**International Journal of Electrical and Computer Engineering (IJECE)**
Vol. 8, No. 3, June 2018, pp. 1324~1330
ISSN: 2088-8708, DOI: 10.11591/ijece.v8i3.pp1324-1330 

# Data Analysis for Solar Energy Generation in a University Microgrid

**Junghoon Lee[1], Seong Baeg Kim[2], Gyung-Leen Park[3]**
[1,3]Dept. of Computer Science and Statistics, Jeju National University, Rep. of Korea
[2]Dept. of Computer Education, Jeju National University, Rep. of Korea

**Article Info**

***Article history:***

Received Dec 21, 2017
Revised Feb 8, 2018
Accepted Mar 1, 2018

***Keyword:***

Big data analysis
Prediction model
Smart grid
Solar energy
Stream orchestration

**ABSTRACT**

This paper presents a data acquisition process for solar energy generation and then analyzes the dynamics of its data stream, mainly employing open software solutions such as Python, MySQL, and R. For the sequence of hourly power generations during the period from January 2016 to March 2017, a variety of queries are issued to obtain the number of valid reports as well as the average, maximum, and total amount of electricity generation in 7 solar panels. The query result on all-time, monthly, and daily basis has found that the panel-by-panel difference is not so significant in a university-scale microgrid, the maximum gap being 7.1% even in the exceptional case. In addition, for the time series of daily energy generations, we develop a neural network-based trace and prediction model. Due to the time lagging effect in forecasting, the average prediction error for the next hours or days reaches 27.6%. The data stream is still being accumulated and the accuracy will be enhanced by more intensive machine learning.



***Corresponding Author:***

Junghoon Lee,
Jeju National University,
Jejudaehakno 102, Rep. of Korea.
+82-64-754-3594
Email: jhlee@jejunu.ac.kr

## 1. INTRODUCTION

Increased environment contamination is one of the most urgent problems we are facing these days. Especially, the industrialization of China makes air quality worse and worse. A great deal of air pollutants come from burning fossil fuels to obtain electricity [1]. Renewable energy, such as wind and sunlight, is the most promising solution to this problem, as they can generate energy without greenhouse gas emissions. However, its intermittent nature prevents itself from being seamlessly integrated into the current energy grid or entirely replacing legacy energy generation mechanisms. Indispensably, the renewable energy integration needs electricity reserve units to cope with the time disparity between energy generation and consumption. Their efficient management is the key not only to blend more renewable energy in our power systems but also to reduce the cost of excessive reserves [2].

A grid can make an energy generation plan according to the forecast on how much renewable energy will be available on the next day or the next few hours in addition to the traditional demand forecast [3], [4]. Generally, the prediction of energy availability can be done based on historical statistics or on relevant spatial and temporal parameters [5]. In the example of solar energy, irradiance will be the most important entity. As for history data analysis, most modern renewable energy generators are able to capture their operation status to report to a central manager or store for further analysis [6]. Those datasets allow us to conduct diverse analysis to better understand the operation of facilities and make a prediction model. Particularly, solar energy generation is deeply dependent on climate conditions. Hence, we can enhance the accuracy of prediction models by the integration of diverse data streams. Here, the prediction model will be

different region by region, making it is necessary to select a best modeling scheme appropriate to the target region and dataset [7].

This paper begins with collecting data streams from multiple solar panels in a microgrid, specifically, Jeju National University, Republic of Korea. Their operation behaviors are traced to develop a prediction model of the amount of solar energy generation for the next hour or day. This approach takes open software solutions for data management, analysis, and visualization. The data stream is stored in MySQL database and a series of queries are designed and issued to this database table. Additionally, the query results, significantly cut down in size, are given to the R statistics package to invoke advanced machine learning APIs and elaborate visualization tools [8]. Specifically, the query results are loaded to the R space either directly via the RMySQL library or by the import command towards the text file downloaded from the MySQL machine. In addition, for the sake of applying a more efficient machine learning library, namely, FANN (Fast Artificial Neural Network), query results are cooked to learning patterns specified by FANN [9].

The rest of this paper is organized as follows: After outlining the paper in Section 1, Section 2 describes the data acquisition process. Section 3 extensively investigates the observation parameters and discusses the result. Finally, Section 4 concludes this paper with a brief introduction of future work.

## 2. DATA ACQUISITION

Figure 1 shows our data acquisition process. After getting an endorsement from the facility management office, our research team obtains the operation records of 7 solar panels over the period from January 2016 to March 2017. The archives are given as Microsoft Excel files. We implement a data parser by Python, which provides comprehensive data interfaces for Excel, Jason, XML, and many others. The parser reads each station record and field one by one to create a series of SQL insert statements. Now, the SQL statements are uploaded to the MySQL machine and executed to insert each record sequentially. In this process, we define a database table containing timestamps and the current amount of energy generation at 7 places. Any queries can be issued to this table via the R package or in the command line interpreter. It must be mentioned that some fields has been corrupted and their values get out of the valid range, that is, the power generation capacity. Those fields will be simply nullified in the database, as missing value interpolation is not our concern [11].

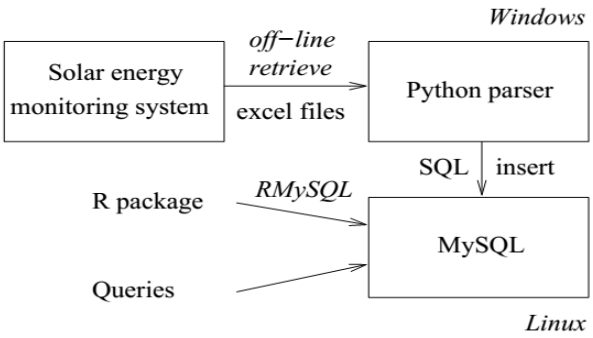
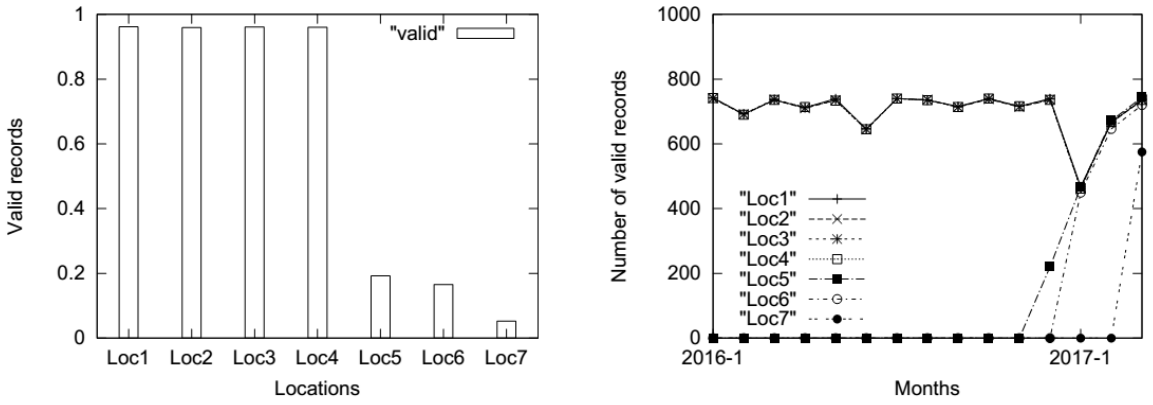



Figure 1. Data acquisition

## 3. DATA ANALYSIS

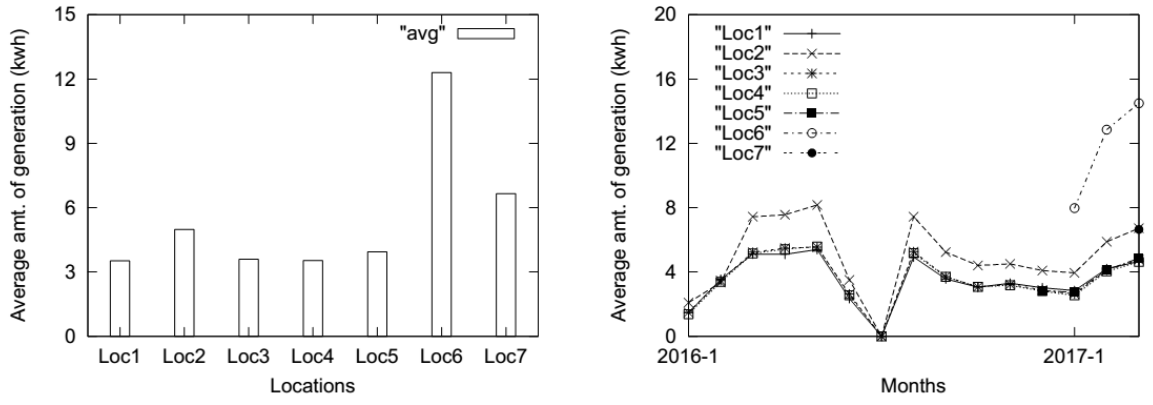
We name each solar panel facility from Loc1 to Loc7, and they are located over the university area. Their power capacities are 30, 40, 30, 30, 30, 30, 90, and 60 *kw*, respectively. Among these, the last three began their operations last December, January, and March, respectively. For them, the amount of electricity generation during the non-working period will be null. Figure 2 plots the valid record ratio since 2016 January. For the first 4 places, ratios are almost 1.0, while the invalidity comes from the intermittent malfunction of the acquisition equipment. On the contrary, for the last 3 places, valid record ratios are quite low, as they have been working just for a few months. Anyway, Figure 2(a) shows the data characteristics in our system. Additionally, Figure 2(b) shows the monthly number of valid records. The number of records becomes nonzero only after a facility begins its operation. The management system has undergone system upgrade, failure remedy, and safety investigation. The number of valid records is most affected by this common factor, while the device-level malfunction is not significant.




(a) Overall validity ratio (b) Monthly validity

Figure 2. Validity ratio statistics

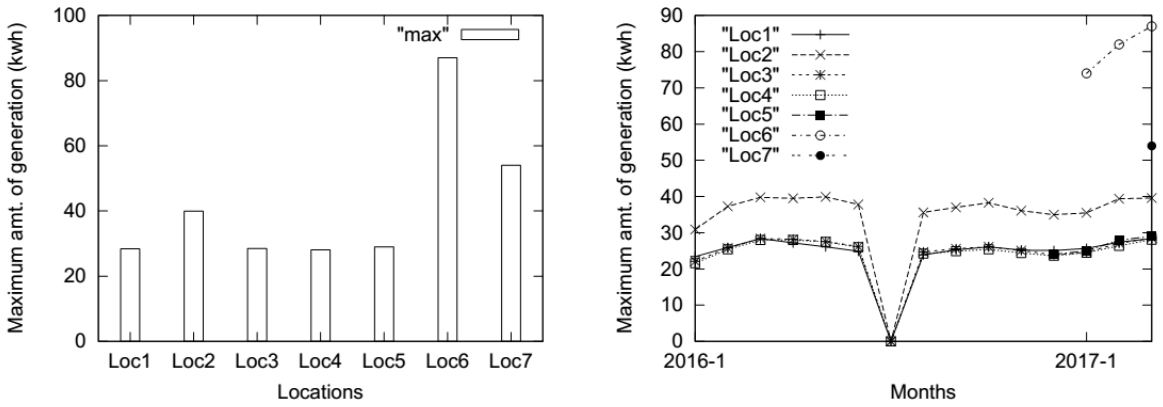
Next, Figure 3 shows the average amount of energy generation for each facility. The value is obtained by averaging the generation amount for each record. Actually, during the night time, solar energy generation cannot take place at all. Hence, the human operator manually turns off the monitoring system from time to time. Those time intervals are excluded in calculating the average. According to Figure 3(a), Loc6, having the generation capacity of 90 *kw*, shows the highest average. We can see that those places having the same capacity show almost the same generation amount. This result indicates that equipment-level difference is quite negligible. In addition, Figure 3(b) shows the monthly average of each facility in a university-scale microgrid. Here again, the monthly behaviors are almost equal for the facilities having the same capacity. The maximum difference is observed to be 7.1% in April 2016. We can see the same pattern in all of the time series.




(a) Overall average amount (b) Monthly average changes

Figure 3. Average amount of energy generation

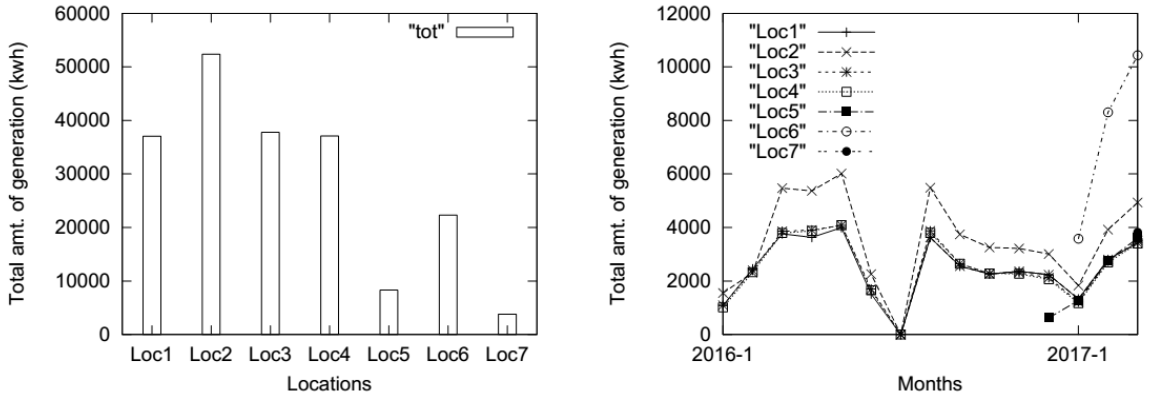
Figure 4 traces the maximum generation amount for all facilities during the whole operation periods. As expected, the maximum amount is limited by the power capacity for each facility. According to Figure 4(a), each generator approaches its full capacity by from 90.0 to 99.7%. The natural worn-out of equipment will worsen this ratio. It will give a guideline on when to replace the equipment. In those days having the best conditions for solar energy generation, that is, high-insolation days, each facility reaches its maximum. In addition, Figure 4(b) shows the monthly maximum generation. In August 2016, the monitoring system has been shut down for some component exchanges. Even during the operation time, records have been hardly valid. Hence, as in the case of Figure 3(b), Figure 4(b) also has a hole in this month.




(a) All-time maximum amount (b) Monthly maximum amount changes

Figure 4. Maximum amount of energy generation

Figure 5 plots the total amount of electricity generation for each facility. As each record contains just the snapshot value at a specific time instant, the exact amount can be different. However, as the solar energy generation does not change sharply in a single day, the add-up of each snapshot value provides sufficiently accurate estimation. During the whole investigation period, the accumulated amount is almost linear to the full capacity of a generator, as shown in Figure 5(a). The last 3 have smaller amount in all-time generation, as they begun working recently. Figure 5(b) shows the monthly amount of energy generated at each solar panel. 7 panels show the similar curves, but the panel having larger capacity seems to drop more sharply. That is, there exist those days in which the energy can be generated between 30 an 90*kw*.




(a) Accumulated amount of generation (b) Monthly statistics

Figure 5. Total amount of energy generation

Now, Figure 6 shows the day-by-day amount of genrration along the whole investigation period for 4 panels working from the beginning. We can see the period in which the monitoring system has stopped working for 45 days. Even though the figure traces the daily total generation, the average and maximum amounts also show the same pattern. The daily difference definitely comes from different climate conditions, particularly, insolation and cloudiness. According to our observation, the solar energy generation lasts at most 11 hours a day. For only a few hours around 1 PM, the generation amount approaches the full capacity. For Loc2 having the full capacity of 40*kw* shows 290*kwh* on the day of best climate condition. With 4 panels, the microgrid obtains up to about 800*kwh* at maximum in a day. It's not so much, compared with the total daily consumption in the university, namely, about 50*Mwh*. However, as more panels (Loc5, Loc6, and Loc7) are installed, the coverage of renewable energy will increase more than double. Up to now, intermittency does not matter, as the solar energy generation takes place during the hour of hot consumption in the university and can be entirely consumed.

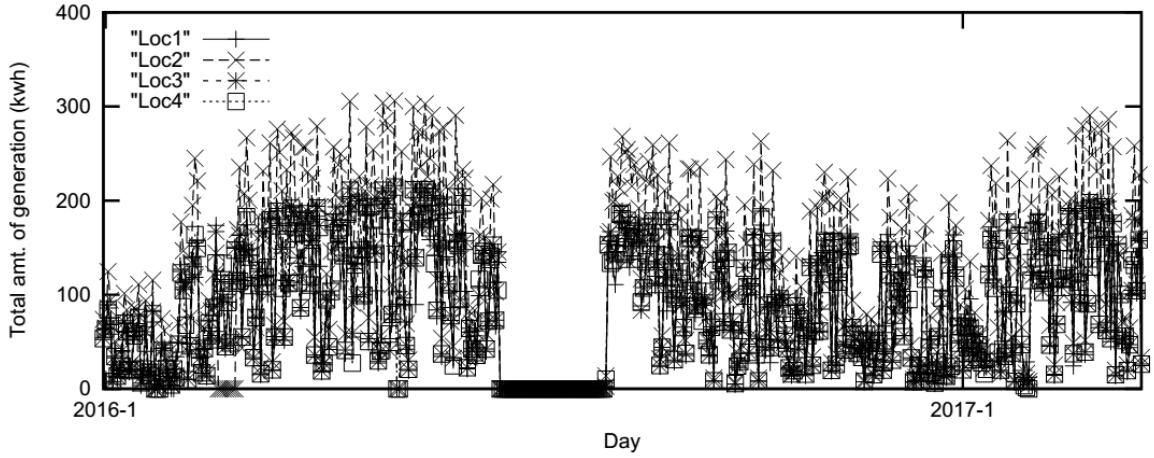



Figure 6. Daily generation pattern

Finally, Figure 7 and Figure 8 build trace and prediction models for Loc1 and Loc2, whose full generation capacities are 30 and 40*kw*, respectively. As the graph will look so dense if we plot from January 2016, these graphs show from 2016-08-08, when the monitoring system restarted after the system upgrade. The prediction model is built exploiting the FANN (Fast Artificial Neural Network) library, which provides comprehensive APIs regarding ANN-based machine learning [9]. Based on the principle of learn by example, the ANN model consists of input, hidden, and output layers as well as the links between them [10]. We think as the climate changes not so instantaneously and the current weather is correlated with previous ones, the generation amount will behave much likely. The sequence of daily generation is converted to a set of learning records. The modeling process takes the generation amounts for the 4 previous days as inputs and the current day generation as output of the neural network. The number of nodes in the hidden layer is empirically selected to be 30.

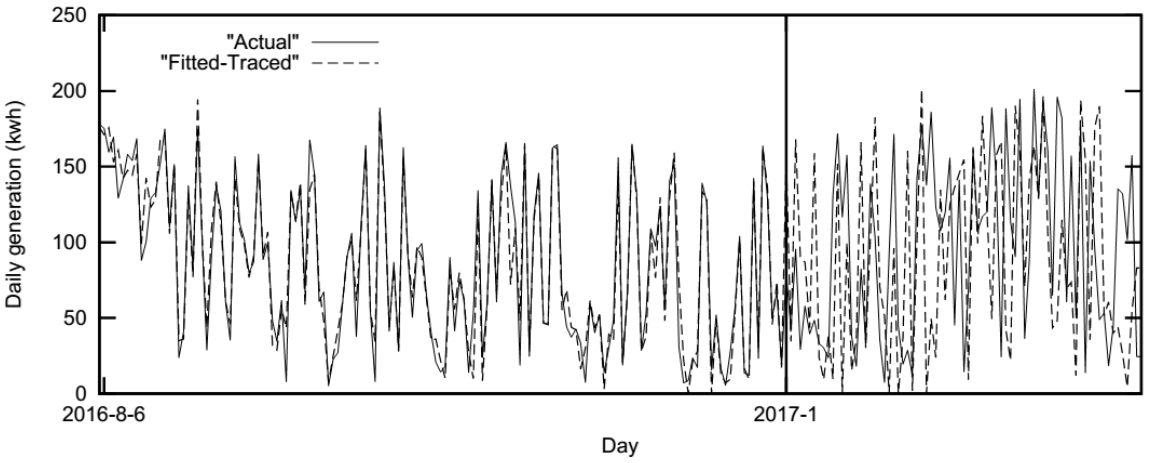



Figure 7. Trace and prediction model for Loc1 (30*kwh* capacity)

Each of Figure 7 and Figure 8 includes a vertical bar on 01-01-17. Daily generations from this day are not used in the learning phase. As can be seen in those figures, the fitting error is quite small and comes from the time lagging effect. The average error is larger in Loc2 having higher capacity. On the contrary, in the prediction part shown in the right-hand side of the vertical bar, the difference between the two curves gets quite severe from time to time. The maximum errors reach 147*kw* and 265*kw* on Loc1 and Loc2, mainly due to time lagging in tracing the changing pattern. The average errors for two places are 26*kw* and 47*kw*, which correspond to 27.6 and 29.6%, respectively. On the first day of a pattern change, the error size gets higher,

for example, the model predicts the maximum amount of generation but the actual generation is the minimum, or vice versa. Anyway, we will integrate the weather forecast to this model and has found the temporal dependency in the data stream of solar energy generation.

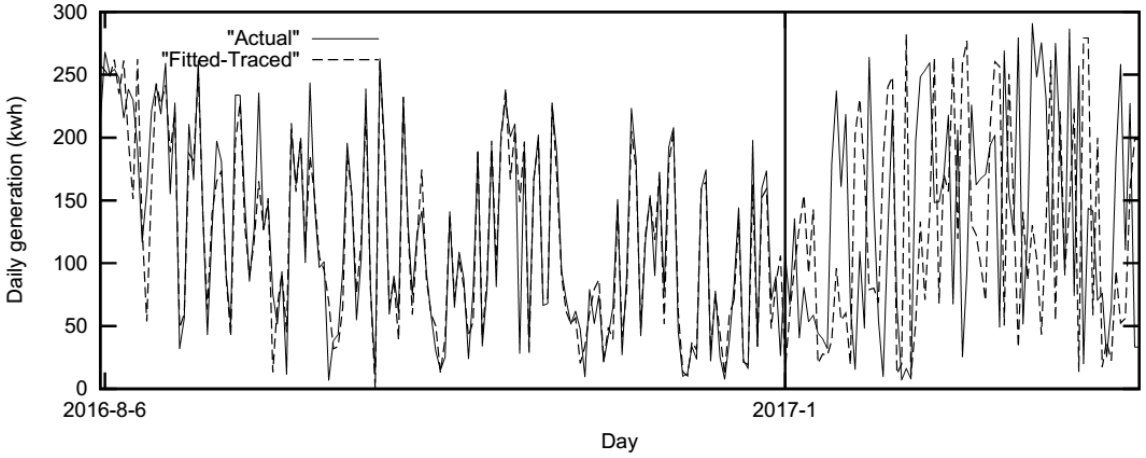



Figure 8. Trace and prediction model for Loc1 (40 *kwh* capacity)

## 4. CONCLUSION

In this paper, we have described out data acquisition process from the management system of solar energy generations. For the hourly generator outputs on 7 panels during the period from January 2016 to March 2017, the average, maximum, and total amounts of generated electricity are analyzed on all-time, monthly, and daily basis. An ANN-based prediction model, built upon the sequential record set, shows average prediction error of 27.6%, mainly due to time lagging, making it necessary to integrate more machine learning process and other parameters.

Currently, we are conducting an analysis on the behavior of EV (Electric Vehicle) chargers, aiming at integrating more renewable energy for EV charging [12]. By combining solar energy generation and EV charging demand, it will be possible to shift the charging demand across the group of chargers belonging to a microgrid having solar energy plants [13], [14].

## ACKNOWLEDGEMENTS

The financial support from Jeju Green Environment Center (JGEC, Korea) is gratefully acknowledged.

## REFERENCES

[1] A. Ipakchi, F. Albuyeh, "Grid of the Future," *IEEE Power & Energy Magazine*, pp. 52-62, 2009.
[2] A. Calazans, M. Kelly, G. Chaudhry, M. Siddiki, "Economic Analysis of a Photovoltaic System Connected to the Grid in Recife, Brazil," in *IEEE Photovoltaic Specialist Conference*, 2015, pp. 1230-1233.
[3] Z. Asad, M. Chaundhry, "A Two-Way Street: Green Big Data Processing for a Greener Smart Grid," *IEEE System Journal*, 2016.
[4] K. Iwamura, H. Tonooka, Y. Miznuo, Y. Mashita, "Big Data Collection and Utilization for Operational Support of Smarter Social Infrastructure," *Hitachi Review*, vol. 63, no. 1, 2014.
[5] S. Haupt, B. Kosovic, "Variable Generation Power Forecasting as a Big Data Problem," *IEEE Transactions on Sustainable Energy*, vol. 8, issue 2, pp. 725-732, 2017.
[6] V. Prasadarao, K. Rao, P. Rao, T. Abishai, "Power Quality Enhancement in Grid Connected PV Systems using High Step Up DC-DC Converter," *International Journal of Electrical and Computer Engineering* vol. 7, no. 2, pp.720-728, April 2017.
[7] B. Kausika, W. Folkerts, W. Sark, B. Siebenga, P. Hermans, "A Big Data Approach to the Solar PV Market: Design and Results of a Pilot in the Netherland," in *European Photovoltaic Solar Energy Conference and Exhibition*, 2013, pp. 4030-4033.
[8] Brunsdon, C., Comber, L. 2015. An Introduction to R for Spatial Analysis & Mapping. SAGE Publication Ltd.
[9] S. Nissen, Neural Network Made Simple. available at http://fann.sourceforge.net/fann en.pdf, Software 2.0, 2005.

[10] B. Benlahbib, F. Bouchafaa, S. Mekhilef, N. Bouarroudj, "Wind Farm Management using Artificial Intelligent Techniques," *International Journal of Electrical and Computer Engineering* vol. 7, no. 3, pp.1133-1144, August 2017.
[11] V. Jyothi, T. Muni, S. Lalitha, An Optimal Energy Management System for PV/Battery Standalone System *International Journal of Electrical and Computer Engineering* vol. 6, no. 6, pp.2538-2544, Dec. 2016.
[12] J. Lee, G. Park, Y. Han, S. Yoo, "Big Data Analysis for an Electric Vehicle Charging Infrastructure using Open Data and Software," in *ACM eEnergy*, 2017, pp.252-253.
[13] J. Lee, G. Park, "Integrated Coordination of Electric Vehicle and Renewable Energy Generation in a Microgrid," *International Journal of Electrical and Computer Engineering* vol. 7, no. 2, pp.706-712, April 2017.
[14] J. Samir, B. Sami, C. Adnane, "Prioritizing Power demand response for Hydrogen PEMFC-Electric Vehicles using Hybrid Energy Storage," *International Journal of Electrical and Computer Engineering* vol. 7, no. 4, pp.1789-1796, August 2017.