Correlation Study between Photovoltaic Power Output and Environmental Variables Using an Embedded IoT System

Jason Y. Corona-Ventura, Oscar Lobato-Nostroza, Gerardo M. Chávez-Campos, Rafael Lara-Hernández, Yvo M. Chiariada-Masseli*, Adriana C. Téllez-Anguiano and Miguelangel Fraga-Aguilar
Tecnológico Nacional de México / IT Morelia, Morelia, Michoacán, México
*Instituto Nacional de Telecomunicações, Santa Rita, Brazil
gmarx_cc@itmorelia.edu.mx

***Abstract*—The sunlight is one of the most important renewable sources. However, the systems designed for its usage are not a continuous power delivering option. A photovoltaic panel is the most common collecting device for solar energy, regardless of having low efficiency at high temperatures. The present paper shows the pre-processing stage for data analysis to study the relationship between the sunlight and weather variables, with the power output of a photovoltaic panel for further estimation models. In order to obtain representative and enough data, an embedded IoT system was proposed to collect and store measurements. Then, data were analyzed by using scatterplots matrix and a non-parametric correlation calculation. The correlation factors probe that data can be classified as fuzzy data with a high confidence level to be used for model developing.**

I. Introduction

Nowadays the sunlight is one of the most important renewable sources due to its major presence compared with others [1]. Mexico has a daily solar radiation between $3kWh/m^2$ and $5kWh/m^2$ on average [2], [3] this radiation could be exploited by enhancing its harvesting infrastructure. These improvements would lead to social benefits helping to reduce the carbon footprint [4], [5], [6].

A photovoltaic (PV) panel is the most common collecting device for solar energy. This panel allows converting solar radiation into electric energy by using the photovoltaic effect. Nevertheless, this effect is disturbed by weather parameters [7], [8], [9], [10], sunlight incidence angle [11] and panel's temperature [12], [13]. Furthermore, the panel's temperature has been reported to be the most critical factor because of its semiconductors nature [12], [13]; a semiconductor modifies its conductivity, either by adding impurity or because of temperature effects, making the semiconductors temperature-depending electronic devices.

Thus, the PV panel delivers energy intermittently due to the previously described factors [14]; this behavior produces none desired panel's output on stand-alone or array applications [15]. For that reason, several research works had been modeled PV panels to study and estimate how these factors affect the panel's energy output. In [16], a fast and straightforward model was developed based only on the panel's technical data, and its output is not affected by weather or temperature factors. More accurate models have been developed using Neural Networks (NN). These models can use as data input weather, light, and temperature measurements. Nonetheless, to enhance the NN's output, it is well known that long term data is necessary [17], [18]. Unfortunately, the long term data measurements for PV panels are challenging to obtain. Consequently, most of the models have been obtained using non-real nor *in-situ* measurements. In this sense, in [19] and [20], commercial log systems were used to monitoring specific applications, these systems have been based on hardware and software for autonomous data gathering and management.

Hence, a data-logger system had been preferred to study and model the system's behavior as a function of its external parameters, and considering the high increase of smart devices, the IoT-systems have become the most suitable monitoring solution [19], [20]. These IoT-devices work over networks to collect and transfer data for future analysis, allowing bidirectional interaction and real-time responses possibilities [21], [22], [23].

The present paper introduces an IoT-based system for measuring *in-situ* the sunlight intensity levels (Lx), humidity (%), temperature (°C), and the energy output (W) of a photovoltaic panel. Data were collected during five months since October 2018 until February 2019. This data will be used to eventually develop a model for short-term forecasting of the power produced by the panel. In this first stage, the data collected were used to study the correlation between the measured variables: sunlight, temperature, humidity, and the solar panel's power by using the Spearman non-parametric correlation method.

II. Material and Methods

The current research required sampling data *in-situ* from a solar panel exposed to the real weather in order to perform an observational study. The solar panel was mounted in a two-level building at the roof, together with an IoT system designed and build for this specific purpose.

The IoT-system's hardware, flow programming, and user interaction is presented in II-A, II-B, II-C subsections, respectively. Then, subsection II-D covers the IoT-system's installation, subsection II-E explains the data collected by the

978-1-7281-2898-6/19/$31.00

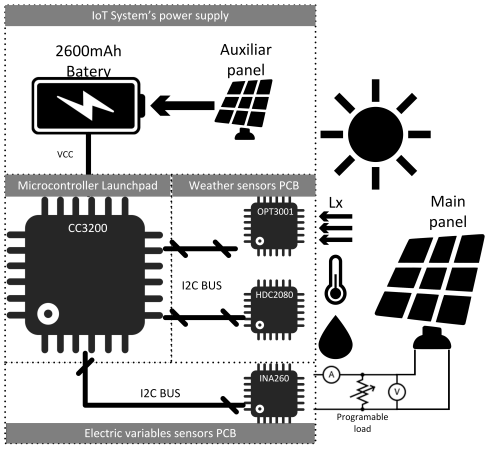

Fig. 1. Proposed IoT monitoring system.

system, and subsection II-F explains the correlation methods used for this study.

A. Hardware

A general schematic of the proposed IoT system is shown in Fig. 1. The system was based on the CC3200 Launchpad from Texas Instruments [24]. The launchpad was integrated with three digital sensors, a 2600mAh battery, and a small solar panel to charge the system.

The CC3200 microcontroller (µC) has embedded protocols including Wi-Fi and HTTP, enabling remote communication and I^2C to interact with the sensors. The µC also offers low power modes and serial flash storage.

The sensors INA260, OPT3001, and HDC2080 have high precision measurements with a typical accuracy of 99.8% and 14-bits resolution. The INA260 can sense up to $15\,\mathrm{A}$ and $40\,\mathrm{V}$, considering that the panel's maximum output is $16\,\mathrm{V}$ and $0.600\,\mathrm{A}$. On the other hand, the OPT3001 was improved to detect an extended range up to 120,000 Lx by using a glass with a 0.55 attenuation factor. The power supply was composed by a 2600mAh rechargeable battery and a USB charger adapted to receive the power from an auxiliary panel attached to the system. This auxiliar panel produces a maximum output of $300\,\mathrm{mA}$.

B. Embedded Software

A FreeRTOS Real-Time Operating System (`RTOS`) was used for task handling and modular programming [25]. The system's behavior was controlled by simultaneous tasks execution regarding priority and semaphores.

The code on the system enables the Wi-Fi and HTTP subsystems to create an access point (AP) and offers an HTML page to interact with the final user. Eventually, the system tries to connect using the collected configuration information, and if the connection is successful, the I^2C subsystems and Low Power Mode (LPM) will be initialized.

Consequently, the system enters into a measurements loop by enabling LPM and idle modes. The acquisition time was controlled by a $32.768\,\mathrm{kHz}$ Real-Time Clock (RTC). This RTC allows logging data with a full date format.

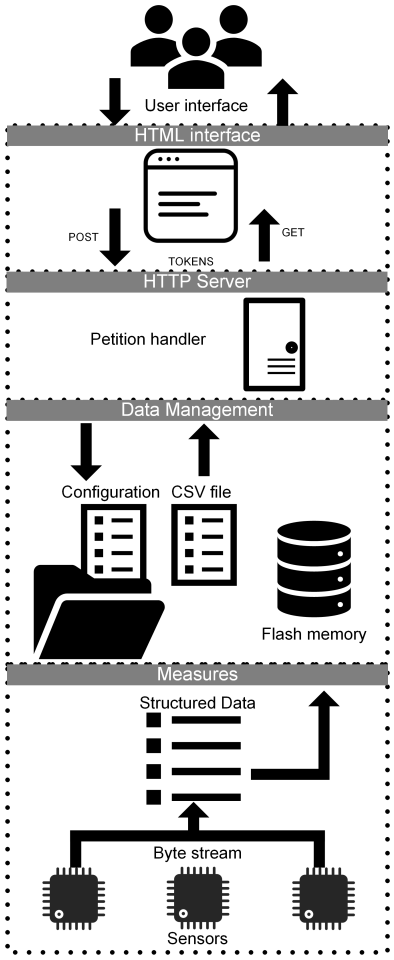

Fig. 2. User interface communication hierarchy.

Each measure was temporally saved on RAM variables, and when data was required, it is delivered as an array separated by commas to indicate columns and lines feed for rows. A `*.csv` file was selected as a container, considering its compatibility with most of the data analysis software. The data files were created once the user requests it or after 288-row entrances.

C. User Interaction

The IoT-system can interact with the final user through a HTML-interface based on the `GET` and `POST` instruction tokens. The tokens were used in the IoT-system together with an HTTP server to implement a petition handler as the main interaction block. The Fig. 2 shows how the HTTP server gets the structured data from sensors to create the `CSV` file by reading the flash memory. Besides the HTTP server can access to configure the IoT-system's Wi-Fi parameters and sampling time; see Fig. 3. When the system is online, a data management web-page shows the actual measurements and allows to download logs or the previous daily stored backup files.

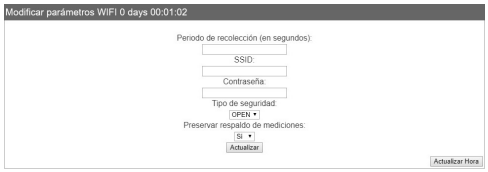

Fig. 3. Wi-Fi station mode configurable settings.

Fig. 4. Available data file download.

D. The IoT-System

The previous hardware, software, and user interaction description sections conclude with the IoT-system device for long-term measurements. The Fig. 5 shows the IoT-system installed on the PV panel: On the back of the PV panel, the power supply PCB with the battery (1) and the INA260-sensor PCB (2) were installed; in front of the top of the main PV panel, the auxiliary PV panel (3) and the weather-sensors PCB were fixed (4).

The system's power consumption varies depending on the operation mode. In AP mode the system consumes around 80mA, with pikes of 120mA during the connection process. In logger mode the power consumption is nearby 70mA, and the LPM feature allows reducing the consumption of idle mode to less than 25mA, providing more autonomy to the system.

E. The IoT-System measurements

The sensors used in the IoT-system have around 99.8% of accuracy. A sample data were collected during a week to compare it with weather data from a meteorological station at the university. The collected data showed similar tendency compared with the station, and the small value differences are attributed to the location of each device.

The data files were downloaded daily and concatenated by week to review data. The Fig. 6 shows the measurements of luminance in Luxes logged in 24 Hrs. The plotted data shows 288 points, one per measurement each 5 minutes, from 15:00 Hrs to 15:00 at next day. As the plot shows there are several points at 0 Luxes from the evening until the dawn of the next day.

The observational study upon the panel's output power was done using around 150 days of data. The collected data were the panel's output power(voltage and current), temperature, humidity, and sunlight intensity. This data was filtered to avoid none representative measurements when: panel's output voltage under 0.1V, low current values under 5mA and nighttime measurements.

Fig. 5. Physical setup: (1) Battery, (2) electrical parameters sensor, (3) auxiliary PV panel, (4) luxometer on top and temperature/humidity sensor on bottom.

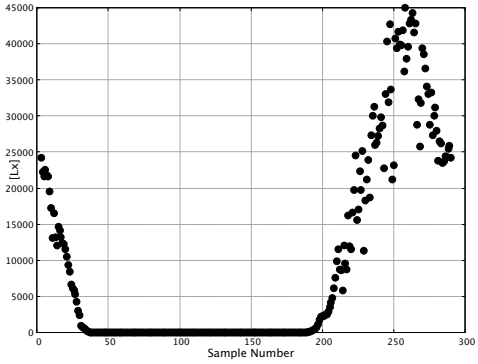

Fig. 6. 24 Hrs light intensity [Lx] measurements; light sample data.

F. Correlation Matrix

The complete set of data was analyzed using R for statistics and gnuplot for data visualization-analysis. In R the correlation matrix was calculated to study the dependence between the multiple measured variables. The resulting matrix contains the correlation coefficients between the related variables.

The most common correlation analysis used is the Pearson correlation test, which measures the linear dependence between two variables. However, there two more analysis options: the Spearman and Kendall correlations. These correlation are used for non-parametric data. The data is consider non-parametric when it is almost parametric but contains outliers, multiple peaks, a shift, or some other feature [26]. Thus, non-parametric statistics analysis like rank-based is

required. The rank-based format means sorting data and assign a rank from 1 to n for each unique value in the data-set. In this study both Pearson and Spearman correlations were used.

The Pearson correlation (r) measures a linear dependence between two variables (x and y), and it is defined by (1):

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} \tag{1}$$

here x and y are vectors of length n (number of observations), $\bar{x}$ and $\bar{y}$ are the means of x and y vectors, respectively.

The Spearman's correlation formula measures the correlation between the rank of x and the rank of y variables with the coefficient ρ defined by (2):

$$\rho = \frac{\sum[rank(x)-\bar{x}][rank(y)-\bar{y}]}{\sqrt{\sum[rank(x)-\bar{x}]^2 \sum[rank(y)-\bar{y}]^2}} \tag{2}$$

here $rank(x)$ and $rank(y)$ are the data values ordered based in the x and y values.

The selection between the Pearson and Spearman correlation was done by observing data in scatter plots. The plots are presented in next section.

III. Results

The correlation matrix requires the data in a concatenated way. Then, a scatterplot plot is commonly used to study the relationship between two variables to plot points (x, y), where x and y are related to the values of the variables. If the scatterplot shows some pattern it can be measured by the correlation.

However, considering the number of variables, for this study an scatterplot matrix was used. The scatterplot matrix allows to explore any relationship between the variables in the data-set. The scatterplot matrix is shown in Fig. 7.

The Fig. 7 shows 12 sub-plots, where the sub-scatterplots are arranged in the next way. Consider the "Lux" text-row, it gives the x label to figures a), b), and c), while the y labels are given by "Temp", "Hum", and "Pow" text-columns, respectively. Thus, Fig. 7.b shows the "Lux" versus "Hum" scatterplot. For this study, the upper right-hand side figures were plotted using small filled circular points; this helps to visualize the general data tendency. Then, the lower left-hand side figures are the mirror of the previous plots, but values are circular transparent gray-filled points, using this kind of points allows to see gray spots on dispersed data, while representative and repetitive points produce dark areas.

By reviewing plots, the 7.c scatterplot (Luxes vs. Power) shows the most relevant relationship for this study; the panel's output power as a function of light intensity. This plot presents a non-linear but positive tendency, and its mirror plot (7.j) shows that data is highly repetitive and thus representative. Also, the scatterplot 7.f shows a non-linear positive tendency for power, it means that at high temperatures values high output power. However, the plot 7.k shows a darker cloud at high temperatures, meaning that most of the time the panel was exposed at high temperatures to obtain the power plot showed in 7.c.

Another interesting behaviors are shown in figures 7.a and 7.b. These plots show how the light intensity depends on temperatures 7.a, and how the humidity affects it 7.b. Then, 7.g shows that humidity was changing from low values to high values (100%) during measurements, the most representative areas are at mid-low humidity values for greatest light intensity records, and at 100% of humidity for the lower light intensity values; this data probably were logged previous to the night, or during cloudy days.

The scatterplots matrix show a non-linear tendency in all plots. Hence, the Spearman's correlation coefficients formula was applied to the complete data set. The equation (3) shows the calculated Spearman's coefficients:

$$\rho = \begin{bmatrix} & Lux & Temp & Hum & Pow \\ Lux & 1.0000 & 0.8455 & -0.7595 & 0.9852 \\ Temp & 0.8455 & 1.0000 & -0.8737 & 0.8226 \\ Hum & -0.7595 & -0.8737945 & 1.0000 & -0.7360 \\ Pow & 0.9852 & 0.8226 & -0.7360 & 1.0000 \end{bmatrix} \tag{3}$$

and equation (4) shows the Pearson's coefficients formula:

$$r = \begin{bmatrix} & Lux & Temp & Hum & Pow \\ Lux & 1.0000 & 0.8178 & -0.7264 & 0.9612 \\ Temp & 0.8178 & 1.0000 & -0.8897 & 0.8136 \\ Hum & -0.7264 & -0.8897 & 1.0000 & -0.7308 \\ Pow & 0.9612 & 0.8136 & -0.7308 & 1.0000 \end{bmatrix} \tag{4}$$

Note that Spearman's coefficients have a better correlation than Pearson's. Thus, the collected data fit better to a non-parametric behavior. For example by comparing the most relevant and less dispersed data, Luxes versus Power (Fig. 7.c), the Spearman's coefficient $\rho(1,4) = 0.9852$ is higher than Pearson's coefficient $r(1,4) = 0.9612$. On the other hand, the lower value in both matrices are $\rho(3,4) = -0.7360$ and $r(3,4)) - 0.7308$, fitting again better the Spearman coefficient than Pearson. This low value can be attributed to the dispersion of data, as shown in Fig. 7.i.

Considering the high values for Lux vs. Pow and Temp vs. Pow, at Spearman's matrix, it is important to visualize how these three variables are related. In Fig. 8, the scatterplot shows the three parameters: in x axis the light intensity measurements in Luxes, in y axis the power in Watts, and the temperature was used as an intensity color value mapped into a false grayscale palette. Note that the correlation coefficients show their relationship in the scatterplot. The higher values of power occur at high temperature (gray and whit filled points) and light intensity. The dark-gray or dark filled points belongs to low power and low-intensity values.

Another interesting plot was obtained by observing the Spearman's coefficients for humidity and power(-0.7360), and temperature with humidity (-0.8737). Both coefficients are negative, meaning an inverse relationship with power and humidity, respectively. The Fig. 9 was plotted using again temperature as false gray scale palette. The figure shows how the panel's power was affected by humidity, in fact how the light intensity was affected by high humidity values, most probably during cloudy or rainy days. Also, this could happen during measurements next to the sunset time. The temperature

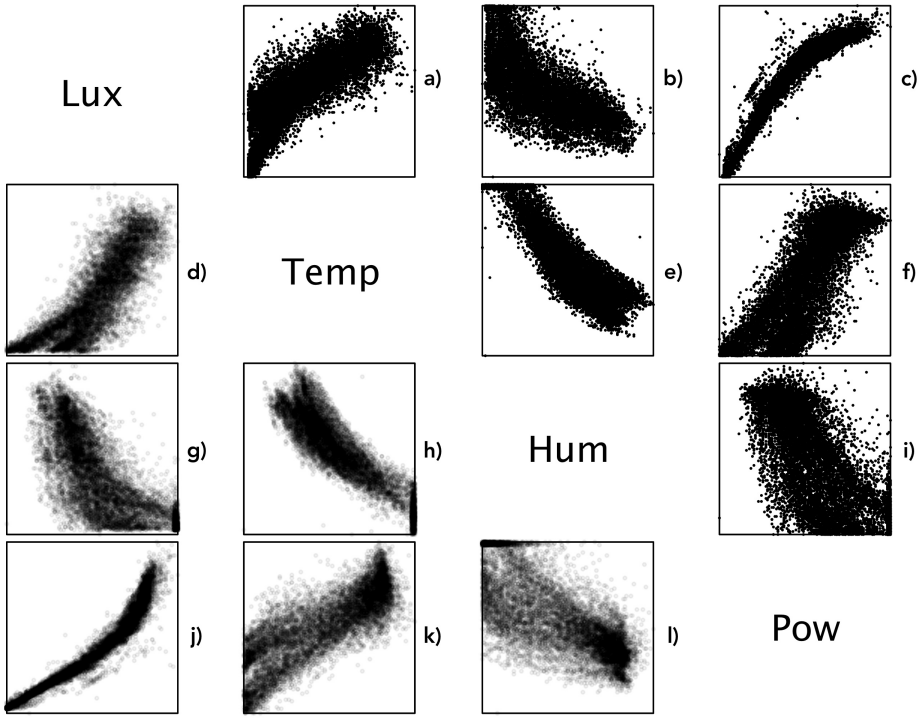

Fig. 7. Scatterplot matrix in (x, y) plots: a)Luxes vs Temperature, b)Luxes vs Humidity c)Luxes vs Power, d)Temperature vs Luxes, e)Temperature vs Humidity, f)Temperature vs Power, g)Humidity vs Luxes, h) Humidity vs Temperature, i) Humidity vs Power, j) Power vs Luxes, k)Power vs Temperature, and l)Power vs Humidity.

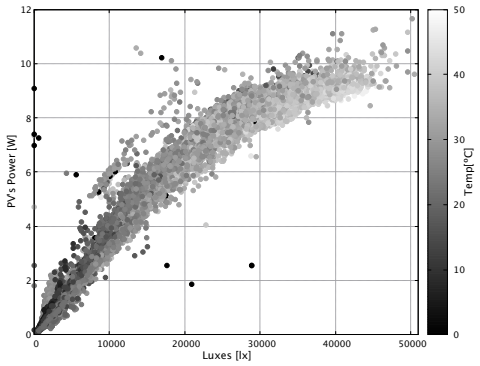

Fig. 8. Scatterplot for light intensity measurements versus panel's output power, with temperature as parameter-palette.

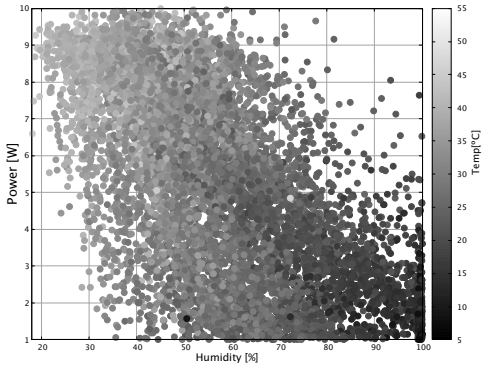

Fig. 9. Scatterplot for humidity measurements versus panel's power, with temperature as parameter-palette.

values support this conjecture, for example, at 100% humidity values with low power there are low temperatures, while the opposite occurs for high power values during normal humidity values(25%-45%) at high temperatures.

IV. Conclusions

The IoT-system was able to collect more than 40,000 entrance for humidity, temperature, light intensity, and panel's power; more than 160,000 measurements.

Also, the system improves the form that data was collected and prepared for further analysis. The complete data-set was useful to create a scatterplot matrix to observe the relationship between all the variables measured. The scatterplot matrix allows visualizing the level of relationship between variables. By using transparent gray-filled points, plots also show how repetitive were the data used in the study; and for further analysis or models. The dispersion in temperature and humidity scatterplots versus power can be reduced if data are used together as input in the modeling stage. Thus, by knowing the weather, models can estimate how the panel's output power will be affected.

Even when some scatterplots seem similar to a linear relationship, the non-parametric correlation coefficients fitted better. The Spearman coefficients matrix was selected, considering how this method ranked the data, and its calculation.

The analysis over coefficient values allows exploring a more complex connection about how power measures were affected directly or indirectly by the different parameters.

Furthermore, the high values of the Spearman matrix classify data as fuzzy data. Then, this data can be used to estimate the panel's output power as a function of the measured variables by using machine learning.

V. Acknowledgment

The authors would like to thank the National Laboratory "SEDEAM" for providing the equipment, tools and electronic components to build the system prototype through the projects No. 235780, 271878 and 282357. Also to the "Tecnológico Nacional de México" TecNM by supporting this project through the founding 6127.17-P.

References

[1] J. Twidell and T. Weir, *Renewable energy resources*. Routledge, 2015.

[2] C. Hargreaves, "The Worldwide Distribution of Solar Resources by Nation by Act on Climate," Tech. Rep., 2015. [Online]. Available: http://www.wrforum.org/wp-content/uploads/2015/10/SS1-Hargreaves.pdf

[3] CONAGUA, "Irradiación Solar," 2012. [Online]. Available: http://www.conagua.gob.mx/CONAGUA07/Contenido/Documentos/presentacion1.pdf

[4] N. Panwar, S. Kaushik, and S. Kothari, "Role of renewable energy sources in environmental protection: A review," *Renewable and Sustainable Energy Reviews*, vol. 15, no. 3, pp. 1513–1524, 4 2011. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1364032110004065

[5] T. B. Johansson and L. Burnham, *Renewable energy : sources for fuels and electricity*. Island Press, 1993.

[6] I. International Energy Agency, "Key world energy statistics," 2016.

[7] V. Omubo-Pepple, "Effects of temperature, solar flux and relative humidity on the efficient conversion of solar energy to electricity," ... *Journal of Scientific* ..., vol. 35, no. 2, pp. 173–180, 2009. [Online]. Available: http://content.imamu.edu.sa/Scholars/it/VisualBasic/ejsr_35_2_02.pdf

[8] S. Mekhilef, R. Saidur, and M. Kamalisarvestani, "Effect of dust, humidity and air velocity on efficiency of photovoltaic cells," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 5, pp. 2920 – 2925, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032112001050

[9] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, "Predicting solar generation from weather forecasts using machine learning," in *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*. IEEE, 2011, pp. 528–533.

[10] T. Verma, A. Tiwana, C. Reddy, V. Arora, and P. Devanand, "Data analysis to generate models based on neural network and regression for solar power generation forecasting," in *Intelligent Systems, Modelling and Simulation (ISMS), 2016 7th International Conference on*. IEEE, 2016, pp. 97–100.

[11] S. R. Potnuru, D. Pattabiraman, S. I. Ganesan, and N. Chilakapati, "Positioning of PV panels for reduction in line losses and mismatch losses in PV array," *Renewable Energy*, vol. 78, pp. 264–275, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.renene.2014.12.055

[12] S. Dubey, J. N. Sarvaiya, and B. Seshadri, "Temperature dependent photovoltaic (PV) efficiency and its effect on PV production in the world - A review," *Energy Procedia*, vol. 33, pp. 311–321, 2013.

[13] V. J. Fesharaki, M. Dehghani, J. J. Fesharaki, and H. Tavasoli, "The Effect of Temperature on Photovoltaic Cell Efficiency," *Proceeding of the 1st International Conference on Emerging Trends in Energy Conservation-ETEC*, no. November, pp. 20–21, 2011.

[14] B. Parida, S. Iniyan, and R. Goic, "A review of solar photovoltaic technologies," *Renewable and Sustainable Energy Reviews*, vol. 15, no. 3, pp. 1625–1636, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032110004016

[15] M. R. Maghami, H. Hizam, C. Gomes, M. A. Radzi, M. I. Rezadad, and S. Hajighorbani, "Power loss due to soiling on solar panel: A review," *Renewable and Sustainable Energy Reviews*, vol. 59, pp. 1307 – 1316, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032116000745

[16] D. Dondi, A. Bertacchini, D. Brunelli, L. Larcher, and L. Benini, "Modeling and optimization of a solar energy harvester system for self-powered wireless sensor networks," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 7, pp. 2759–2766, 2008.

[17] T. Khatib, A. Mohamed, and K. Sopian, "A review of solar energy modeling techniques," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 5, pp. 2864–2869, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032112000767

[18] M. Hadjab, S. Berrah, and H. Abid, "Neural network for modeling solar panel," *International Journal of Energy*, vol. 6, no. 1, pp. 9–16, 2012.

[19] HyperThings, "Solar monitoring," 2019. [Online]. Available: http://www.hyperthings.in/index.php/iot-solutions/solar-monitoring

[20] PowerWise, "Monitor solar production," 2019. [Online]. Available: https://www.powerwisesystems.com/monitoring-control-solutions/monitor-solar-production-and-performance/

[21] K. K. Patel and S. M. Patel, "Internet of things-iot: definition, characteristics, architecture, enabling technologies, application & future challenges," *Int. J. Eng. Sci. Comput*, vol. 6, no. 5, 2016.

[22] R. van Kranenburg, E. Anzelmo, A. Bassi, D. Caprio, S. Dodson, and M. Ratto, "The Internet of Things," *1st Berlin Symposium on Internet and Society*, no. October, pp. 26–36, 2011. [Online]. Available: https://www.researchgate.net/profile/Matt_Ratto/publication/228360933_The_Internet_of_Things/links/0912f513755ebd1e87000000.pdf

[23] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645–1660, sep 2013. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0167739X13000241

[24] Texas Instruments, "CC3200 SimpleLink™ Wi-Fi® and Internet-of-Things Solution, a Single-Chip Wireless MCU," p. 71, 2013. [Online]. Available: http://www.ti.com/lit/ds/symlink/cc3200.pdf

[25] Real Time Engineers Ltd., "The Free RTOS™ Reference Manual," 2016. [Online]. Available: https://www.freertos.org/Documentation/FreeRTOS_Reference_Manual_V10.0.0.pdf

[26] J. Brownlee, *Statistical Methods for Machine Learning*, 2019.