AJ ACADEMIC JOURNALS
expand your knowledge

SCIENTIFIC RESEARCH AND ESSAYS

Full Length Research Paper

New application of principal component regression in estimation of electrical energy consumption in an abnormal automatic meter reading system

Visavat Kantikoon* and Vijit Kinnares

Department of Electrical Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Thailand.

This paper proposes a new application of principal component regression (PCR) for estimating electrical energy consumption in case of abnormal automatic meter reading (AMR) systems. These events occur in a delivery metering system such as problems from mistakenly setting and connecting meters in electrical systems, broken metering accessories, etc. The estimation is performed by using MATLAB. The unclean sampled input data is used to estimate the target output data. The mean absolute percentage error (MAPE) is used as estimation performance. In this proposed estimation, load profiles obtained from the AMR are used as input data for training to create estimation model and for testing to validate model. Estimated results are verified by comparison between the proposed PCR application and other applications such as simple linear regression (SLR), multiple linear regression (MLR). The proposed PCR gives the best error results of MAPE for the lost electrical energy estimation.

Key words: Automatic meter reading (AMR), load profiles, principal component regression (PCR), multiple linear regression (MLR), simple linear regression (SLR).

INTRODUCTION

Presently, technologies of energy meter have developed rapidly (Alahakoon and Yu, 2016). In particular, the cost of energy meter technology is greatly reduced. The reason is that most electrical energy providers pay attention in the energy meter technology development and the energy consumption data record system for the customer monthly payment operation.

In the past, most electrical energy providers chose a mechanical or electronic meter for the electrical energy consumption measurement of the customer and the monthly energy payment operation, because of the price which was lower than a smart meter. But in the present, those electrical energy providers have required database technology, energy management and electrical energy consumption history. An automatic meter reading (AMR) is used for those requirements and designed to be used with smart meters (Paris et al., 2014; Stephen et al., 2014).

A smart meter has more advantages than the electronic meter in electrical energy consumption history capacity that data history period limit is 45 days and the cost of both gets nearly closer in the present.

So this is inclination that the mechanical meter usage is decreased and the mechanical meter will be replaced by the smart meter in the present day and the near future.

The prominent feature of electrical energy data recording in the smart meter is specified to be used with the AMR which the electrical energy consumption data in the energy meters for all customers will be recorded at the database of the AMR.

Smart grids are the extension, which themselves are built upon AMR systems (Arif et al., 2013; Khan et al., 2014). AMR systems are achieved by using many communication technologies, including power-line, radio frequency and mobile network such as GPRS/GSM. As many energy providers look for upgrade progress toward smart grids, replacing mechanical meters by the entire infrastructure may not be economical. For that reason, most smart grid developing providers have formulated priority of suitable smart grid development region.

In Thailand, provincial electricity authority (PEA) has formulated framework of smart grid (Meenual and Thongchai, 2009). Current smart grid development of PEA is in the preparation stage, including suitable PEA smart grid Technologies selection and adaptation, implementation plan setting, and necessary PEA smart grid foundation development. PEA intends to aim at pilot project of PEA smart grid in the near future.

Currently, a soft computing method is widely used to determine the reliability of the distribution system or to predict the electricity load demand (Singh et al., 2013). The classification of distribution system loss in case of the economical profit can be defined into two items, namely technical loss and non-technical loss. The technical loss occurs by loss of transmission line and equipment in distribution line like copper loss. Unlike technical loss, non-technical loss is caused by abnormal metering equipment or violation and meter tampering by customer.

There are several methods used for the non-technical energy loss estimation which is aimed to claim payment from customers in case of energy theft or damaged equipment by customer or forced majeure. One method commonly used by most providers to estimate is the average monthly electricity consumption of the previous three months abnormal, or three months after an abnormal ending for consideration of that estimation. Those methods have limitations in affecting the performance and reliability of the estimation. Especially, the condition of the electrical consumption behavior will be relatively stable and continuous.

There is rarely research on the lost energy estimation in abnormal metering of electricity customer such as using simple linear regression (SLR) with a focus on a minimum variable input reduction for estimation of output, but some significant data is limit. The multiple linear regression (MLR) (Black and Henson, 2014) was used to estimate for improving the limitation of SLR application. However, MLR is still limited in the case of uncleanness

of input data to be used for simulation of the model in some times.

In this paper, the sampled data is considered from the energy consumption data of the industrial energy customers in Thailand. These energy data will be managed by using AMR system (Ladarat and Naetiladdanon, 2015) of PEA in Thailand. The energy data history recorded in the data center of AMR will be used to increase the efficiency of the energy data management for customers and providers. Thereafter, the load profile of energy consumption will be used as input data for the real energy estimation in case of the abnormal metering system.

OPERATION OF AUTOMATIC METER READING

In the present study, most energy providers have more interest in the automatic meter reading (AMR) (Figure 1). This is the reason why many business factors have become a necessity for the advantage in business competition, such as the cost saving capacity of energy providers and the energy saving by monitoring and management of customers.

The main part of an AMR system is electrical energy usage measure and recording equipment. The smart meter is an important electrical measure equipment, like a joint of service and communication between providers and customers in real time.

The communication system is necessary for AMR as it plays an important part in energy consumption data transmission automatically, such as load profiles, monthly billing data, and alarm logs of the smart meter. These data will be transmitted to a meter interface unit (MIU) and be saved in the database of AMR system (AMR data center) for every fifteen minutes. Therefore, the energy provider can take advantage of the energy consumption data monitoring for non-technical loss or illegal electricity usage detection (Erdene et al., 2013). Similarly, the customers can observe and manage the self-energy consumption demand.

PRINCIPAL COMPONENT REGRESSION

Linear regression analysis for the electrical fields has been widely used in the load demand forecasting. Normally, the algorithm of linear regression is classified into two types according to the number of input data, including single linear regression (SLR) in case of single input data and multiple linear regression (MLR) in case of multiple input data. In this case study, both SLR and MLR are used for a comparison with the proposed algorithm.

Figure 2 shows the proposed algorithm of the principal component regression (PCR) models consisting of MLR and the principal component analysis (PCA).

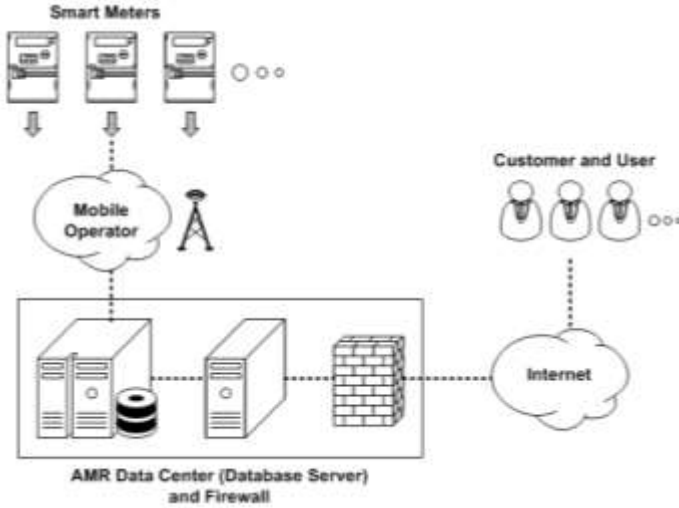The estimate procedure of regression has two parts,

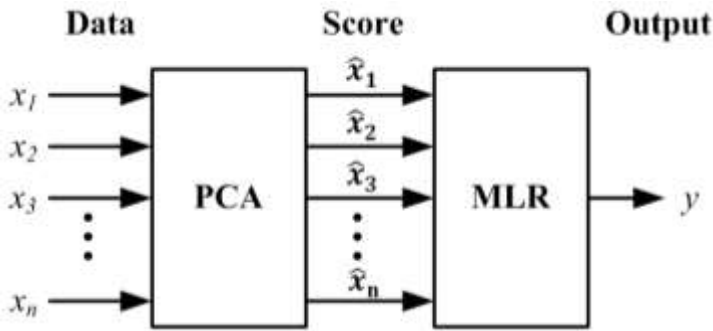**Figure 1.** Structure of the AMR system.

Figure 2. PCR models.

including model creation and estimation. The regression learns a function that maps input variables to their target output. Conventionally, that function is a static function enabling the estimation of responses for new input variables. The multiple regression model can be expressed as

$$y_j = W_j \mathbf{x} + \varepsilon_j, \quad \text{for } j = 1, 2, \ldots, n, \qquad (1)$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_m]^T \in \mathfrak{R}^m$ is the input variables matrix, $\mathbf{y} = [y_1, y_2, \ldots, y_n]^T \in \mathfrak{R}^n$ is the output variables matrix, $W_j = [W_{j1}, W_{j2}, \ldots, W_{jm}]$ is the jth row of the regression matrix W, and $\varepsilon_j \sim N(0, \sigma^2)$ denotes the estimation error in which σ^2 is the variance of the observation.

In step of estimation, a training part for model creation of input/output variables set pairs $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^N$, is supposed

to be a linear relationship between these two sets of variables W. It can be estimated as

$$\min \sum_i^n (\mathbf{y}_i - W\mathbf{x}_i)^2, \qquad (2)$$

The equivalent matrix norm equation is expressed as

$$\min \| \mathbf{Y} - W\mathbf{X} \|^2, \qquad (3)$$

where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$, $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N)$, and $\| \cdot \|$ denotes the matrix 2-norm. This is an optimization problem which can be simply resolved as

$$W = (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X}^T \mathbf{Y}. \qquad (4)$$

With the estimated regression coefficient, the output variables can be predicted for the input variables. Generally, Equation 4 requires that $m = rank(\mathbf{X}\mathbf{X}^T) < N$, or $\mathbf{X}\mathbf{X}^T$ is invertible. However, this condition may not always be satisfied.

To avoid the unclean input data for training and the irreversible problem of MLR, PCR replaces the input variables by principal components to estimate the output variables. Particularly, PCR firstly projects $\mathbf{x}$ onto a low-dimensional subspace

$$\mathbf{x} = P^T . \hat{\mathbf{x}}, \qquad (5)$$

where $\hat{\mathbf{x}} \in \mathfrak{R}^k$ with $k < rank(\mathbf{X}) < m$, and the value of P stands for the eigenvectors of subspace $\mathfrak{R}^k$ which can be obtained by eigenvalue decomposition techniques. Therefore, the solution reduces to

$$W_p = (\hat{\mathbf{X}}\hat{\mathbf{X}}^T)^{-1} \hat{\mathbf{X}}^T \mathbf{Y}, \qquad (6)$$

where $\hat{\mathbf{X}} = (\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_N)$.

In case of appropriate k eigenvectors which are chosen, $\hat{\mathbf{X}}\hat{\mathbf{X}}^T$ is invertible. The optimization problem can be simply resolved.

The transformation of P is obtained by choosing k eigenvectors of the column subspace of $\mathbf{X}$. The singular value decomposition of $\mathbf{X}$ can be considered as

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \qquad (7)$$

where $\mathbf{U}$ is an orthogonal matrix of scaled principal component scores, $\mathbf{V}$ is an orthonormal matrix of eigenvectors, and $\mathbf{S}$ is a diagonal matrix of the singular values, like the dimension of $\mathbf{X}$, respectively.

Table 1. Parameters of the model simulation using the proposed application

Abnormal variables	Input variables	Input PCs variables	Output variables
I_a	$I_b, I_c,$ V_a, V_b, V_c	PC_1, PC_2, PC_3	kWh
I_b	$I_a, I_c,$ V_a, V_b, V_c	PC_1, PC_2, PC_3	kWh
I_c	$I_a, I_b,$ V_a, V_b, V_c	PC_1, PC_2, PC_3	kWh
V_a	$I_a, I_b, I_c,$ V_b, V_c	$PC_1, PC_2, PC_3,$ PC_4, PC_5	kWh
V_b	$I_a, I_b, I_c,$ V_a, V_c	$PC_1, PC_2, PC_3,$ PC_4, PC_5	kWh
V_c	$I_a, I_b, I_c,$ V_a, V_b	$PC_1, PC_2, PC_3,$ PC_4, PC_5	kWh

Where I_a, I_b, I_c: is the current of phase a, b, and c ; V_a, V_b, V_c: is the voltage of phase a, b, and c ; $PC_1,..., PC_5$: is the principal component of 1st to 5th ; and kWh: is the electrical energy used in one hour.

MODELING

In this case study, the aims are to estimate the energy consumption data of an industrial electricity customer of PEA in Thailand. This will be used in case of the non-technical loss energy which is the results of the abnormal metering from the AMR system in every fifteen minutes. And then, the data selection and statistic error will also be shown here in this work.

Data selection

The proposed PCR is used for the electrical energy estimation which is created in two models, including training model and testing model.

The variables of training model consist of the sampled input data and the sampled output data. The sampled input data are the voltages and the currents, whereas the sampled output data is the energy equivalent to one kilowatt of power sustained for one hour (kWh). Those are recorded before abnormal metering data period.

The variables of testing model consist of the sampled input data. These are the voltages and currents recorded during abnormal metering data period.

The parameters of the model simulation are used in the proposed PCR application. As shown in Table 1, items are the abnormal and normal input variables, the input PCs variables, and the output variables in units of kilo-
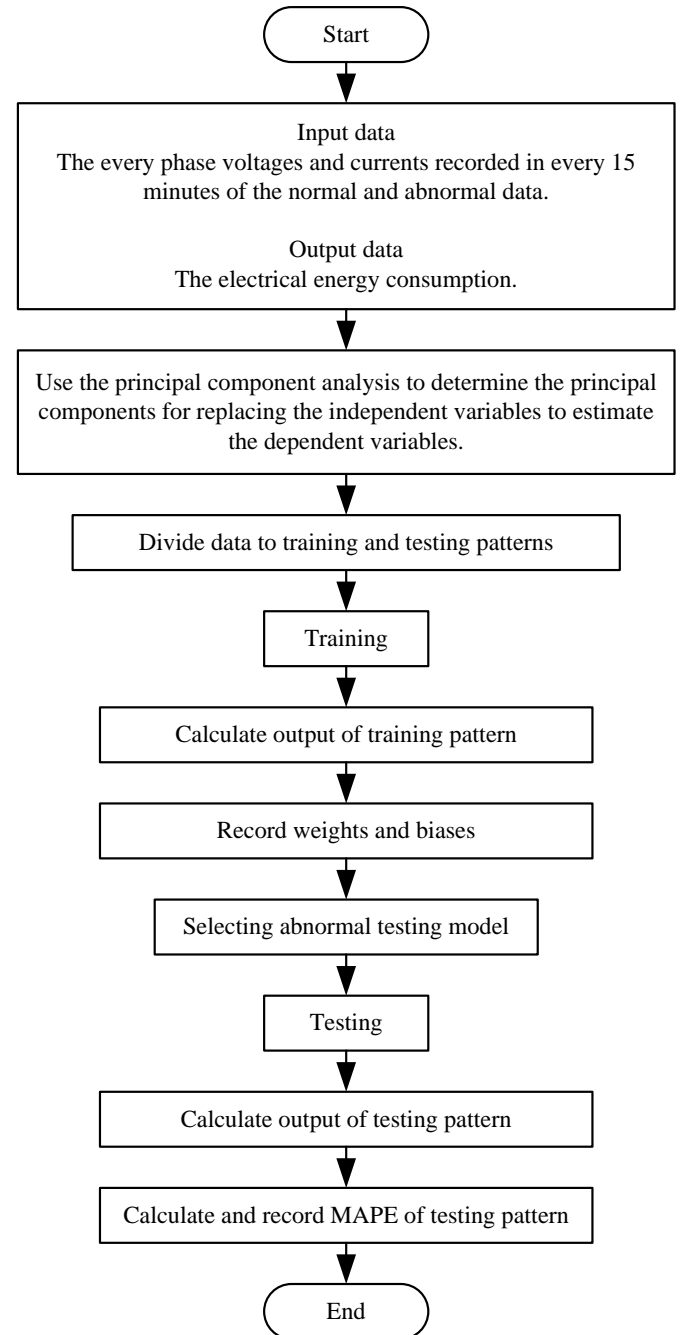
Figure 3. Flowchart diagram for the proposed estimation.

watt hours, which are used for the estimation.

The flowchart for the process estimation of electrical energy consumption by the PCR application is proposed as shown in Figure 3.

Statistic error

In this paper, the statistic error can be calculated by using
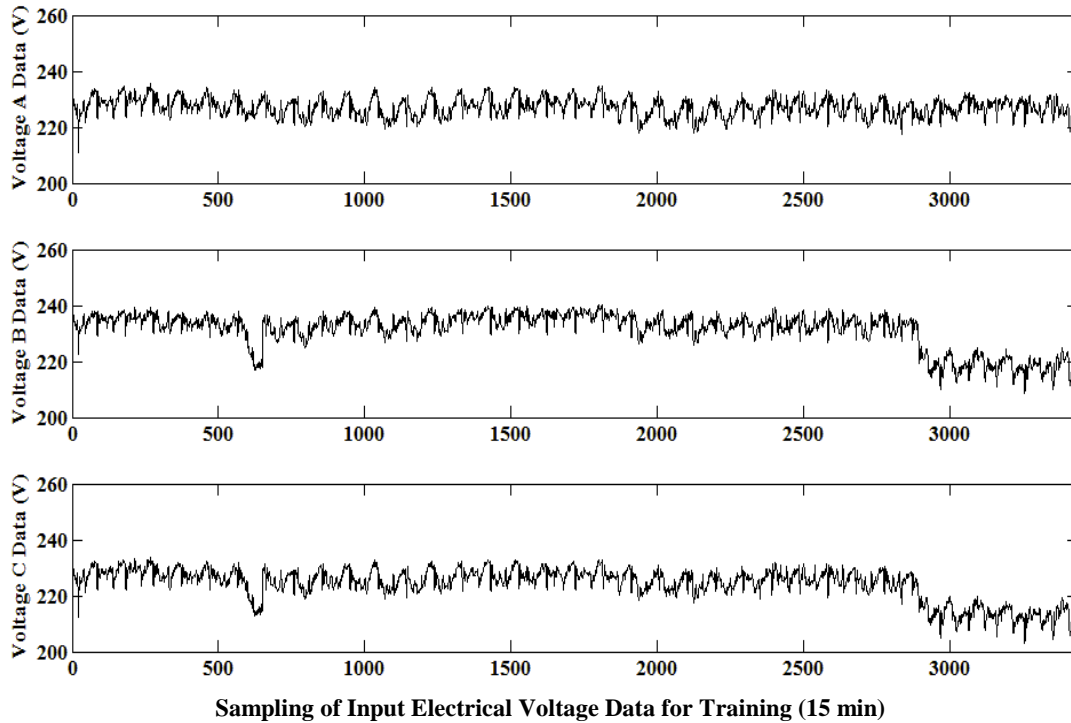
Figure 4. The input voltage data of the training.

mean absolute percentage error (MAPE) which is widely used in forecasting fields.

Because the results of this case study are for every fifteen minutes energy estimates, MAPE was also used to evaluate the estimate accuracy for every fifteen-minute energy. The formulas are expressed as follows,

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_{actual_i} - y_{estimate_i}}{y_{actual_i}} \right|, \tag{8}$$

where n is the number of observations, at the 15 min interval i , y_{actual_i} represents the actual electrical energy consumption, and $y_{estimate_i}$ represents the estimated electrical energy consumption, respectively.

SIMULATION RESULTS

This simulation uses the estimation for the abnormal voltage and current of any phase. PCR using 5 principal components (PCs) of the input parameters in case of the abnormal voltage of any phase, and 3 principal components (PCs) of the input parameters in case of the abnormal current of any phase is employed. The proposed estimation results will be used to compare with the estimation results using MLR for 5 normal input parameter and SLR for only 1 normal input parameter.

In order to demonstrate the error of the clear estimation, the unclean input voltage data that reflect the advantages of the method over other method is used. The unclean input voltage data can be found frequently in the case of an oxide at the joint of the electrical circuit wires or the electrical wire terminals of the meter, especially in case where the meter is installed in areas with high pollution.

In the training, the input and output data are achieved from the AMR system before abnormal metering occur. For example, the recorded input voltage data and current data of phases a, b, and c are illustrated in Figures 4 and 5, respectively. Also, the recorded electrical energy consumption data or target data is shown in Figure 6.

The input data is used for MLR in order to confirm the accuracy in comparison with PCR. The results of the estimation error are shown in Figure 7 in the event of irregularities such as the abnormal input voltage data of phases a, b, and c .

The estimation error results in using MLR are shown in Figure 8, in case of the abnormal input current data of phases a, b, and c.

The same output and input data with MLR for training is modeled in the proposed PCR by using MATLAB program, 5PCs matrix for the input training data in case of the abnormal voltage of any phase are shown in Table 2. 3PCs matrix for the input training data to create an MLR model in case of the abnormal input current data of any phase is shown in Table 3.
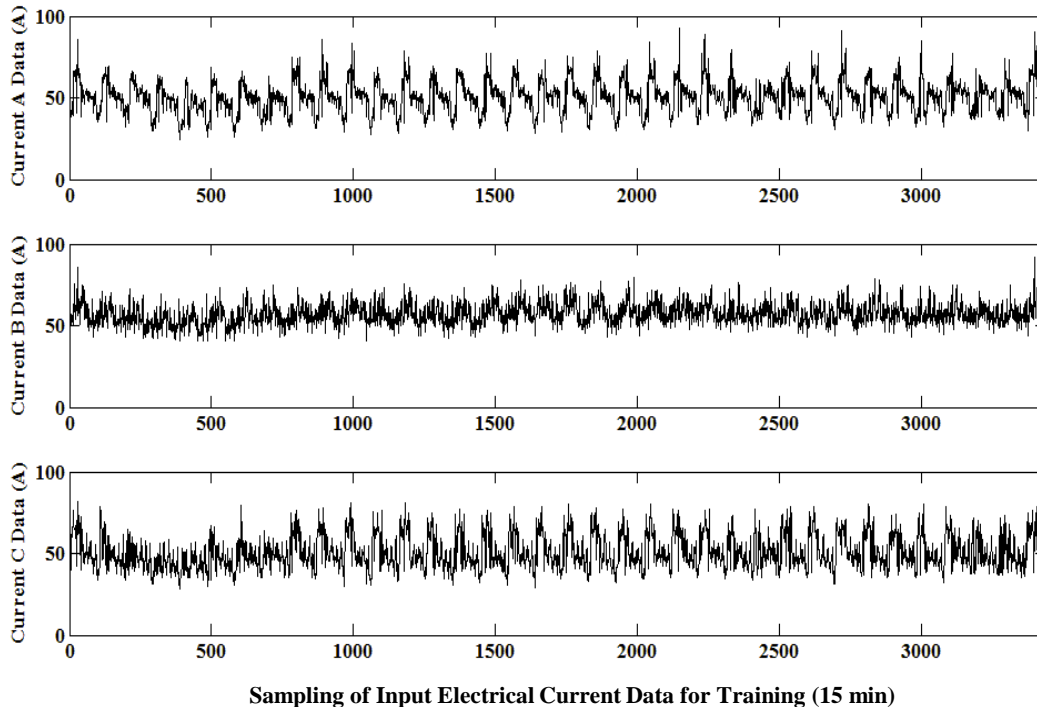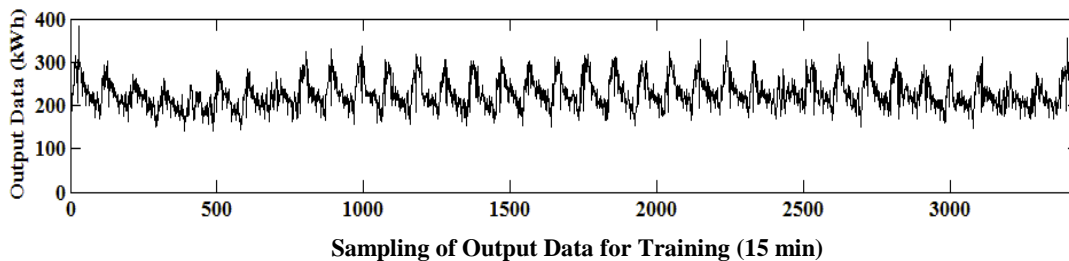
Figure 5. The input current data of the training.

Figure 6. The output energy data of the training.

The regression matrix W, and the estimation error matrix ε derived from the model simulation with the input training data are shown in Table 4, for using as a weight and bias of the testing model which is used to estimate the output testing data in case of abnormal input voltage data of any phase.

Also, the parameters of the regression matrix W and the estimation error matrix ε for estimating the output testing data in case of the abnormal input current data of any phase are shown in Table 5.

The testing data obtained from the AMR consists of the voltages and currents data of phases a, b and c for testing as shown in Figures 9 and 10, respectively. The measured electrical energy consumption for testing are shown in Figure 11.

The input and output data for testing measured at other times is used to confirm the accuracy of the estimation by using the proposed PCR application and other applications.

The estimated energy results and the estimation error of the proposed PCR in case of the abnormal input voltage data of phases a, b, and c are shown in Figures 12 to 14, respectively.

Figures 15 to 17 show the estimated energy consumption and the estimation error in case of the abnormal input current data of phases a, b and c by the proposed PCR, respectively.

According to results from Figures 12 to 17, they are in close estimation. They show that the proposed energy estimation method offers satisfactory performance.

For MAPE results of the proposed PCR, as shown in Table 6, it is found that MAPE is 4.48% in case of the abnormal input voltage data of phase a, 3.55% in case of the abnormal input voltage data of phase b, and 3.23% in
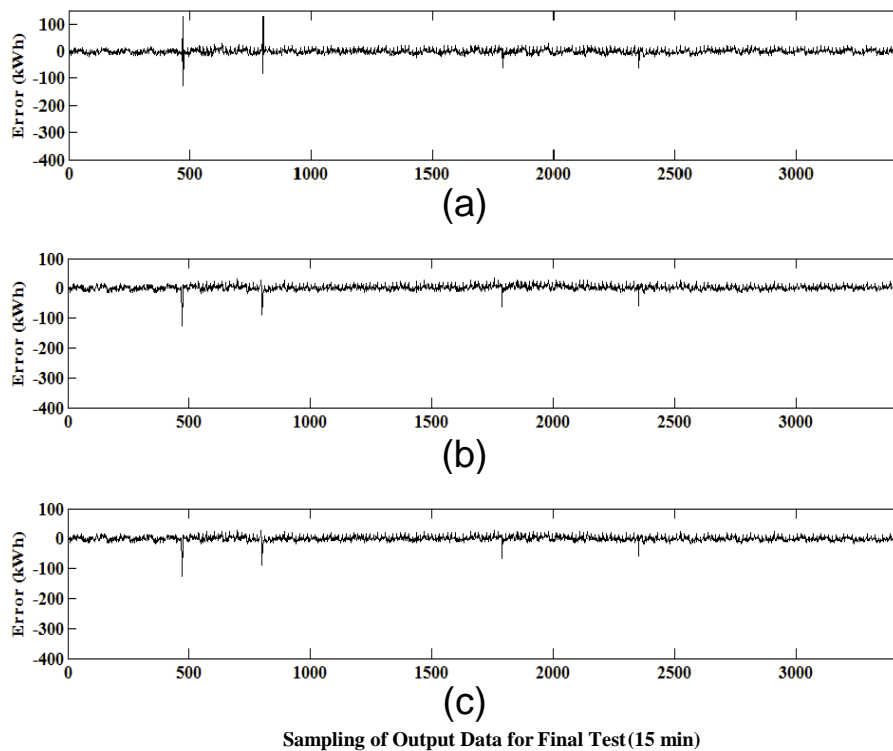
Figure 7. The output estimated energy errors by MLR for the abnormal input voltage data of each phase (a, b, c).
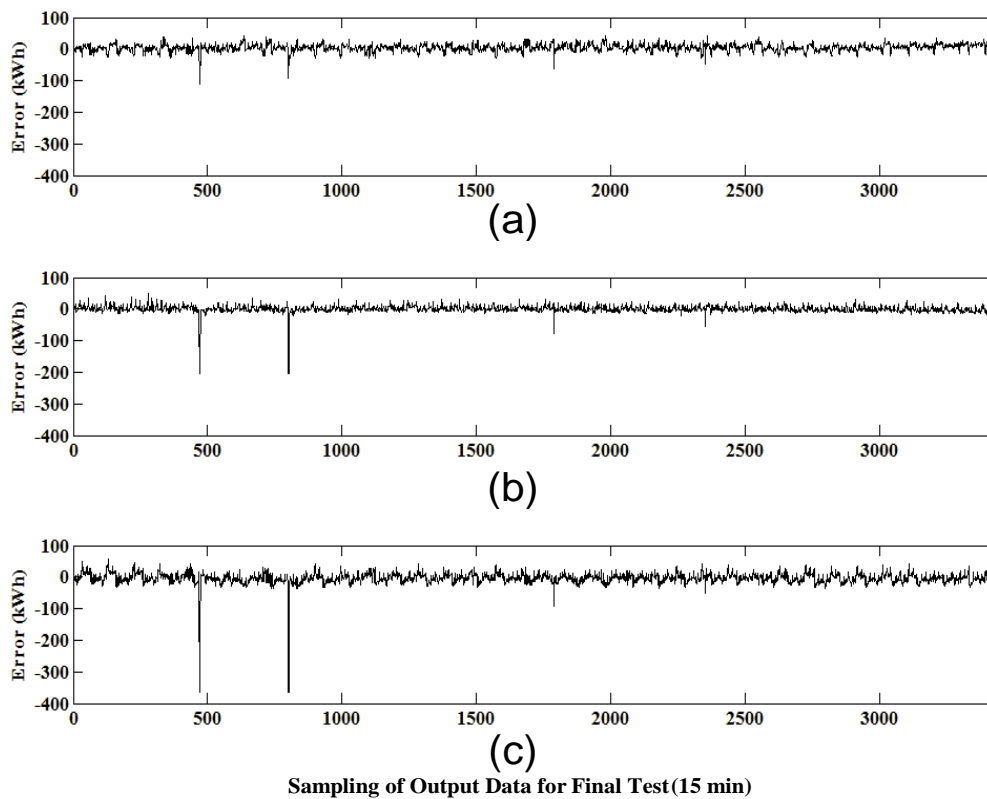

Figure 8. The output estimated energy errors by MLR for the abnormal input current data of each phase (a, b, c).

Table 2. 5 PCs matrix of the simulated model using PCR in each case of abnormal input voltage data.

Abnormal voltage of	PCs matrix				
Phase a	0.6242	-0.2429	0.1602	0.7250	-0.0111
	0.5109	0.8399	0.0599	-0.1714	0.0220
	0.5787	-0.4845	-0.0224	-0.6554	0.0160
	-0.0780	-0.0073	0.7197	-0.1047	-0.6818
	-0.0914	-0.0252	0.6725	-0.0672	0.7309
Phase b	0.6167	-0.2470	0.6694	-0.3287	-0.0505
	0.5095	0.8387	-0.0720	0.1717	0.0478
	0.5777	-0.4843	-0.5011	0.4107	0.1091
	-0.1368	-0.0188	0.3965	0.4435	0.7918
	-0.0872	-0.0254	0.3720	0.7050	-0.5969
Phase c	0.6186	-0.2421	0.6756	-0.3100	-0.0792
	0.5094	0.8410	-0.0940	0.1343	0.0800
	0.5786	-0.4836	-0.5068	0.3895	0.1508
	-0.1350	-0.0124	0.3798	0.3522	0.8446
	-0.0698	0.0061	0.3658	0.7811	-0.5012

Table 3. 3 PCs matrix of the simulated model using PCR in each case of abnormal input Current data.

Abnormal current of	PCs matrix		
Phase a	0.6961	0.7143	0.0296
	0.6558	-0.6639	0.3392
	-0.2042	0.1055	0.2756
	-0.1362	0.1453	0.6546
	-0.1583	0.1296	0.6162
Phase b	0.6928	0.2407	-0.6779
	0.6898	0.0278	0.7022
	-0.1474	0.3412	-0.0958
	-0.1009	0.6556	0.1553
	-0.1106	0.6286	0.1187
Phase c	0.6737	0.4666	-0.5725
	0.7054	-0.1945	0.6753
	-0.1483	0.3146	0.1227
	-0.1070	0.5762	0.3364
	-0.1229	0.5599	0.2967

case of the abnormal input voltage data of phase c in which MAPE is equal to that of the MLR application.

In addition, SLR using the best only 1 input current data of the model (which is not the abnormal current) will be used for estimating the output energy data in all cases of the abnormal input voltage data, for which MAPE is

Table 4. W and ε matrix of PCR for the simulated model in each case of abnormal input voltage data.

Abnormal voltage of	W and ε matrix
Phase a	$W = \begin{bmatrix} 0.6672 \\ -0.1339 \\ 0.1876 \\ -0.1930 \\ -0.3201 \end{bmatrix}$, $\varepsilon = [0.0103]$
Phase b	$W = \begin{bmatrix} 0.6675 \\ -0.1345 \\ -0.0489 \\ 0.2495 \\ -0.3194 \end{bmatrix}$, $\varepsilon = [-0.0205]$
Phase c	$W = \begin{bmatrix} 0.6688 \\ -0.1299 \\ -0.0498 \\ 0.2598 \\ -0.2236 \end{bmatrix}$, $\varepsilon = [0.0017]$

Table 5. W and ε matrix of PCR for the simulated model in each case of abnormal input current data.

Abnormal current of	W and ε matrix
Phase a	$W = \begin{bmatrix} 0.7154 \\ -0.2224 \\ 0.3646 \end{bmatrix}$, $\varepsilon = [-0.0021]$
Phase b	$W = \begin{bmatrix} 0.7016 \\ 0.1891 \\ 0.1617 \end{bmatrix}$, $\varepsilon = [0.0124]$
Phase c	$W = \begin{bmatrix} 0.6529 \\ 0.2624 \\ -0.1413 \end{bmatrix}$, $\varepsilon = [0.1024]$

10.06% in case of the abnormal input voltage data of phases a, b and c.

For MAPE results of the proposed PCR, as shown in Table 7, it is found that MAPE is 4.58% in case of the abnormal input current data of phase a, 3.75% in case of the abnormal input current data of phase b, and 8.03% in case of the abnormal input current data of phase c. MAPE for the proposed PCR is the least error value in contrast to the other applications.

From MAPE results for MLR, the normal current and voltage of all phases will be used for estimating the output energy data in case of the abnormal input current
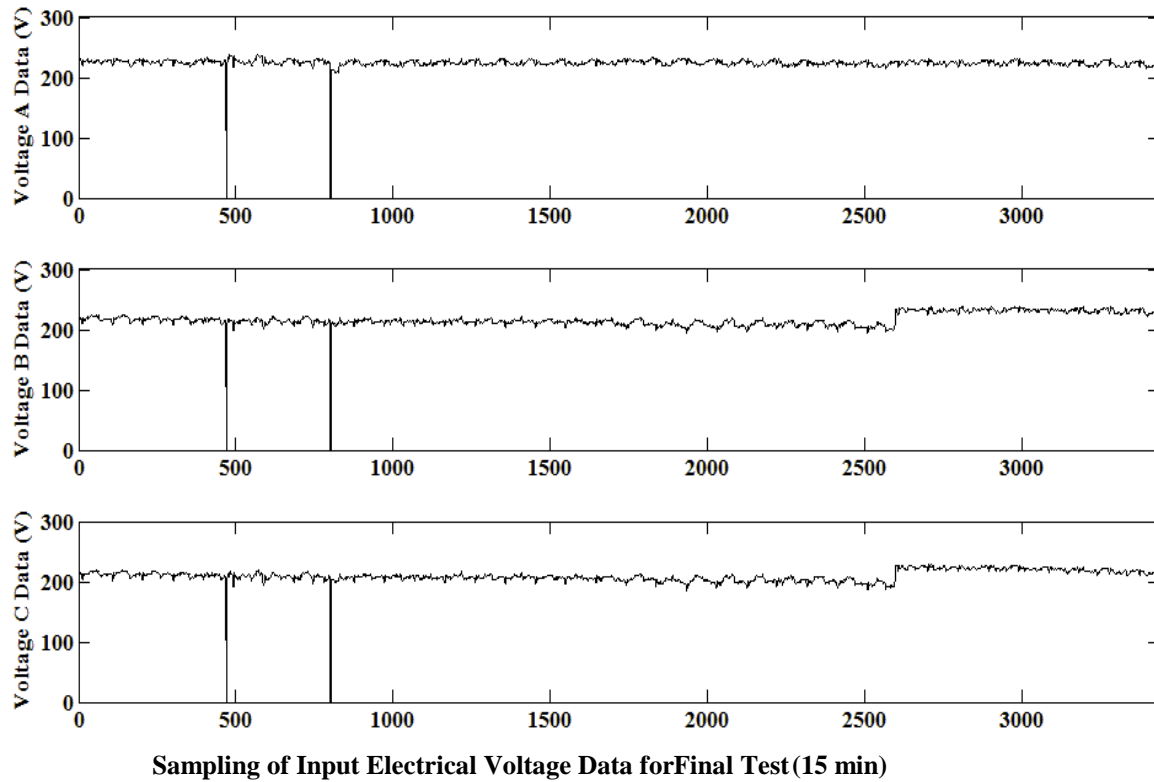
Sampling of Input Electrical Voltage Data for Final Test (15 min)

Figure 9. The input voltage data of the testing.

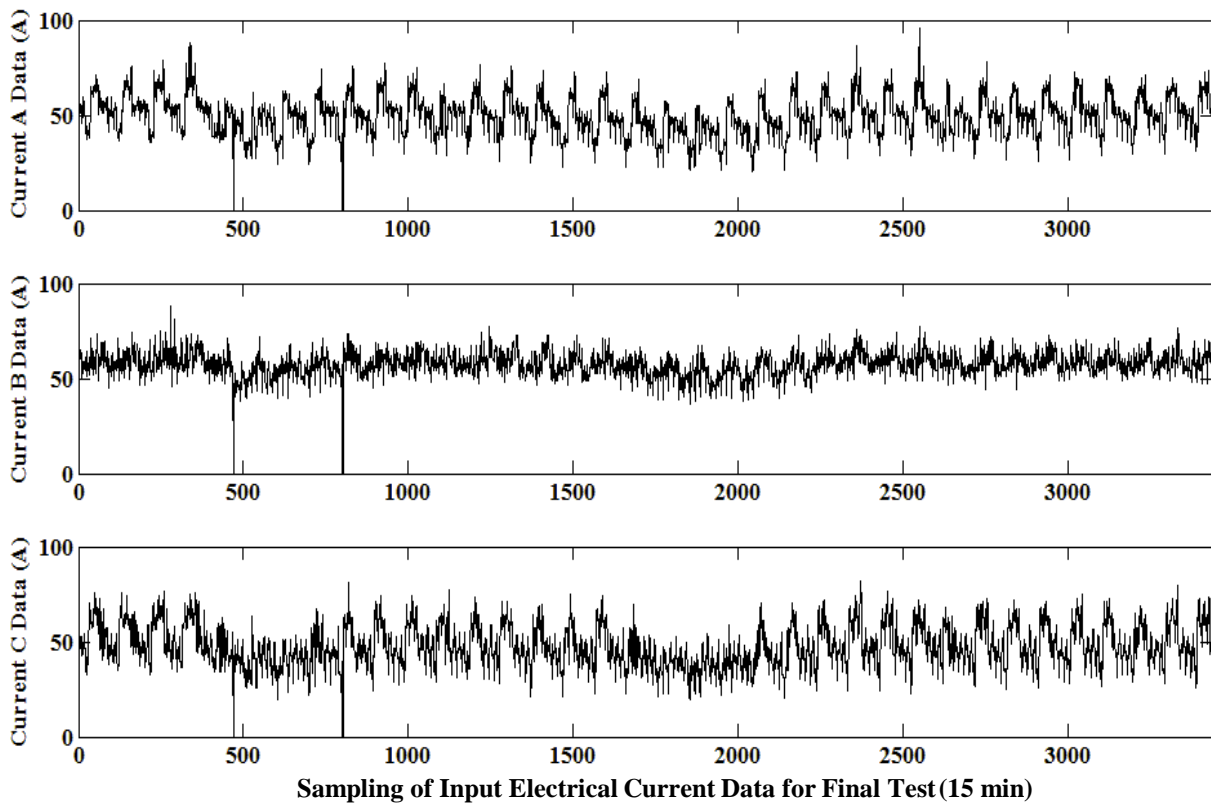**Sampling of Input Electrical Current Data for Final Test (15 min)**

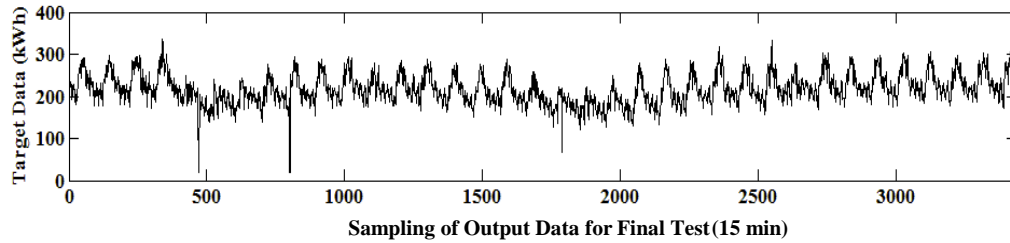**Figure 10.** The input current data of the testing.
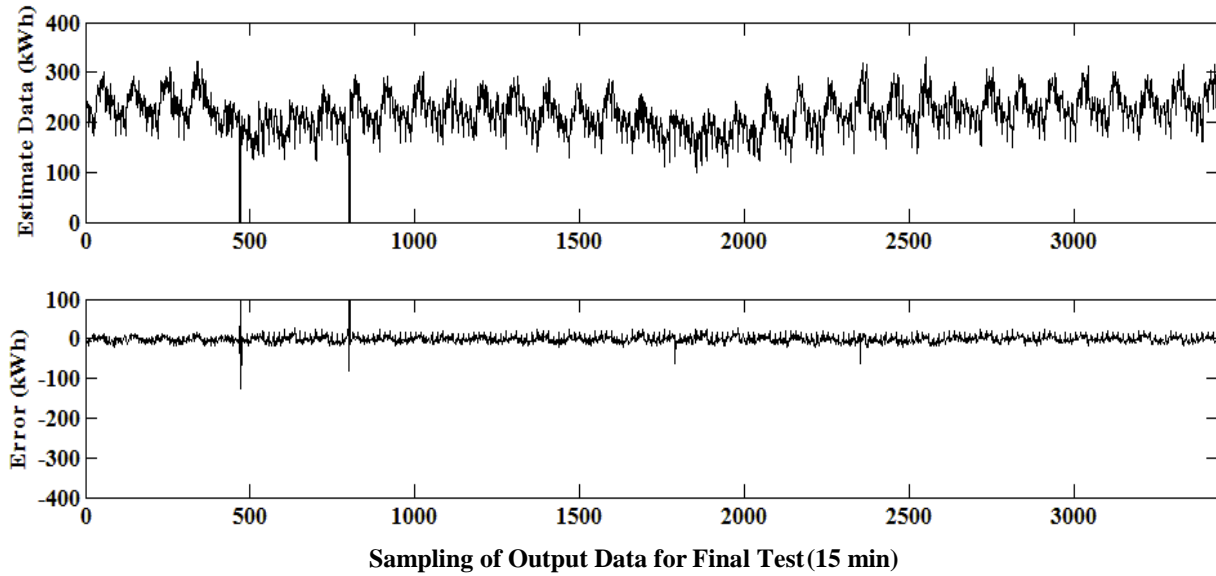
Figure 11. The target data of the testing.

Figure 12. The estimated output data by the proposed PCR for the abnormal voltage of phase a and corresponding error compared with the measured output.
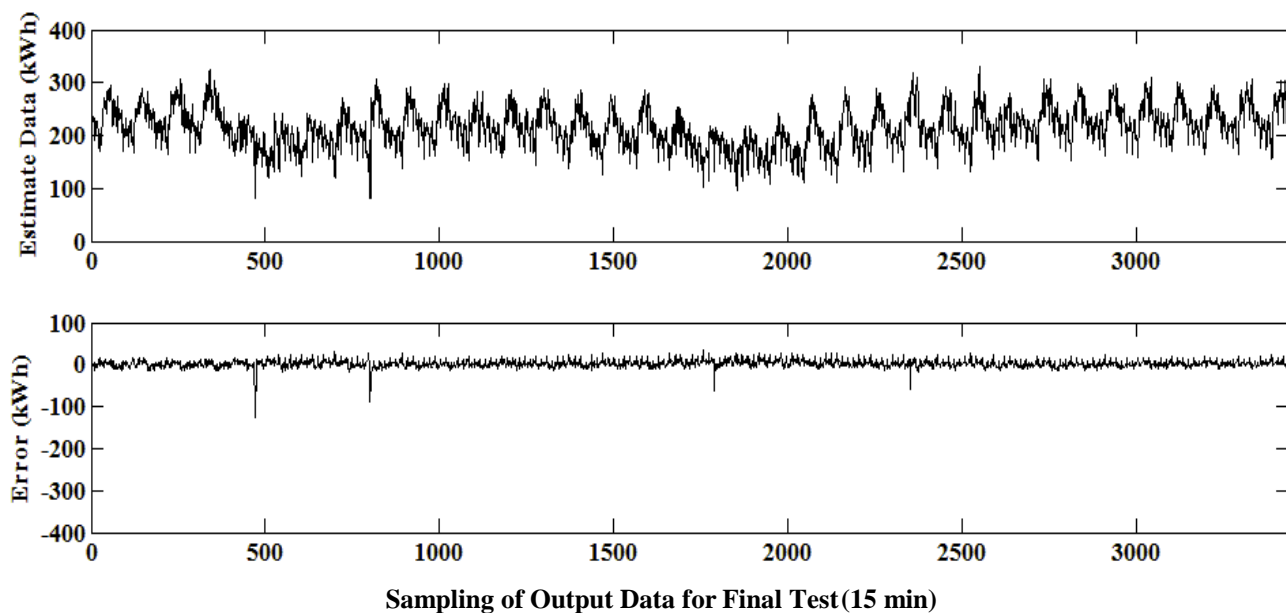

Figure 13. The estimated output data by the proposed PCR for the abnormal voltage of phase b and corresponding error compared with the measured output.
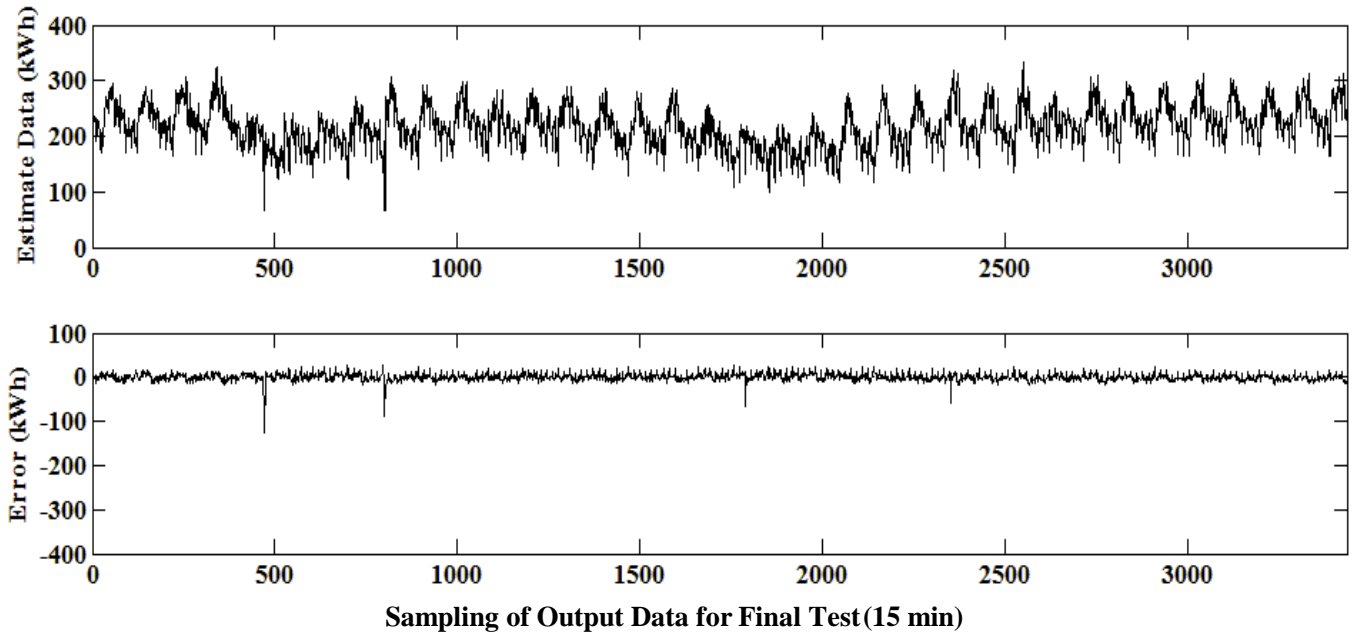
Figure 14. The estimated output data by the proposed PCR for the abnormal voltage of phase c and corresponding error compared with the measured output.
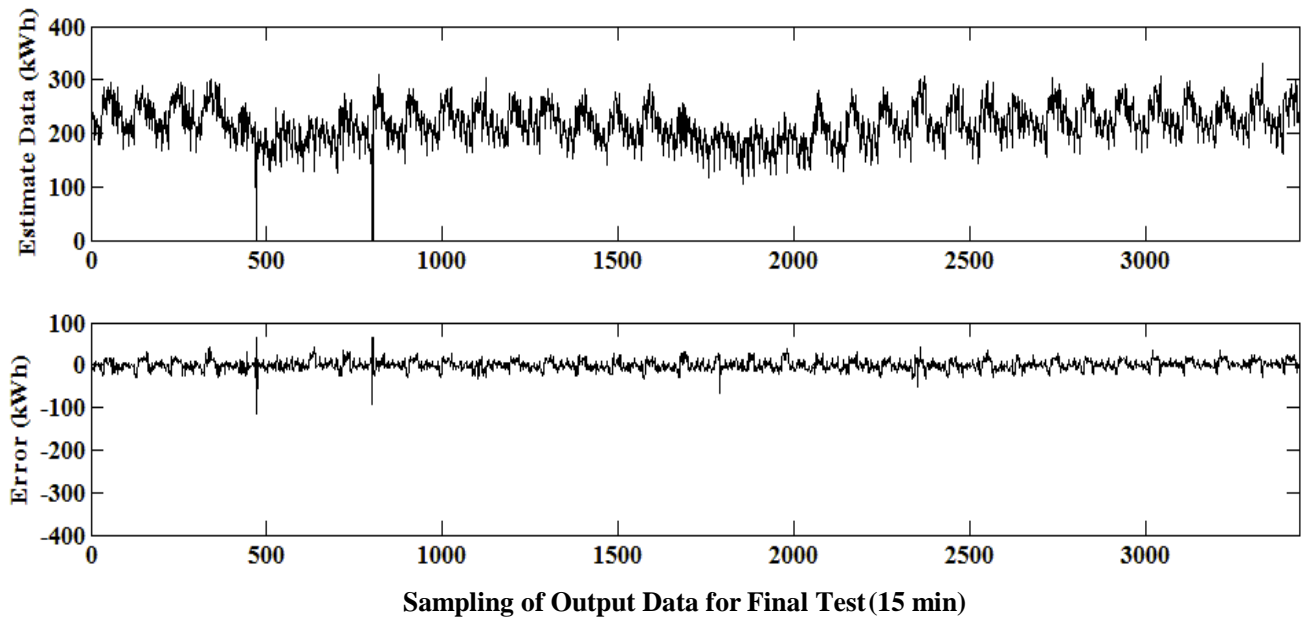

Figure 15. The estimated output data by the proposed PCR for the abnormal current of phase a and corresponding error compared the measured output.

data of any phase for which MAPE is 5.00% in case of the abnormal input current data of phase a, 5.21% in case of the abnormal input current data of phase b, and 9.22% in case of the abnormal input current data of phase c.

For MAPE results by SLR, the best only 1 input current data of the model (which is not the abnormal current) will be used for estimating the output energy data in all cases of the abnormal input current data, for which MAPE is 13.86% in case of the abnormal input current data of phase a, and 10.06% in case of the abnormal input current data of phases b and c.
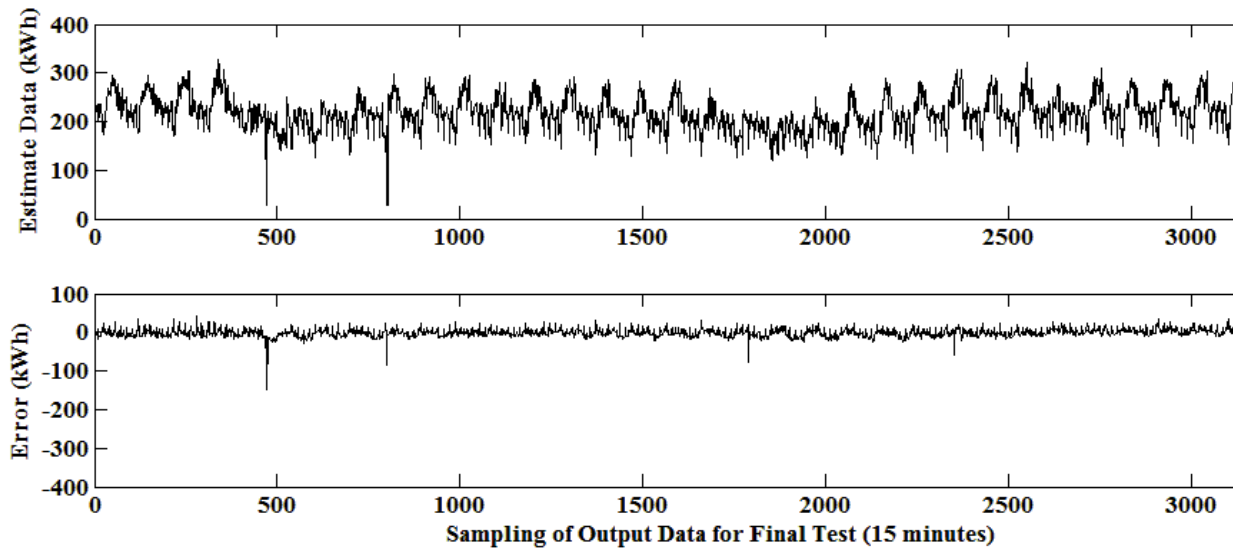
Fig. 16. The estimated output data by the proposed PCR for the abnormal current of phase b and corresponding error compared the measured output.
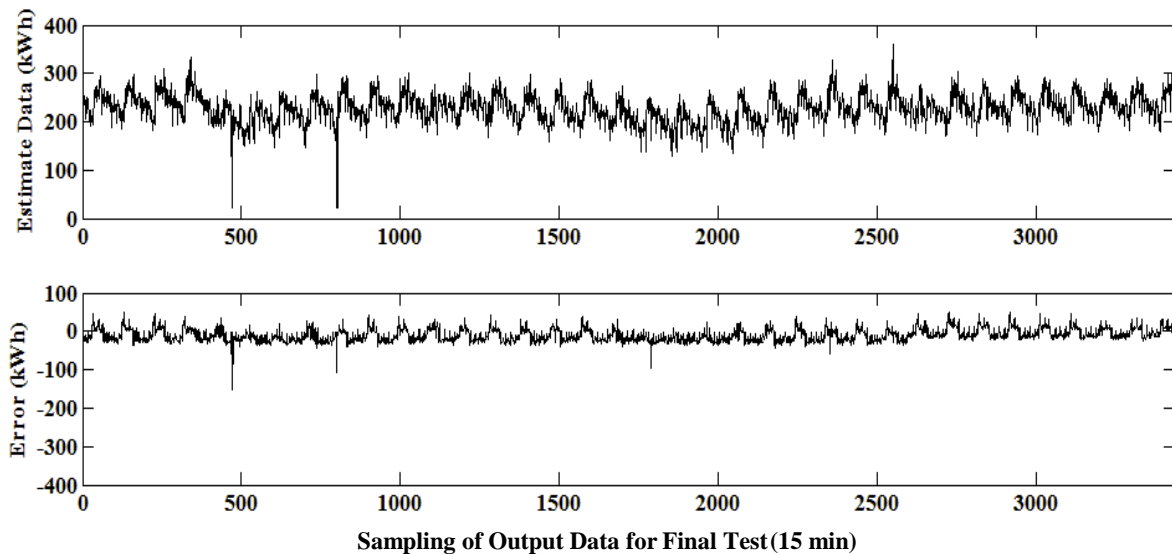

Figure 17. The estimated output data by the proposed PCR for the abnormal current of phase c and corresponding error compared with the measured output.

Table 6. MAPE comparison of the proposed PCR application with MLR and SLR in case of abnormal input voltage data at any phase.

Abnormal variables	MAPE (%)		
	By proposed PCR application (for 5 PCs input)	By MLR application (for 5 input)	By SLR application (for 1 input)
V_a	4.48	4.48	10.06
V_b	3.55	3.55	10.06
V_c	3.23	3.23	10.06

Table 7. MAPE comparison of the proposed PCR application with MLR and SLR in case of abnormal input current data at any phase.

Abnormal variables	MAPE (%)		
	By proposed PCR application (for 3 PCs input)	By MLR application (for 5 input)	By SLR application (for 1 input)
I_a	4.58	5.00	13.86
I_b	3.75	5.21	10.06
I_c	8.03	9.22	10.06

Conclusion

This paper has proposed a new application of PCR in estimation of electrical energy consumption in case of abnormal metering in an AMR system. The error results in this paper show effective performance of the application using the proposed PCR. The simulation results and comparison results between the proposed PCR, MLR and SLR have shown that the MAPE for the proposed application is the best error when compared to that for MLR and SLR of the estimation in case of the abnormal input voltage or current data at any phase. In the difference, MLR is estimated by using 5 normal input data, or SLR is estimated by using the best only 1 input data selection for the output energy estimation.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ACKNOWLEDGEMENTS

REFERENCES

Alahakoon D, Yu X (2016). Smart Electricity Meter Data Intelligence for Future Energy Systems: A Survey. IEEE Transactions on Industrial Informatics,12(1):425-436.
Paris J, Donnal JS, Cox R, Leeb WS (2014). Hunting Cyclic Energy Wasters. IEEE Trans. Smart Grid 5(6):2777-2786.
Stephen B, Isleifsson FR, Galloway S, Burt GM, Bindner HW (2014). Online AMR Domestic Load Profile Characteristic Change Monitor to Support Ancillary Demand Services. IEEE Transactions Smart Grid 5(2):888-895.
Arif A, Hussain MA, Mutairi NA, Ammar EA, Khan Y, Malik N (2013). Experimental study and design of smart energy meter for the smart grid. International Conference RenewABLE Sustainable Energy, pp. 515-520.
Khan TA, Khan AB, Babar M, Taj TA, Ijaz I (2014). Smart meter incorporating UWB technology. IEEE NW Rus. Young Research Electrical Electronic Engineering Conference. pp. 75-78.
Meenual T (2009). The PEA SmartGrids framework for effective distribution. 20th International Conference Electricity Distribution pp. 1-5.
Singh AK, Ibraheem SK, Muazzam M, Chaturvedi DK (2013). An Overview of Electricity Demand Forecasting Techniques. National Conference Emer. Trending Electrical Institute Computer Engineering 3(3):38-48.
Black JD, Henson WL (2014). Hierarchical Load Hindcasting Using Reanalysis Weather. IEEE Trans. Smart Grid 5(1):447-455.
Ladarat W, Naetiladdanon S (2015). PEA Automatic Meter Reading system: Progress and lessons learned. 12th International Conference Electrical Engineering Electronic Computer Telecommunication. Information Technology pp. 1-5.
Erdene BB, Lee B, Kim MY, Ahn TH, Kim D (2013). Extended Smart Meters-Based Remote Detection Method for Illegal Electricity Usage. IET Generation, Transmission and Distribution, 7(11):1332-1343.