approaches such as those that employ fuzzy logic, artificial neural network and least square based regression.

The rest of the paper is as follows. In section 2, a detailed description of proposed method is discussed. In section 3, the method is applied in a case study and is compared with other methods. Section 4 provides a conclusion to the paper and discusses topics for further research.

2. Proposed Method for Load Estimation

This section fully describes the proposed approach for load estimation of distribution networks using a limited set of available real-time data. Fig.1 shows a block diagram representation of the method. Initially, the conventional PCA method is applied to the past load data of all network nodes in order to extract the governing patterns among the loads. Next, based on the past load data IPCA method is used to establish the relation models between the known and unknown load variables; this leads to some proper models devised so that the load distribution network becomes observable. Finally, with IPCA method the loads of all network nodes are estimated in real time with a desired accuracy using: real-time data measurements and the principal components extracted earlier.

In short, the proposed method is aimed to determine a robust linear relationship between the known and unknown load variables, in order to relax the need for availability of a large set of real-time data.

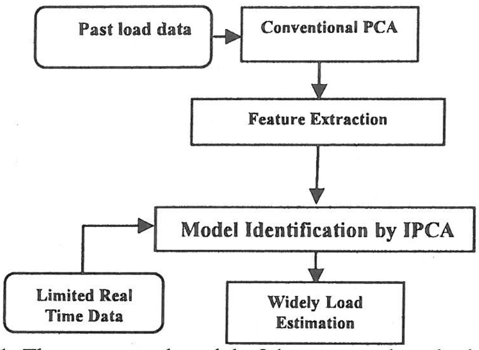

Fig.1: The conceptual model of the proposed method.

2.1 Pattern extraction of load variables

Assuming the availability of historical load data, the network data matrix "H" can be defined as:

$$H = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,n} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,n} \\ \vdots & \ddots & \ddots & \ddots \\ h_{r,1} & h_{r,2} & \cdots & h_{r,n} \end{bmatrix} \quad (1)$$

where,

n: number of load nodes

r: number of load samples in each node

$h_{i,j}$: i-th sample of the j-th node load

The so-called adjusted load matrix, H_A is then calculated as,

$$\bar{h}_j = \frac{1}{r}\sum_{i=1}^{r} h_{i,j} \quad \& \; j \in \{1,2,...,n\}$$
$$Row_i(H_A) = Row_i(H) - \bar{h} \quad \& \; i \in \{1,2,...,r\} \quad (2)$$

In which, $\bar{h}$ denotes the average value of samples of nodal loads, and $Row_i(H)$ represents the i-th row of matrix "H". Now, the covariance matrix "C" is calculated by,

$$C = \frac{1}{r}(H_A^T \cdot H_A) \quad (3)$$

It is a fact that the above matrix is positive semi-definite with "n" unit eigenvectors which are corresponding to "n" non-negative real eigenvalues [21]. By selecting a set of larger eigenvalues, we can compose a matrix denoted by "T" as given below in order to build the principal components of matrix "C".

$$T = [\phi_1 \quad \phi_2 \cdots \quad \phi_i \quad \cdots \quad \phi_k] \quad (4)$$

In the above, ϕ_i is the i-th eigenvector corresponding to the i-th eigenvalue and k is the number of selected eigenvalues which satisfies $k << n$. The matrix "T", which is called as "PCA transformation matrix", defines a strong linear relation between the variables of data matrix.

Now, by using matrix "T", we can transform the adjusted data matrix "H_A", to a reduced dimension matrix of PCA as follows:

$$H_{PCA} = H_A \cdot T \quad (5)$$

where, (H_{PCA}) denotes the projection of "H_A" to PCA.

It should be noticed that a suitable selection of the number of eigenvectors "k", will have a key role in performance of the proposed method. This issue will be discussed later in section 3.

2.2 The proposed IPCA method

In the previous section, by using the conventional PCA method, the governing patterns of data were extracted and then used in transformation matrix "T". In order to estimate the network load in the real-time of t_0, it is necessary to estimate accurately n load variables corresponding to n buses. To do so, we first represent the load variables of the network as:

$$Q(t_0) = [Z(t_0) \quad X(t_0)]$$
$$Z(t_0) = [z_1(t_0) \quad z_2(t_0) \quad \cdots \quad z_m(t_0)] \quad (6)$$
$$X(t_0) = [x_1(t_0) \quad x_2(t_0) \quad \cdots \quad x_{n-m}(t_0)]$$

where:

$Q(t_0)$: vector of nodal loads in t_0

$Z(t_0)$: vector of known loads in t_0

$X(t_0)$: vector of unknown loads in t_0

$z_i(t_0)$: i-th node known load in t_0

$x_j(t_0)$: j-th node unknown load in t_0

m : number of load measurements

The adjusted value of vector $Q(t_0)$ is calculated by Eq.7 and then it is transformed by matrix "T" as given in Eq.8.

$$A(t_0) = Q(t_0) - \bar{h} \quad (7)$$
$$A_{PCA}(t_0) = A(t_0).T \quad (8)$$

Here, $A_{PCA}(t_0)$ whose i-th element is calculated by Eq.9 represents the projection of $A(t_0)$, which denotes the vector of load variables.

$$A_{PCA}(t_0)_{(i)} = \sum_{j=1}^{m} z_j . T_{(1,j)} + \sum_{j=1}^{n-m} x_j . T_{(1,j+m)} \quad (9)$$

Since the ϕ_i vectors, which construct the matrix "T", are orthonormal, it can be shown that the pseudo inverse of T is equal to its transpose [21]. Consequently, the return primary vector of load variables Q', A' can be determined as follows:

$$Q'(t_0) = A_{PCA}(t_0).T^T + \bar{h}$$
$$Q'_j(t_0) = (\sum_{i=1}^{k} A_{PCA}(k).T_{(i,j)}) + \bar{h}_j \quad (10)$$

$$A'_j(t_0) = \sum_{i=1}^{k}(\sum_{j=1}^{m} z_j.T_{(1,j)} + \sum_{j=1}^{n-m} \hat{x}_j.T_{(1,j+m)}).T_{(i,j)} \tag{11}$$

It should be noted that the uniqueness of solution given by Eq.11 is guaranteed. Furthermore, each element of vector $Q'(t_0)$ is obviously a linear combination of unknown variables, which are related to the hidden relations among the variables obtained by the PCA transformation, see Eq.10. It is expected that the return Q, $Q'(t_0)$ should be deviated to form $Q(t_0)$. This deviation can be ignored by a proper choice of "k" leading to,

$$Q_j(t_0) = \sum_{i=1}^{k}(\sum_{j=1}^{m} m_j.T_{(1,j)} + \sum_{j=1}^{n-m} x_j.T_{(1,j+m)}).T_{(i,j)} + \bar{h}_j \tag{12}$$

By rewriting the above Eq.12 for "n" nodes of the network, we will end up with the following.

$$\begin{pmatrix} f_1(Z) \\ f_2(Z) \\ \vdots \\ f_n(Z) \end{pmatrix} = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1,n-m} \\ u_{21} & u_{22} & \cdots & u_{2,n-m} \\ \vdots & \vdots & \cdots & \vdots \\ u_{n1} & u_{n2} & \cdots & u_{n,n-m} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-m} \end{pmatrix} \tag{13}$$

$$\Rightarrow f(Z) = h(X)$$

$$f_i(Z) = \sum_{j=1}^{m} \alpha_{i,j} \cdot Z_j$$

where,

$f(Z)$: measurement model

$h(X)$: measurement function

$\alpha_{i,j}$: j-th coefficient of i-th measurement model

Therefore, the inter-relation between known and unknown variables can be properly explored via IPCA method. Then, the estimated load values of the unknown loads can be determined by:

$$X(t_0) = U^{+} \cdot f(Z) \tag{14}$$

Here, U^{+} denotes the pseudo inverse of matrix U.

Fig. 2 represents the flowchart of the proposed. As it can be seen, the estimated values of second stage are used as new data in the first stage, which improves the performance of the algorithm.

2.3 Proposed method for the "k" selection

As it was discussed earlier, selection of the number of eigenvectors "k" can have a great effect on the accuracy of the proposed load estimation method. In conventional *PCA* method, the transformation matrix "T" can be arranged by minimum selected number of eigenvectors to obtain a suitable projection of the primary data. This projection can be used for returning the primary data with a minimum error. In conventional *PCA* method, it is recommended that the number of eigenvectors, "k", be selected according to the following Eq.15.

$$\frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \geq 0.9 \tag{15}$$

In this equation, λ_i, is the i-th eigenvalue, from the set of eigenvalues which have been sorted from highest to lowest.

In conservation *PCA* method, it is assumed the all data are available. Therefore, by increasing the number of eigenvectors, "k", for formation of "T" matrix, we can reduce the error between primary data and returned data. However, in our case, due to restrictions, all data is not available; so increasing of "k" (i.e., selection of smaller eigenvalues) to certain level, results in extracting of weak relations, which increases the error of load estimation.

In this paper, we have proposed a new concept for suitable selection of "k". The concept is based on a trade-off between the extracted features and load estimation error. As it can be seen in next section, the results of numerical study which is represented in Table 1, confirms that suitable selection the number of eigenvectors decreases the error between the actual and estimated load. For this propose, a new concept is addressed here.

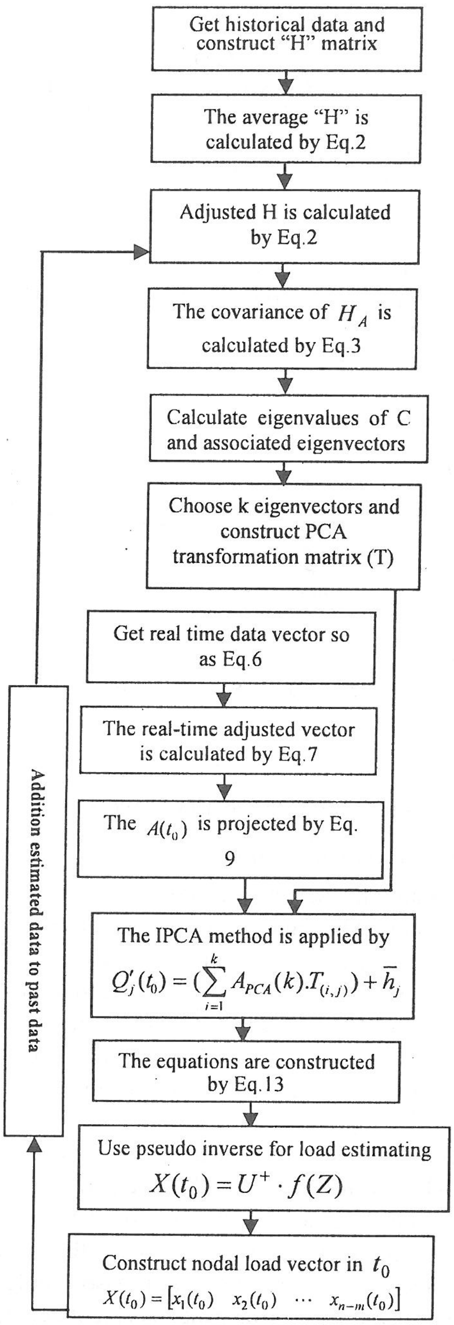

Fig 2: Flowchart of the proposed method.

Briefly, the procedure for selection of "k" is as follows:

Step 1 Get the historical data,

- Construct the "C" matrix
- Determine eigenvectors of the "C" matrix
- Set $k=1$ and Err = infinite

Step 2 Assessment of "k"

- Construct the "T" matrix
- Apply *IPCA* and calculatdoad estimation
- Calculate error (Err_k) between estimated and actual value

- Step 3 Comparison and consider
- If $Err_k < Err$ then $k=k+1$, $Err=Err_k$ and go to step 2, else go to step 4

Step 4 The suitable "k" is "$k-1$" and stop.

According to the experimental results, the proposed procedure can be converged after 3-7 iterations. The flowchart of the algorithm of the procedure is represented in Fig. 3.

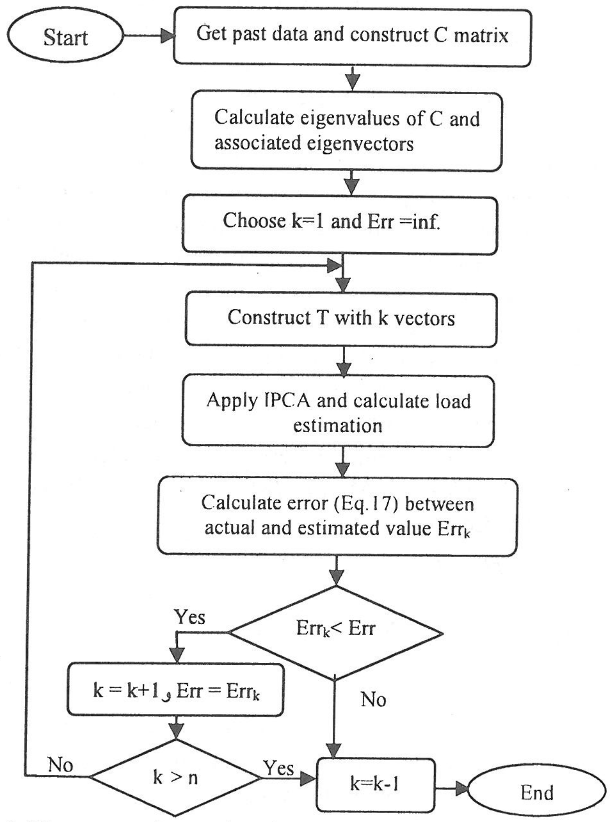

Fig.3: The proposed procedure for the "k" selection.

3 NUMERICAL STUDIES

3.1 The network under study

In order to evaluate the performance of the proposed method, an actual network which is a part of central Tehran distribution network with 24 load nodes is used. A single-line diagram of the underlying system is shown in Fig.4. There are 28500 consumers with a total peak of 35MW connected to this network.

The historical load data of a six month period are collected from low voltage side of distribution substations. It is tried to select a part of the collected data with a minimum effect of factors such as, temperature, special days, etc. Therefore, load data of 43 days were selected for the study. Furthermore, the load data of weekend days are separated from the above data. The maximum temperature difference in this period was 5^{0C} (27^{0C}~32^{0C}). It should be mentioned that in this system, distribution automation equipments are allocated according to the reliability concerns for power interruption. However, optimum allocation of measurement equipments would result in lower load estimation error. In the network under study, there are six substations which are equipped with remote real-time load measurement facilities which collect the active load data of the substations. Other substations have data loggers without remote facilities.

3.2 Selection of "k"

In order to study the effect of the number of eigenvectors "k", on the level of error between the primary data "H" and the returned primary data "H'", both the conventional PCA method and the developed IPCA method are investigated. Here, the primary matrix "H" is with 43 rows and 24 columns where 24 refers to the number of buses and 43 indicates the number of measurement samples. It is also assumed that the real-time measurement data are located in the first six columns of the data matrix. The returned data of PCA method are calculated according to Eq.16. For comparing between effects of the "k" value on PCA method and proposed method, it is assumed that values of all data are available for PCA method, but in the proposed method only six nodes' data are available and the other data are estimated by IPCA. The estimation error is calculated based on standard index of MAPE (Mean Absolute Percentage Error) as Eq. (17).

$$H' = H_A \cdot T^T \tag{16}$$

$$MAPE = \frac{1}{n \times r}\left(\sum_{i=1}^{r}\sum_{j=1}^{n}\frac{|h_{i,j} - h'_{i,j}|}{h_{i,j}}\right) \tag{17}$$

where, $n \times r$ in Eq.17 is the number of data samples. Fig.5 shows the principal components (eigenvalues) and Table 1 represents the error for different "*k*".

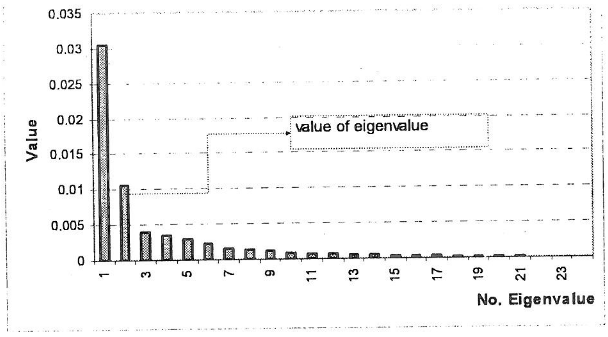

Fig. 5: Principal component of past data matrix.

As it can be seen in conventional *PCA* method, the returned data error is decreased by selecting larger "*k*" (it is assumed that all data are available.), but in the proposed *IPCA* method, the estimated error increases when "*k*" is larger than a definite value (if real-time data is restricted) see Fig.6. However, if the procedure described in section 2 for selecting of eigenvectors be used, more accurate estimation will be obtained. In our network study, *k=4* is a suitable selection for matrix formation "*T*", for which Eq.15 equals to 0.77.

TABLE 1:
EFFECTS OF "*K*" ON RETURNED PRIMARY AND ESTIMATED DATA.

The k value	MAPE value	
	PCA method	IPCA method
2	2.60%	2.10%
3	2.40%	2.07%
4	2.20%	2.03%
5	2.00%	2.05%
6	1.60%	2.38%
7	1.00%	2.60%
20	0.20%	3.07%
24	0.00%	5.02%

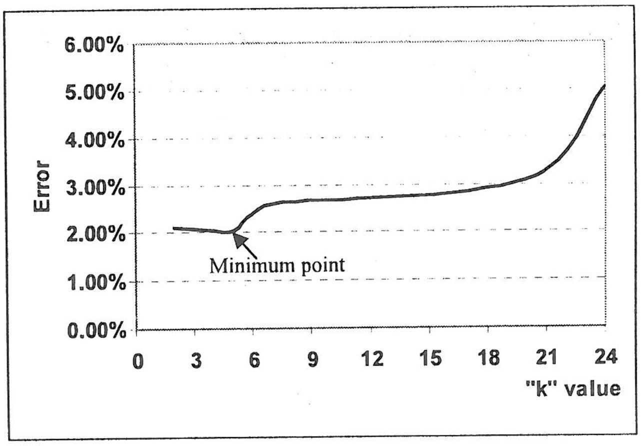

Fig. 6: Effect of "*k*" on estimated data by *IPCA*.

3.3 Results of load estimation

In this study, the selected data of 43 days are separated into two groups. The data of 30 days were used as the past data for feature extracting, and the remaining data of 13 days were used for testing of the load estimation performance by *IPCA* method. The six values of each day of second group data were known values (real-time data measurements) and other data (18 values) were unknown load data.

According to the results that were obtained, the error index MAPE for test data samples is equal to 2.9% which shows a good accuracy for the proposed method. Fig. 7 shows the cumulative error between the actual and estimated load (for 13 days test data). Also, Table 2 shows the per unit percentage error which is calculated according the following Eq.18.

$$Err^{Pu} = \frac{|P - \hat{P}|}{P_b} \tag{18}$$

In this equation, "*P*" is the actual load, $\hat{P}$ is the estimated load and "P_b" is the maximum load for 13 days period.

In order to investigate the performance and accuracy of the proposed method, the result of load estimation by the proposed method has been compared with other load estimation techniques including fuzzy logic, neural network and linear regression with minimum least square criteria. Fuzzy logic consists of the Gaussian membership functions for each input and sixteen rules between six inputs data and each output data. The neural network is a two layer perceptron network that consists of seven neurons. The perceptron network is learned by Levenberg-Marquardt method. Table 3 shows the error index *MAPE* for different methods applied to the study network. As it can be seen, the proposed method *IPCA* has lower error in comparing with the other methods.

TABLE 3:
ERROR INDEX MAPE FOR DIFFERENT METHODS

Method	IPCA	regression	Neural network	Fuzzy logic
Err. value	2.9%	3.5%	4.5%	4.3%

The reason of such accuracy relies on the fact that the proposed method can extract the features and interrelation of load variables very precisely.

4 CONCLUSION

In this paper, a new method for load estimation of distribution network was developed. *IPCA* approach presented here is an extension of *PCA* method which is capable of solving a large set of problems of high dimension and restricted with limited real-time data. Therefore, it opens a new solution space for similar problems with limited real-time data. The proposed approach is an analytical method, and the results of each stage of the solution can be used for other problems such as allocation of real-time measurement equipments which are suitable for load estimation.

Numerical studies on an actual distribution network confirm the accuracy of the method and its superiority to other methods.

Further work is under study regarding optimum allocation of real time measurement equipment as a necessary step for the extension of this research.

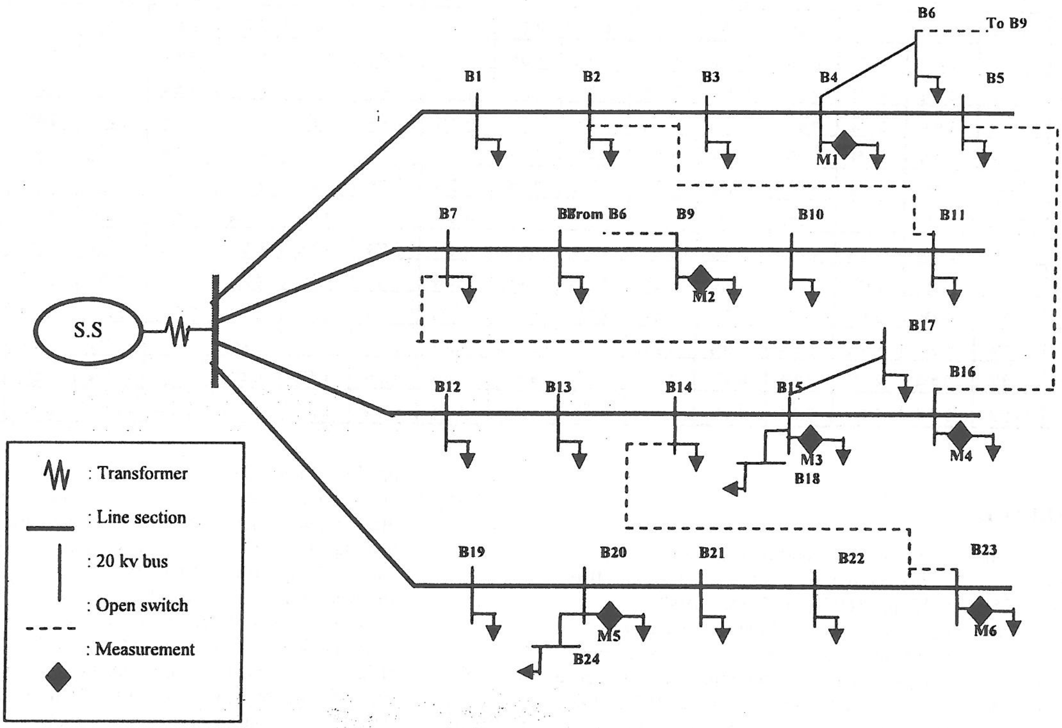

Fig. 4: Single-line diagram of the network under study.

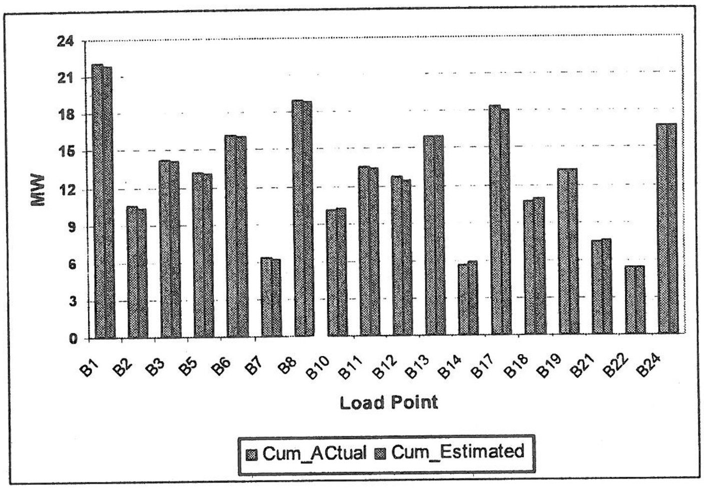

Fig. 7: The cumulative error result.

TABLE 2:
THE PER UNIT PERCENTAGE ERROR RESULT

Day	B1	B2	B3	B5	B6	B7	B8	B10	B11	B12	B13	B14	B17	B18	B19	B21	B22	B24
1	1.8%	0.5%	0.8%	1.6%	1.6%	0.4%	2.9%	4.1%	0.6%	1.4%	1.2%	1.6%	0.4%	2.9%	1.4%	2.2%	2.5%	0.9%
2	3.4%	0.4%	1.8%	2.1%	2.8%	3.1%	1.0%	2.0%	1.0%	1.0%	4.3%	2.0%	1.6%	3.2%	1.3%	0.8%	2.3%	3.3%
3	2.2%	1.3%	4.3%	2.0%	0.2%	0.0%	0.4%	1.0%	3.2%	2.8%	0.2%	0.6%	2.4%	0.5%	0.7%	1.0%	1.1%	0.1%
4	0.9%	0.7%	1.0%	3.1%	0.6%	1.9%	0.9%	1.0%	0.6%	2.4%	1.0%	1.4%	1.4%	1.7%	0.4%	0.4%	0.9%	1.9%
5	4.7%	1.2%	1.3%	5.6%	0.5%	2.5%	4.4%	2.9%	1.0%	6.1%	2.2%	0.8%	5.0%	0.1%	1.3%	0.4%	0.9%	3.2%
6	2.5%	2.6%	1.0%	2.3%	0.7%	1.1%	1.5%	0.2%	0.7%	2.8%	0.7%	2.6%	0.2%	1.6%	0.4%	0.1%	2.7%	1.7%
7	2.5%	2.6%	1.0%	2.3%	0.7%	1.1%	1.5%	0.2%	0.7%	2.8%	0.7%	2.6%	0.2%	1.6%	0.4%	0.1%	2.7%	1.7%
8	1.2%	0.2%	2.8%	0.8%	0.3%	0.3%	3.8%	1.4%	1.2%	4.0%	1.3%	0.3%	5.3%	1.3%	2.5%	1.6%	1.7%	3.7%
9	2.5%	3.4%	1.1%	1.6%	2.1%	1.3%	4.3%	0.5%	0.4%	0.4%	2.7%	3.0%	2.2%	1.8%	1.3%	1.0%	0.8%	3.2%
10	2.7%	2.2%	0.4%	2.1%	2.4%	1.8%	1.1%	0.2%	0.3%	1.3%	0.1%	2.1%	0.0%	0.1%	2.2%	0.5%	0.4%	3.5%
11	2.7%	1.3%	1.2%	1.3%	0.3%	1.9%	4.2%	1.1%	0.2%	1.9%	2.1%	1.7%	3.6%	1.7%	2.7%	1.4%	1.9%	2.2%
12	3.5%	0.9%	0.9%	2.1%	0.1%	0.6%	1.3%	0.1%	0.5%	1.2%	1.2%	1.8%	0.8%	1.1%	0.1%	1.5%	0.3%	0.4%
13	1.2%	3.0%	0.7%	0.9%	0.4%	2.9%	2.4%	2.9%	0.4%	4.4%	0.6%	0.9%	3.5%	0.5%	1.5%	2.8%	2.3%	1.6%

5 REFERENCES

[1] A. Bose and K.A.Clements, "Real-time modeling of power networks", Proceedings of the IEEE Volume 75, Issue 12, pp. 1607 – 1622, Dec. 1987.

[2] V. Borozan and N.Rajakovic, "Minimum loss distribution network configuration: Analyses and management", Electricity Distribution Part1 : 14th International Conference and Exhibition on Electricity Distribution (CIRED 1997 - Distributing Power for the Millennium) ,No. 438, Vol.6, pp.6-18, June 1997.

[3] R.P.Broadwater, A.H.Khan, H.E.Shaalan and R.E.Lee, "Time varying load analysis to reduce distribution losses through reconfiguration", IEEE Trans. Power Delivery, Vol. 8, No. 1, pp. 294 – 300, Jan. 1993.

[4] H.Wang, N. Suhulz, "A load modeling algorithm for distribution system state estimation", T&D Conference and Exposition, 2001IEEE/PES,Vol:1, pp.102-105, 2001.

[5] M.E.Baran and A.W. Kelley, "State estimation for Real-time monitoring of distribution systems", IEEE Trans. Power Systems, Vol. 9, No.3, pp. 1601-1609, Aug. 1994.

[6] T.Wang and M.Fan, "A Novel load estimation method in distribution network", Proceeding 1998 International Conference on Power Sys. Tech., Vol. 1, No. 1, pp. 567-571, Aug 1998.

[7] Atish K. Ghosh, David L, Lubkeman, Rober H. Jones, "Distribution state estimation using A probabilistic approach", IEEE Trans. on Power Systems, Vol. 12, No. 1, pp. 45-51, Feb. 1997.

[8] D. M. Falcao and H. O. Henriqes, "Load estimation in radial distribution systems using neural network and fuzzy set techniques", Power Engineering Society Summer Meeting Volume 2, pp. 1002 - 1006, 2001.

[9] Jie Wan and Karen Man Miu, "A zonal load estimation for unbalance, radial distribution networks", IEEE Trans. Power Delivery, Vol. 17, pp. 1106 - 1112, Oct. 2002.

[10] H.Liu, D.Yu, H.D. Chiang, "A heuristic meter placement method for load estimation", IEEE Trans. On Power Systems. Vol. 17, No. 3, pp. 913 - 917, Aug. 2002.

[11] Jie Wan and Karen Man Miu, "Weighted least squares methods for load estimation in distribution networks", IEEE Trans. On Power Systems, Vol. 18, No.4, pp. 1338 – 1345, Nov. 2003.

[12] Shigenori Naka, Takamu Genji, Toshiki Yura, and Yoshikazu Fukuyama, "A hybrid particle swarm optimization for distribution state estimation", IEEE Trans. On Power Systems, Vol. 18, No. 1, Feb. 2003.

[13] M.E. Baran, Lavelle A. A. Freeman and Frank Hanson, "Load estimation for load monitoring at distribution substation", IEEE Transactions on Power Systems, Vol. 20, No. 1, Feb. 2005.

[14] A.Yaghoti, M. Parsa and V. J. Majde, "State estimation in distribution systems",18TH Conference and Exhibition on Electricity Distribution, IEE , June 2005.

[15] Yongjian Fu; "Data mining" ,Potentials, IEEE, Vol. 16, Issue 4, pp. 18 – 20, Oct-Nov 1997.

[16] Zohreh Nazeri, Jianping Zhang, "Mining aviation data to understand impacts of severe weather On airspace system performance", Coding and Computing, 2002. Proceedings International Conference, pp. 518 – 523, April 2002.

[17] I. Higuchi, S. Eguchi, "Robust principal component analysis with adaptive selection for tuning parameters", Journal of Machine Learning Research5, pp. 453-471, 2004.

[18] M. Tipping and C. Bishop, "Probabilistic principal component analysis". Journal of the Royal Statistic Society, 61(3), pp. 611-622, Sep.1999.

[19] Vidal, R. Yi Ma Sastry; S. , "Generalized principal component analysis (GPCA)", Computer Vision and Pattern Recognition, Proceedings, Vol:1, pp: I-621- I-628 , June2003.

[20] J.W. Taylor, L.M. Menezes and P.E.McSharry, "A comparison Of univariate method for forecasting electricity demand up to a day ahead", Elsevier, International Journal of Forecasting 22, pp. 1– 16, 2006.

[21] R.K.LIVESLY , "Mathematical methods for engineering" Ellis Horwood Limited, 1989 .