Data Article

Solar panels simulation data generated using LTSpice under different operating conditions

Kanika Sood\textsuperscript{a}, Nathaniel Ruppert\textsuperscript{b}, Rakeshkumar Mahto\textsuperscript{b,}\textsuperscript{*}

\textsuperscript{a} Department of Computer Science, California State University, Fullerton, CA 92831, United States
\textsuperscript{b} Department of Electrical and Computer Engineering, California State University, Fullerton, CA 92831, United States

\textbf{ARTICLE INFO}

\textit{Article history:}
Received 20 July 2022
Revised 17 August 2022
Accepted 5 September 2022
Available online 11 September 2022

Dataset link: Simulation Dataset of Partial Shading and Fault of a Photovoltaic Module (Original data)

\textbf{Keywords:}
Photovoltaics
Spice
Machine learning
Artificial Intelligence
Renewable Energy

\textbf{ABSTRACT}

This paper presents a detailed description of the circuit simulation data obtained using LTSpice, a high-performance SPICE simulation software for easing the simulation of solar panel circuit data. The data represents the photovoltaic modules with different configurations and cells placed in series and parallel. The data are obtained from automated Python creation and simulation of PV cells in specified formats. The collected data are then organized in a CSV (Comma Separated Value) file. Each file contains properties of the photovoltaic cell, such as individual cell voltage, current, and others. The data are then used to evaluate each of the photovoltaic modules of an array.

Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

\textbf{Specifications Table}

| Subject | Computer Science, Engineering |
|---------|--------------------------------|
| Specific subject area          | Artificial intelligence, Photovoltaic energy, Optimization, Machine learning |
| Type of data                    | Table, Image, Figure |

* Corresponding author.
E-mail address: ramahto@fullerton.edu (R. Mahto).

\url{https://doi.org/10.1016/j.dib.2022.108581}
2352-3409/Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Value of the Data

- The data presented in this work can interest a broader community of researchers involved in designing and modeling machine learning algorithms for optimizing solar panels.
- With this data, the readers can evaluate the behavior of PV systems under different shading and fault conditions at various temperatures to better understand the performance of PV panels in different conditions.
- In this research, we compare a variety of machine learning (ML) algorithms to identify faults in PV systems. These results can be helpful for comparing with other ML algorithms to enhance the performance of PV based systems to power drones, lookout towers, small electronic devices, embedded systems, and many others. Drones have shown their effectiveness during natural or human-made hazards such as forest fires, nuclear accidents, floods, and others. In such operation before the deployment of solar powered drones, the presented data will enable to test the performance of PV modules in different hazardous operating conditions. Even smart sensors installed at a remote location can be powered by PV based system. The presented dataset is useful for simulating its performance in various operating conditions.
- Recently, machine learning [1–5] and fuzzy logic-based [6–9] maximum power point tracking (MPPT) have been used to ensure the PV systems can operate at optimal efficiency. All the techniques require a dataset to test their algorithm’s effectiveness before deploying it on an existing PV system. The data presented in this work is beneficial in testing the MPPT algorithm in a realistic virtual environment.

1. Data Description

The collected dataset presents ten attributes that describe the solar panel. Each solar cell is represented as an individual instance in the training dataset. The solar cells are modeled using...
the doubled diode based equivalent model shown in Fig. 1. For every instance, a unique serial number is assigned, incremented by one, and includes features associated with a PV cell.

Fig. 2 presents a snippet of the attributes of the solar panel that are used as descriptive features and fed as input to the machine learning algorithms. The complete dataset used in this work is available at the Mendeley Data repository [10]. The descriptive features are mapped using a hypothesis and the relationship between the input features and the target label is captured by the mapping function. The mapping function can be used to make predictions on new unseen data. Fig. 3 highlights the correlation between the input features. The correlation matrix highlights that the 'series cells' feature is highly correlated with voltage and parallel cells.

Additionally, power and current are highly correlated as well. Fig. 4 shows the relationship between variables categorized by their fault types. This work explores three types of faults in solar panels: shaded, short circuit, and open circuit. In the dataset, we use 0 for shaded, 1 for short circuit, and 2 for open circuit. Fig. 5 presents the features plotted based on temperatures ranging from 20 to 50 °Celsius.
Fig. 3. Correlation matrix for the input features.

Fig. 4. Feature Visualization plot by type.
2. Experimental Design, Materials and Methods

For the ML based classifier to be effective, it is essential to have an accurate dataset with plenty of data points to train the ML classifier. However, generating such an extensive dataset with experimental data with all possible operating conditions probable during the mission is tedious. Hence, to simplify this lengthy task SPICE modeling technique can be used to generate the dataset. In [11], to generate the dataset for different operating conditions of PV module, Python, LTSPICE XVII (Version 17.0.35) [12], and 2-diode based equivalent PV module modeling technique is used, as shown in Fig. 1.

To create a dataset for training the ML algorithm, the following conditions include solar irradiation, various PV panel configurations, fault conditions (Short, Open), partial shading conditions, and temperature. The different PV configurations used resulted in the simulation of 6965,234 different scenarios. For emulating the partial shading conditions, we consider the solar irradiation from 500 to 1000 W/m². An ideal unshaded PV cell is with solar irradiation of 1000 W/m². For emulating the distinct types of shading conditions throughout the day, the solar irradiation of shaded PV cells varies from 900 W/m² to 500 W/m². Later, the Python program designed for this paper programmatically creates all the required scenarios via the following equations:

\[
C_{\text{open}}[\text{Ns},\text{Np}] = (\text{Np}-i) \times N_S
\]  

(1)
Fig. 6. Steps involved in Dataset Generation.

\[
C_{\text{short}\{N_s,N_p\}} = N_p \times (N_s - j) \quad (2)
\]

\[
C_{\text{shade}\{N_s,N_p\}} = N_p \times N_s \quad (3)
\]

\[
S_{N_s,N_p} = C_{\text{open}\{N_s,N_p\}} + C_{\text{short}\{N_s,N_p\}} + C_{\text{shade}\{N_s,N_p\}} \quad (4)
\]

\[
S_{\text{tot}} = \sum (S_{N_s,N_p}) \times T_s \times I_s \quad (5)
\]

where \( C \) is the number of different configurations for \( N_p \) is the total number of PV cells from parallel and \( N_s \) is the total number of PV cells in series, which are considered from 1 to 10. The variable \( i \) in the Eq. (1) is given by \( 0 < i \leq N_p - 1 \) and variable \( j \) in Eq. (2) is given by \( 0 < j \leq N_s - 1 \). For the short and open circuit emulation, a number of the PV cells in series and parallel were electrically removed, respectively.

Eqs. (1), (2), and (3) describe how the number of necessary simulations is determined for open, short, and shade conditions, respectively. Eq. (4) is the combination of (1), (2), and (3) and is the total simulations to be run based on the various PV panel configuration, with (5) being the total amount of simulations to run, where \( T_s \) is the number of temperature scenarios and \( I_s \) is the amount of solar irradiation intensity scenarios. Once per PV configuration is determined from (4) for a set \([N_s,N_p]\), the Python program generates SPICE netlist (.cir) files for simulation. After all PV panel per cell configurations have been created, LTSPICEXVII is run via the Python program across all .cir files, producing .raw files. The files contain output voltage, generated current, and total power per cell, which are collected via the Pandas library into a single .csv file, which is used in ML processing. The subsequent steps involve using ML tools in preprocessing, training, testing, and classification. The use of, which are collected via the Pandas library into a single .csv file, which is used in ML processing. All steps involved in generating the dataset are shown in Fig. 6.

Ethics Statements

These data are primary data and do not include human subjects, animal experiments, or social media platforms.

CRediT Author Statement

Kanika Sood: Machine Learning Modelling, Writing, Methodology, Software, Conceptualization; Rakeshkumar Mahto: Investigation, Validation, Conceptualization, Data curation, Writing; Nathaniel Ruppert: Visualization, Data analysis, Data Simulation.
Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Simulation Dataset of Partial Shading and Fault of a Photovoltaic Module (Original data) (Mendeley Data).

Acknowledgments

Funding: This work is supported by the start-up fund provided by the Computer Science department and Summer Undergraduate Research Academy (SURE-A) funding offered by the Office of Research and Sponsors Project.

References

[1] K. Sood, R. Mahto, H. Shah, A. Murrell, Power Management of Autonomous Drones using Machine Learning, in: 2021 IEEE Conference on Technologies for Sustainability (SusTech), Apr. 2021, pp. 1–8, doi:10.1109/SusTech51236.2021.9467475.
[2] G.M. Tina, C. Ventura, S. Ferlito, S. De Vito, A state-of-art-review on machine-learning based methods for PV, Appl. Sci. 11 (16) (2021) 7550.
[3] M.S. Nkambule, A.N. Hasan, A. Ali, J. Hong, Z.W. Geem, Comprehensive evaluation of machine learning MPPT algorithms for a PV system under different weather conditions, J. Electr. Eng. Technol. 16 (1) (2021) 411–427.
[4] C. Kalogerakis, E. Koutroulis, M.G. Lagoudakis, Global MPPT based on machine-learning for PV arrays operating under partial shading conditions, Appl. Sci. 10 (2) (2020) 700.
[5] M. Takruri, et al., Maximum power point tracking of PV system based on machine learning, Energies 13 (3) (2020) 692.
[6] A.M. Noman, K.E. Addoweesh, H.M. Mashaly, A fuzzy logic control method for MPPT of PV systems, in: IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society, 2012, pp. 874–880.
[7] X. Li, H. Wen, Y. Hu, L. Jiang, A novel beta parameter based fuzzy-logic controller for photovoltaic MPPT application, Renew. Energy 130 (2019) 416–427.
[8] J.-K. Shiau, Y.-C. Wei, B.-C. Chen, A study on the fuzzy-logic-based solar power MPPT algorithms using different fuzzy input variables, Algorithms 8 (2) (2015) 100–127.
[9] C. Robles Algarín, J.Taborda Giraldo, O. Rodriguez Alvarez, Fuzzy logic based MPPT controller for a PV system, Energies 10 (12) (2017) 2036.
[10] K. Sood, N. Ruppert, and R. Mahto, “Simulation Dataset of Partial Shading and Fault of a Photovoltaic Module,” vol. 1, Aug. 2022, doi:10.17632/3fr92f4xy9.1.
[11] R. Mahto, N. Ruppert, A. Nguyen, G. Kalotra, Fuzzy Logic Based MPT Algorithm for Reconfigurable Photovoltaics, presented at the IEEE Global Humanitarian Technology Conference (GHTC), 2021.
[12] “LTSPICE.” https://www.analog.com/en/design-center/design-tools-and-calculators.html#LTspice. Accessed August 14, 2022.