International Journal of Control and Automation
Vol. 9, No. 8 (2016), pp.349-358
http//dx.doi.org/10.14257/ijca.2016.9.8.33

# A Prediction Model For Solar Energy Generation Built Upon Status Monitoring

Junghoon Lee[1], Jin-hee Ko[1], Chan Jung Park[2] and Gyung-Leen Park[1,*]

[1]*Dept. of Computer Science and Statistics,*
[2]*Dept. of Computer Education,*
*Jeju National University*
*Jeju-Do, Rep. of Korea, 63243*
*{ jhlee, littletomato, cjpark, glpark}@jejunu.ac.kr*

***Abstract***

*This paper first presents how to create a data stream of solar power generation from the climate archive open to the public as well as the operation logs accumulated for internal use. Then, a daily prediction model is built to forecast the amount of the future electricity generation according to the weather parameters such as wind speed, temperature, insolation, and sunshine hours in Jeju City. For the regression model built upon the identification of linear dependency of most parameters to the generated solar energy, its fitting accuracy is evaluated in terms of absolute residuals, standardized residuals, Cook's distance, and error probability distribution. The prediction result shows that the absolute residual stays below 30 kwh and the average at 15.8 kwh, while maximum generation reaching 148 kwh. The prediction model will not only evolve with more record collections, possibly averaging out the effect of some abnormal points, but also make it possible for consumers like electric vehicles to select an energy source out of solar, wind, and legacy grid-providing energies.*

***Keywords****: solar generation, climate effect, log file processing, linear regression, prediction*

## 1. Introduction

Due to the environment problem we are facing, renewable energy is drawing more and more attention all over the world and many cities are constructing power generation facilities targeting at sunlight, wind, and the like. However, their main drawback lies in intrinsic dependency on climate conditions which can never be perfectly controlled by humans. Hence, their deployment must cope with the time disparity in energy generation and consumption. At this stage, a promising solution is to predict the availability of future energy sources and adapt the consumption as much as possible [1]. Meanwhile, modern renewable energy plants commonly provide real time monitoring and status logging for more efficient operation, administration, and management [2]. Besides, the massive volume of history data can help us to develop an efficient prediction model.

Moreover, a fundamentally new type consumer is currently penetrating into our daily lives. It is an EV (Electric Vehicle) [3]. Its battery is charged before driving, that is, actual energy consumption, possibly alleviating the above-mentioned time disparity as well as absorbing the energy overproduced from the uncontrollable renewable sources [4]. Basically, an EV can select the energy to charge its battery, specifically, from the legacy

* Prof. Gyung-Leen Park is the corresponding author.

This research was supported by Korea Electric Power Corporation through Korea Electrical Engineering & Science Research Institute. (Grant number: R15XA03-62)

ISSN: 2005-4297 IJCA

grid, or any of renewable energy sources. For an EV to decide from which to charge, accurate prediction is most important, while the prediction is carried out based on a bunch of past data, as can be seen in most data mining strategies [5]. After all, the penetration of EVs, which is witnessed these days, will contribute to the extensive employment of renewable energy. Here, Jeju City is extending the occupation of solar energies and prompting the deployment of EVs under its ambitious plan of Carbon-Free Island 2030. This area is planning to replace all gasoline-powered vehicles with EVs by 2030.

In most countries, the weather bureau opens the past climate records to the public while almost every power generation facility stores its operation records in log files according to its own file format. The public climate records are very well-structured and it is not difficult to read and interpret them. In addition, we can also parse the log file to get the hourly or daily amount of power generation and combine with the climate records. Moreover, there are diverse high-quality and almost free analysis software tools capable of building a comprehensive prediction model [6]. In this regard, this paper presents how to process the solar energy monitoring records, builds a prediction model with linear regression, and assesses its performance.

This paper is organized as follows: After outlining the main issue of this paper in Section 1, Section 2 reviews some related work and our previous research. Then, Section 3 describes how to process the log file and climate archives. Section 4 builds a prediction model for the future solar energy generation, and Section 5 evaluates the performance of the prediction model. Finally, Section 6 summarizes and concludes this paper with a brief introduction of future work.

## 2. Related Work

As for some related work, a monitoring function is embedded in energy generation systems, not just in a solar panel but also even on the roof of a fast-moving EV, mainly for the detection of abnormal states such as rapid variation and electrical discontinuity [7]. It also keeps generating monitoring records from a set of key inspection points in a generation system, and these records are not discarded but stored for a further analysis. Here, the effect of each climate factor is different region by region, and thus it is necessary to identify critical factors and find an appropriate prediction model [8]. Basically, during the operation of the solar system, it is possible to decide whether to consume the stored energy now or save for later use according to the climate parameters [9].

Our research team has conducted an analysis and developed prediction models regarding the climate effects on renewable energy generation in Jeju City, Republic of Korea. First, in [10], wind speed records are converted to a series of learning patterns and fed into an artificial neural network capable of efficiently fitting the nonlinear behaviors of wind speed changes. This model predicts the wind speed on hourly, daily, and monthly basis to estimate the amount of electricity from regional wind turbines for the sake of better integration of wind energy to the main grid. This model is extended, namely, making three different models predict independently and taking the majority of them as the final prediction [11]. Those models are also autonomous neural networks learned by hourly, daily, and seasonal records. This vote-based scheme can eliminate abnormal mispredictions unexpectedly taking place in a single prediction model.

In addition, [12] examines the correlation between wind speed and sunlight in 4 major regions in Jeju City, taking into multiple renewable energy sources into account simultaneously. The authors have extensively considered the regional effect and other factors such as rainfall and underwater level dynamics. According to the analysis, the negative correlation between them justifies the development of a compensatory generation plan for each source. The correlation level is different in different regions, but it is still possible to build a common strategy with minor weight tuning.

 

## 3. Data Management

To begin with, from the KMA (Korean Meteorological Administration) web site, the hourly archive of climate records is downloaded. As the files have a common well-defined format, we can develop a computer program capable of translating each record to an SQL statement and inserting the recorded information into the predefined database table. It contains all climate columns provided on the web site such as cloudiness, air pressure, and the like, not just those fields necessary in this paper. In addition, the log file from the solar power facility does not conform to a regular grammar, while not only a single polling record is interrupted arbitrarily but also timestamps break in those locations. Our parser implementation first detects the exact boundary between two adjacent records and checks validity of a record via the checksum field. Next, after removing the redundant timestamps in a polling record, we can obtain the hour-by-hour amount of power generation. Here again, all fields obtained from the operation log are stored, including voltage, current, frequency, and system status.

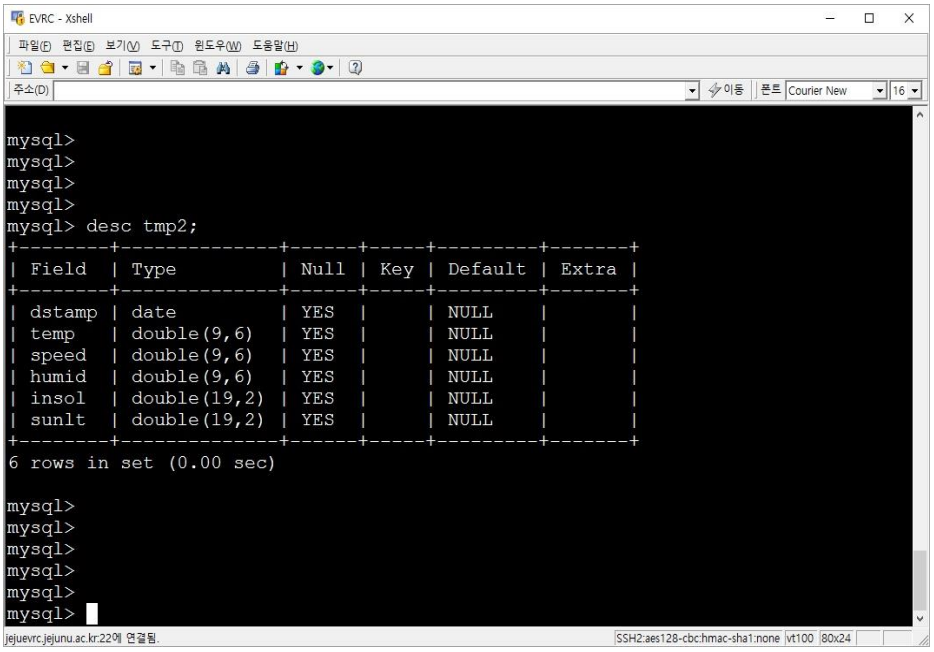


**Figure 1. Database Table Specification**

To retrieve the daily generation amount, the intact earliest and latest records of power generation are most important. After some SQL grouping day-by-day tuples and joining two tables by the common timestamp, we get a new table consisting of daily power generation, temperature, humidity, insolation, and sunlight duration as shown in Figure 1. Currently, the post *id* is not included, as we are focusing on a single facility, however, with the addition of multiple facilities over the whole city area, post *id* and geographic location fields will be included.

Our analysis framework employs MySQL, one of the most widely used open software database systems [13]. Currently, 292 daily records from the time period beginning from 2015-05-08 are inserted in the table, and doubtlessly the number of records will keep growing enough to accommodate a big data processing framework [14]. The database can interface a variety of other applications such as R, which provides a rich set of sophisticated data visualization and state-of-the-art analysis tools [15]. For example, Figure 2 is the result of 3-dimensional plotting for the power generation according to the sunshine hours and insolation, the two most important parameters for the amount of solar power. As 292 daily records are accumulated, the graph has as many points as the number of the records. We can see highly dense distribution near the point (0, 0, 0), in which very little insolation is observed and solar power is barely generated. Those three elements seem to have a strong linear dependency with each other, and the effect of sunshine hours and insolation to the power generation can be straightforwardly traced by a linear regression model, even though a more sophisticated methods such as neural networks is also available [16].

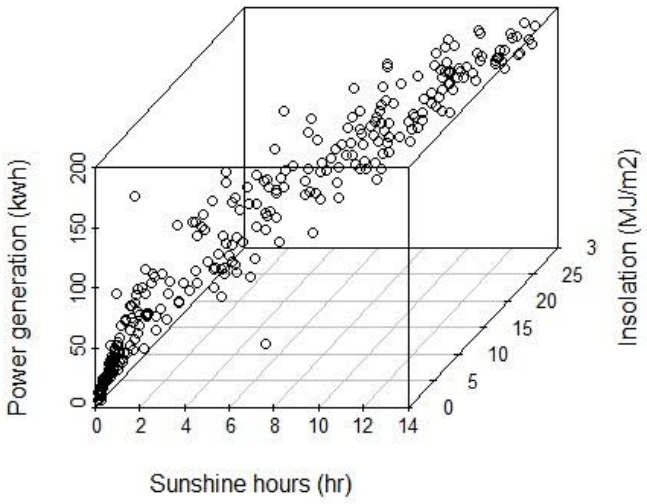


**Figure 2. Effect Of Solar Parameters**

## 3. Linear Regression Model

Our model begins with the identification of critical climate parameters for the solar energy generation. According to the correlation test conducted in the R workspace, the covariance and correlation values of wind speed, temperature, sunshine hours, and insolation to the amount of generated electricity are shown in Table 1. Wind speed is correlated with the generation not so much and negatively. However, it cannot be negligible. Temperature is positively and sufficiently correlated enough to be picked for the prediction model. The others have a very strong correlation, approaching 1.0, to solar energy generation. Actually, in our preliminary analysis, the behavior of the amount of solar energy generation does not seem to be a time series, as a value is not dependent on the set of previous values in Jeju City and the solar energy is inherently discontinuous during the night time.

 

**Table 1. Correlation Factors**

| | covariance | correlation |
|---|---|---|
| wind speed | -17.99 | -0.22 |
| temperature | 256.01 | 0.57 |
| insolation | 485.14 | 0.93 |
| sunshine hours | 237.76 | 0.96 |

As generally known, the linear regression method fits a straight line through the set of points, regardless of dimensions, or the number of variables, in such a way that makes the sum of squared residuals of the model as small as possible. Here, residuals are the differences between the actual and fitted values. This method is usually built upon the least square method and makes a prediction with an arithmetic calculation involved in the evaluation of a linear expression. Almost every statistics package basically supports this fitting process. We need to just invoke the regression API after packing a set of tuples consisting of the above-mentioned climate fields into a data frame. The direct comparison of actual and fitted values is shown in Figure 3. We can see only 3 significant outliers in this figure and most points are lined up around the diagonal which corresponds to the perfect estimation.

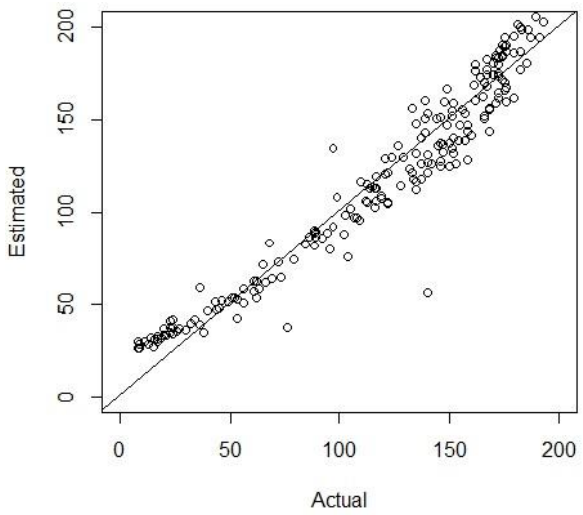


**Figure 3. Comparison of Actual and Fitted Values**

The validity test results of the model are shown from Figure 4 to Figure 7. First, Figure 4 shows the residual errors plotted versus their fitted values. It is desirable that the residuals randomly distribute around the horizontal (dotted in this Figure) line corresponding to zero residual error. It means that residuals are not correlated and thus there should not be a distinct trend in the distribution of residuals. Residuals are slightly more likely to be negative, especially when the solar energy generation is less than 50 *kwh* or larger than 150 *kwh*. However, it can be considered to be large a random distribution, considering that fitted values tend to get smoother.

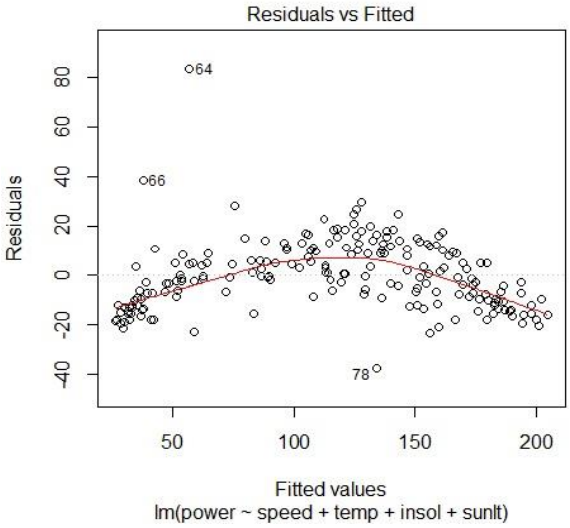


**Figure 4. Residual Distribution**

In addition, Figure 5 shows the absolute values of the standardized residuals as a function of fitted values. Again, residual are required to be uncorrelated and are normally distributed. Just one point, numbered 64, has a value higher than 2.0 and the other points are not located far away from the mean. The standardized residual is the residual divided by standard deviation and it normalizes the given error distribution. Those points around the solid curve contribute the fitting model.

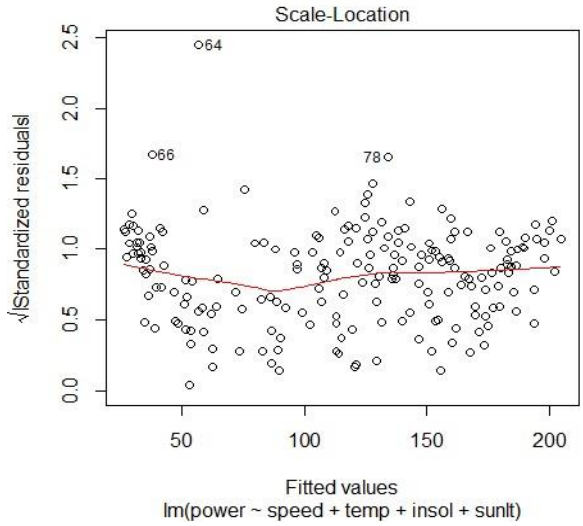


**Figure 5. Standardized Residuals**

Figure 6 shows the leverage of each point, to display a measure of its importance in determining the regression result. The Cook's distances of 3 points have quite larger values than the others, but much less than 1.0, above which careful checking is required. The distance measures the effect of deleting a given point to the regression model.

 

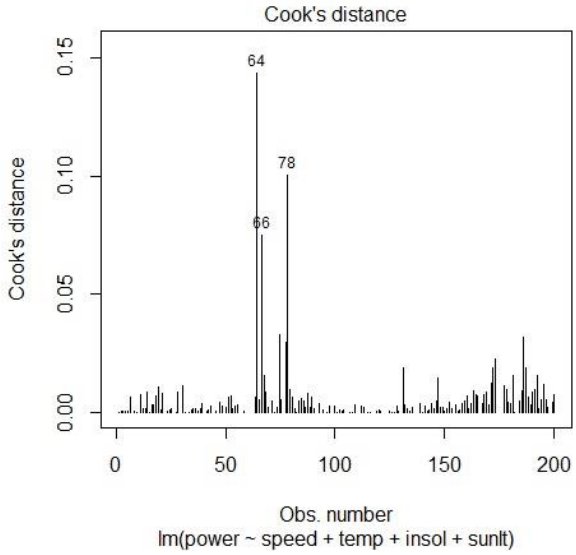


**Figure 6. Cook's Distance**

Next, Figure 7 shows the histogram of residuals to decide whether residuals distribute normally. It is also generated by an R function. Here, each box represents a 2 *kwh* interval in the x-axis and the y-axis shows the number of records in the set of 200 total records. As can be seen in this figure, most points are centered around 0 while the shape is largely symmetric in both sides, provided that we exclude the rightmost extreme point.

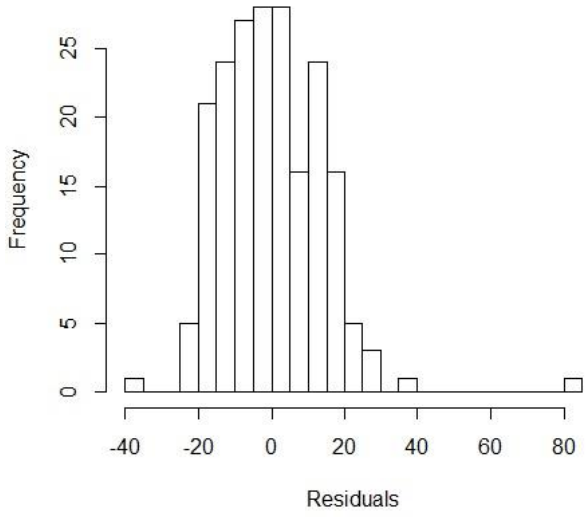


**Figure 7. Histogram Analysis**

## 4. Prediction

To evaluate the prediction accuracy, Figure 8 plots the forecast results for 92 records which have not been used in the model construction. For this set, the maximum amount of energy genertation is 148 *kwh*. The gap between an actual value and its predicted value looks larger compared with the fitting case shown in Figure 3 as in the case of other prediction approaches. Here, the maximum difference reaches 77.97 *kwh*. Those points are mainly found when the actual value is near 0. However, if we exclude those two points having an extraordinarily large error, the absolute residual stays below 30. As can be seen in the figure, when the amount of power generation is less than 75 *kwh*, our model tends to overestimate while underestimate in the other case due to the inherent characteristics of linear regression. Here, more points are located in the area of small solar power generation, mainly due to seasonal effect. With more records, we can develop a better prediction model. With this model, we can estimate the amount of solar energy given that a weather forecast is available [9].

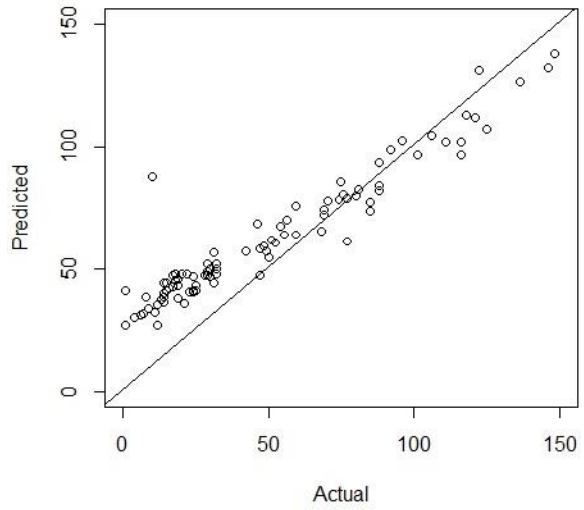


**Figure 8. Prediction Results**

## 5. Conclusions

Renewable energy generation keeps extending its coverage in our daily lives, but due to the unpredictable availability stemmed from the severe dependency on the climate condition, a forecast model of reasonable quality is getting more important. In this paper, we have filtered the essential fields from the operation log files of a solar energy facility as well as the climate archives in Jeju City. Then, a prediction model is developed based on the linear regression method to forecast the amount of daily solar energy generation according to wind speed, temperature, insolation, and sunshine duration. Here, 200 records are used for training while 92 for evaluating the accuracy. The proposed model has achieved the accuracy level, which bounds the absolute residual below 30 *kwh* and makes average error equal to 15.8 *kwh* for the set whose maximum reaches 148 *kwh*. This result indicates that solar energy generation can be forecasted by means of a relatively straightforward and time-efficient method, potentially coping with the ever-growing number of history records.

As future work, we are planning to develop a city-wide prediction model combining spatial information in Jeju City to make is possible for EVs to build an efficient charging plan and to find a tour schedule including battery charging with renewable energies [17]. Here, integrative management of renewable energy generation, legacy power grid, and

 

EV fleets should be systematically organized in this city, along with various EV-related applications. Information technology-base sophisticated algorithms will enrich the service of smart grids and electric vehicles, providing comfortable energy lives [18].

## References

[1] W. Yäici and E. Entchev, "Prediction of the performance of a solar thermal energy system using adaptive neuro-fuzzy inference system", 3rd International Conference on Renewable Energy Research and Applications, **(2014)**, pp. 601-604.
[2] C. Schuss, B. Eichberger and T. Rahkonen, "A monitoring system for the use of solar energy in electric and hybrid electric vehicles", IEEE International Instrumentation and Measurement Technology Conference, **(2012)**, pp. 524-527.
[3] J. Lee, G.-L. Park, Y. Cho, S. Kim and J. Jung, "Spatio-temporal analysis of state-of-charge streams for electric vehicles", 4th ACM/IEEE International Conference on Information Processing in Sensor Networks, **(2015)**, pp. 368-369.
[4] I. Bayram, M. Shakir, M. Abdallah and K. Qaraqe, "A survey on energy trading in smart grid", IEEE Global Conference on Signal and Information Processing, **(2014)**, pp. 258-262.
[5] I. Goiri, K. Le, T. Nguyen, J. Guitart, J. Torres and R. Bianchini, "Green Hadoop: Leveraging green energy in data processing frameworks", Proceedings of Eurosys, **(2012)**.
[6] Y. Zhao, "R and Data Mining: Examples and Case Studies", Elsevier Inc., **(2013)**.
[7] P. Visconti and G. Cavalera, "Intelligent system for monitoring and control of photo-voltaic plants for optimization of solar energy production", IEEE 15th International Conference on Environment and Electrical Engineering, pp. 1933-1938, **(2015)**.
[8] K. Passow, L. Ngan, B. Littmann, M. Lee and A. Panchula, "Accuracy of energy assessment in utility scale PV power plant using PlantPredict", IEEE 42nd Photovoltaic Specialist Conference, **(2015)**.
[9] M. Detyniecki, C. Marsala, A. Krishnan and M. Siegel, "Weather-based solar energy prediction", IEEE World Congress on Computational Intelligence, **(2012)**.
[10] J. Lee, G. Park, Y. Kim, E. Kim and I. Lee, "Wind speed modeling based on artificial neural networks for Jeju area", International Journal of Control and Automation, vol. 5, no. 2, **(2012)**, pp. 73-80.
[11] J. Lee and G. Park, "Vote-based wind speed forecast scheme built upon different artificial neural network models", International Journal of Control and Automation, vol. 9, no. 5, **(2016)**, pp. 99-110.
[12] C. Park, J. Lee, S. Kim, J. Hyun and G. Park, "Spatio-temporal correlation analysis of wind speed and sunlight for prediction power generation in Jeju", International Journal of Multimedia and Ubiquitous Engineering, vol. 8, no. 4, **(2013)**, pp. 273-282, http://www.mysql.com.
[13] H. Rahimi-Eichi and M. Chow, "Big-data framework for electric vehicle range estimation", 40th Annual Conference of the IEEE Industrial Electronics Society, **(2014)**, pp.5628-5634.
[14] C. Brunsdon and L. Comber, "An Introduction to R for Spatial Analysis & Mapping", SAGE Publication Ltd., **(2015)**.
[15] W. Yäici, E. Entchev, M. Longo, M. Brenna and F. Foiadelli, "Artificial neural network modeling for performance prediction of solar energy system", 4th International Conference on Renewable Energy Research and Applications, **(2015)**, pp. 1147-1151.
[16] B. Bhattarai, M. Levesque, M. Maier, B. Bak-Jensen and J. Pllai, "Optimizing electric vehicle coordination over a heterogeneous mesh network in scaled-down smart grid testbed", IEEE Transactions on Smart Grid, vol. 6, **(2015)**, pp. 784-794.
[17] S. Ramchrum, R. Vytelingum, A. Rogers and N. Jenning, "Putting the 'smarts' into the smart grid: A grand challenge for artificial intelligence", Communication of the ACM, vol. 55, no. 4, **(2012)**, pp. 86-97.

 