

Assessing the Modelling Approach and Datasets Required for Fault Detection in Photovoltaic Systems

Max Bird

Imperial College London
mhb316@ic.ac.uk

Salvador Acha

Imperial College London
salvador.acha@ic.ac.uk

Niccolo Le Brun

Imperial College London
niccolo.le-brun11@ic.ac.uk

Nilay Shah

Imperial College London
n.shah@ic.ac.uk

Abstract-- Reliable monitoring for photovoltaic assets (PVs) is essential to ensuring uptake, long term performance, and maximum return on investment of renewable systems. To this end this paper investigates the input data and machine learning techniques required for day-behind predictions of PV generation, within the scope of conducting informed maintenance of these systems. Five years of PV generation data at hourly intervals were retrieved from four commercial building-mounted PV installations in the UK, as well as weather data retrieved from MIDAS. A support vector machine, random forest and artificial neural network were trained to predict PV power generation. Random forest performed best, achieving an average mean relative error of 2.7%. Irradiance, previous generation and solar position were found to be the most important variables. Overall, this work shows how low-cost data driven analysis of PV systems can be used to support the effective management of such assets.

Index Terms— Fault detection, machine learning, photovoltaics, random forest, weather data

NOMENCLATURE

N	Number of data points
$W_{\text{predicted}}$	Predicted PV generation
W_{actual}	Actual PV generation
W_{total}	Total installed PV capacity

I. INTRODUCTION

The use of photovoltaic (PV) systems has experienced rapid growth in recent years. In 2017 alone China installed over 53 GW of solar capacity, equivalent to the grid size of California, home to the world's 5th largest economy. To ensure continued global growth and maximum performance of current installations, it is critical to develop low-cost, effective maintenance solutions. Well maintained assets have extended usable life, increased uptime and, as a result, larger lifetime power generation and return on investment. Commercial PV maintenance strategies are primarily either reactive, where assets are repaired when failure occurs, or preventative, where maintenance is performed after a set period based on an estimated probability of failure. Both approaches require a relatively large time, capital investment, and are labour intensive; thus, a cheaper, more effective solution is desirable.

Many PV monitoring solutions are present in literature; designed to detect, classify and locate faults in systems by comparing monitored system behaviour (output power, voltage, current) to predictions obtained from a reference model. Research efforts in this area can be broadly split into the different approaches used to create this reference model, with a summary of the different types given in [1]. Electrical methods have been popular in the past, which commonly use parametric relationships between expected power output, operating voltage and current [2, 10-11] with extended studies also including incident irradiance and temperature to correct for weather effects [3-4, 12]. These require comprehensive monitoring of

panel behaviour (voltage, current) which introduces additional costs related to data transmission, discussed in [1].

Forecasting power production of PVs is also commonly used to generate reference performance models, generally focusing on the relationship between PV power generation and relevant weather variables. Classical regressive techniques and, more recently, machine learning models have been used, with [5] giving a comprehensive overview of these approaches and their prevalence in forecasting solar generation. Machine learning (ML) techniques are especially popular, due to their ability to predict relationships between input and output variables in highly non-linear systems. Support vector machines (SVM) [6], artificial neural networks (ANN) [7,13] and random forest (RF) [8] have all been used and show success in predicting PV power output. These approaches benefit from the user not needing to know any specifics about the PV system components, as they simply identify the relationship between its power production and the weather it experiences. Other methods [9, 14] also include a pre-processing stage, where day types such as cloudy or non-cloudy are defined, and individual models are trained for each type. The benefits of this approach are unclear and add additional computational requirements, so have not been considered here but could be investigated in an expanded study.

The current work aims to build on the above, by providing an in-depth overview of the dataset and modelling approach required for day-behind, weather-based modelling of PV generation in relation to fault detection. Different ML algorithms, weather variables, training set sizes and distances between PV sites and weather stations are all investigated, and their effects on modelling accuracy compared. This research improves on previous works for several reasons, summarised below:

- Detailed data pre-processing to allow selection of high-quality sites for analysis.
- Use of a much larger dataset, based on 5 years of historical generation from multiple commercial sites and 5 years of historical weather data from weather stations across the UK.
- Comparison between 3 different machine learning techniques, namely SVM, ANN and RF.
- Inclusion of 7 different weather variables, with their effect on modelling performance compared.

II. METHODOLOGY

The proposed methodology consists of a data collection and processing stage to identify viable sites, followed by a training stage and various sensitivity analyses, summarised in Fig. 1.

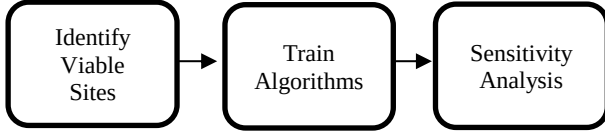


Fig. 1: Flowchart for methodology.

A. Identifying Viable Sites

A large database of over 100 commercial roof-mounted PV installations (sites) was used as the source of historical PV generation data, combined with historical weather data gathered from the Met Office Integrated Data Archives (MIDAS). Generation and weather data are then loaded into an SQL database for easy access and querying. Both datasets cover hundreds of PV sites and weather stations from all over the UK, and so a process is required to select multiple high-quality site/weather station pairings for analysis. A PV installation is defined as viable if:

- It has generation records from January 2013 – 2018;
- It is within 5 km of a weather station which also has 98% of all-weather variables (detailed in the next section) over the same time period.

To satisfy these criteria several python scripts were designed. Firstly, a shortlist of sites with a weather station within a 5km radius was generated. Next, the weather records of each weather station are checked for missing entries and removed if they have less than 98% of any weather type. Due to the huge scale of this weather data, and MIDAS undertaking their own internal data quality checks, no further checks were performed. Finally, the remaining sites historical generation datasets were examined for a number of predefined error types, summarised in Table I.

Table I - Summary of error types.

Error Type	Description
1	Zero generation entries recorded within a day
2	Days with zero generation recorded
3	Nights with non-zero generation recorded
4	Missing entries
5	Negative entries

Table II - Percentage occurrence of each error type.

Site	Percentage Occurrence (%)				
	Type 1	Type 2	Type 3	Type 4	Type 5
A	0.011	6.356	0.000	0.000	0.000
B	0.017	0.329	0.000	0.000	0.000
C	0.005	0.219	0.000	0.008	0.000
D	0.024	0.000	0.000	0.009	0.000

From this, four site/weather station pairs were found to have acceptably low error counts, summarised in Table II, and were selected for analysis. The distance between each site and its weather station are as follows: Site A – 4.73 km, Site B – 4.63 km, Site C – 2.15 km and Site D – 5.15 km. For type 1 errors, the missing entries are calculated using linear interpolation between the nearest two non-zero generation values within that

day. For type 2 and type 4 errors the missing/zero generation is taken to be equal to the generation across the previous day. Although this is an approximation, the occurrence of this error type is extremely limited and should not have noticeable impact on prediction accuracy. Site A is the only site that shows a possible data issue, with 6.4% of its days having no recorded generation. However, all missing days are in the validation year, and as such Site A is still included to strengthen the analysis, but its validation results are not presented.

All analysed sites are installed with Solyndra, a cylindrical panel of copper indium gallium selenide (CIGS) thin film solar cells, with varying capacities [15]. Their location and capacity are summarised in Fig. 2. The daily generation for site C across the 5-year period is illustrated in Fig. 3 to exemplify the data analysed.

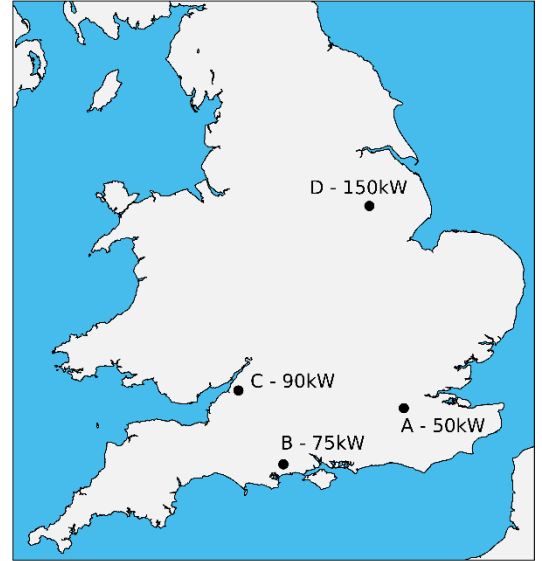


Fig. 2: Location and PV capacity of sites analysed in this study.

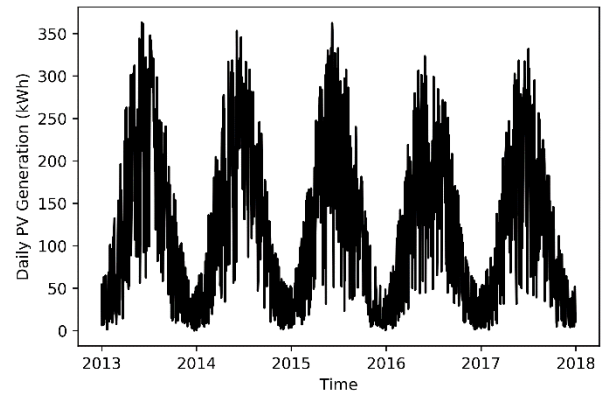


Fig. 3: Daily PV generation for site C.

B. Training Stage Description

Three different popular machine learning algorithms for PV forecasting are compared; a simple support vector machine (SVM), a random forest (RF) classifier and an artificial neural network (ANN) called Adam. An overview of each algorithm, and an in-depth discussion of their prevalence in PV generation forecasting can be found in [5].

SVM is the earliest form of machine learning, constructing a hyper-plane or set of hyper planes into a theoretically infinite dimensional space. It maximises the distance between the hyper-plane and the nearest training points (the support vectors) to lower the generalisation error and make classification as accurate as possible. Random forests on the other hand are built on an ensemble of decision trees, to perform tasks such as regression or classification. Random subsets of training data are taken, a model is fitted, and then the predictions are aggregated across the sub-sets. This approach lowers the variance in predictions and reduces overfitting the training set, resulting in a robust predictive model. The ANN used utilises a multi-layer perceptron (MLP) method called ‘Adam’. It uses multiple hidden layers and optimises the squared loss between predictions and actual values with a stochastic gradient descent method. It is also noted to work well with especially large datasets, as well as being computationally efficient compared to other methods, hence its inclusion in this paper. The scikit-learn python module was used for all analysis, with almost all parameters kept as their default values as the focus of this paper is on input variable selection rather than optimising the machine learning algorithms themselves. The variables chosen are defined as follows: for SVM ($\gamma = 10^{-5}$), for RF (n-estimators = 1000) and for Adam ($\alpha = 10^{-5}$, hidden-layers = (1000,)).

To complement our analysis, the predication results from an existing commercial model, which uses multi-variant linear regression (MVLRL) with sunshine hours, cloud cover and daylight hours as input variables, is also included. As only the MVLRL predictions are available and not the model itself, none of the analysis mentioned below was applied to it, and its predictive accuracy is simply calculated to allow comparison with existing modelling approaches.

For each of the three different machine learning algorithms used, two different ‘model types’ are defined, which vary the input data supplied, summarised in Table III.

Table III - Model type summary

Model Type	Inputs	Output	Time Scale
1	Irradiance, solar position, wind direction and speed, ambient temperature, relative humidity, all for prediction time interval	Panel generation (kWh)	Hourly
2	All weather types above for prediction time interval, all weather types above for the previous time interval, panel generation of previous time interval	Panel generation (kWh)	Hourly

Model 1 simply inputs all weather types from the prediction interval into the algorithm, as is common in PV generation predictions. Model 2 however also includes data from the previous time interval, aiming to capture any inertia in the system, which could lead to improvements in predictive accuracy. Only one previous time interval has been included, to prevent the analysis becoming computationally intractable. For

each algorithm and model type, the entire data set is split into a three-year training set, a one-year test set and a further one-year validation set. Due to site A missing data within the validation year, as well as for brevity, validation results for all sites are not presented here, but have been calculated and agree closely with test year accuracies for all sites.

Once this analysis is carried out, the combination of model type and algorithm with the best predictive accuracy will be selected as our best-case modelling approach. This will then be used in the sensitivity analysis discussed below.

C. Sensitivity Analysis

Initially, each weather type was individually removed from the training and test sets of our best-case modelling approach, it to see its effect on predictive accuracy. Next, three different selections of data inputs were used as a training set, namely 1). irradiance, 2). irradiance and previous generation, and 3). irradiance, previous generation and solar position. These three scenarios were chosen as they only use data which is either integral to predictions (irradiance), or easily obtainable (solar position and previous generation) and should represent the minimum dataset necessary when applying this approach.

Next, the length of the training set was varied using 1 month, 3 month, 6 month, 1 year and 2 year sets, and compared to the results for the standard 3-year set. Each training set was tested on the same year of actual generation (2016) and comprised of data immediately preceding the test year.

Finally, the effect of the distance between the site and its weather station was investigated. First, comparison within the initial 5 km radius was considered. Next, a larger 75 km radius was considered, to see how modelling accuracy drops off over larger distances. Site A is the only site with enough weather stations meeting the viable criteria within this radius, so is the only site used.

The motivation behind these analyses is to work backwards from the initial ‘best-case’ modelling approach defined at the end of Section B, as it uses extensive datasets which are not commonly available. Through this analysis, informed selection of suitable weather variables, length of training set and weather station distances can be made, allowing more effective modelling in future work.

III. MODEL VALIDATION

To evaluate algorithm performance, two popular predictive indicators were applied, namely the mean absolute percentage error (MRE) and the root mean square error (RMSE). The MRE is defined below,

$$MRE = \frac{1}{N} \sum_{t=1}^N \frac{|W_{predicted} - W_{actual}|}{W_{total}} \times 100\% \quad (1)$$

This metric is useful as it is normalised by the total installed capacity of each site, allowing comparison between sites of different capacities. MRE is used here rather than mean absolute percentage error (MAPE); which is commonly found in photovoltaic forecasting evaluation literature [5]. MAPE

divides the forecasting error by the actual generation value, and the resulting errors can become quite large for small forecasting values both at the beginning and end of the day. MRE instead divides by the total installed capacity of the site, preventing this bias.

The root mean square error (RMSE) was also used, which shows the spread of the residuals between the predicted and actual generation; this value depends on the size of the installation thus no averages of this value are presented.

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (W_{predicted} - W_{actual})^2}{N}} \quad (2)$$

Note also that both indicators were calculated only between predicted and actual generation where the generation of the panel is non-zero, to prevent them becoming inflated by night-time predictions accrued from zero values.

IV. RESULTS AND DISCUSSION

Firstly, a comparison between each combination of machine learning algorithm and model type is presented, with results for the existing commercial model (MVLR) also included. Next, several sensitivity analyses are covered, looking firstly at the influence weather variables have, then the size of the training set and then the distance between the sites and weather stations.

Table IV - MRE and RMSE for Model 1

	Algorithm	Site			
		A	B	C	D
MRE (%)	MVLR	6.62	4.69	4.02	6.33
	Adam	4.14	3.04	2.48	4.36
	SVM	4.45	3.75	3.27	4.81
	RF	4.15	3.1	2.54	4.88
RMSE (kWh)	MVLR	3.25	3.43	3.52	9.27
	Adam	2.17	2.35	2.35	6.97
	SVM	2.34	2.93	3.07	7.5
	RF	2.31	2.59	2.53	8.01

Table V - MRE and RMSE for Model 2

	Algorithm	Site			
		A	B	C	D
MRE (%)	MVLR	6.62	4.69	4.02	6.33
	Adam	3.38	2.54	2.77	2.96
	SVM	3.31	2.78	2.48	4.1
	RF	2.88	2.29	1.74	3.88
RMSE (kWh)	MVLR	3.25	3.43	3.52	9.27
	Adam	1.73	1.96	2.39	4.63
	SVM	1.73	2.28	2.34	6.35
	RF	1.64	2.01	1.75	6.39

TABLE VI - Average MRE values for each Model type

Model Type	MRE (%)			
	MVLR	Adam	SVM	RF
1	5.42	3.50	4.07	3.67
2	5.42	2.91	3.17	2.70

A. Model Comparison

Tables IV and V show the MRE and RMSE for Model 1 and 2 respectively, broken down for each site, with average values also presented in Table VI. In general, Model 2 performs significantly better than Model 1, with all algorithms showing improvement in both RMSE and MRE. Random forest for example reduces its MRE from 3.67% to 2.70%. Clearly each algorithm is benefitting from incorporating weather data and generation from the previous time interval.

All machine learning approaches outperform MLVR used in the existing commercial model, reflecting the strength of machine learning techniques when analysing complex non-linear systems. As seen in Fig. 4, SVM, Adam and Random Forest can capture the generation of the panels at a much higher resolution than MVLR, closely tracking the hourly behaviour, rather than just general trends across a day, this makes them much better suited for predictive maintenance.

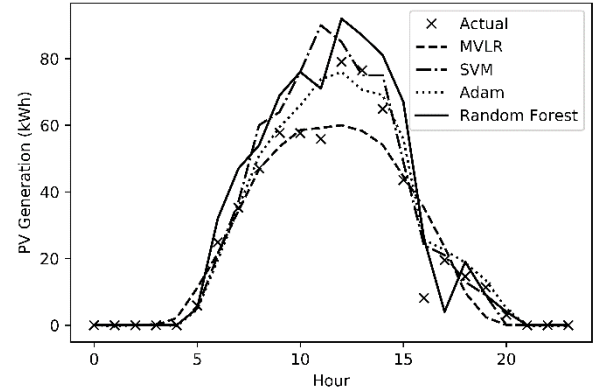


Fig. 4: Site C hourly predictions vs actual generation

Overall RF performs slightly better than both SVM and Adam, achieving an MRE in Model 2 of just 2.70% compared to 3.17% and 2.91% respectively. From this, RF Model 2 is selected as the best modelling approach at this stage and is carried forward into the sensitivity analysis. With more time, and a more in-depth optimisation of the operating parameters, the argument could be made that Adam would be a more favourable choice due to its lower computational requirements and similar performance. This should be investigated in a further study but is outside of our scope.

B. Input Variables

As mentioned in the methodology the initial criteria for viable sites is very steep and only met by a handful of site/weather station pairs. This limits the replicability of a potential low-cost, data-driven, fault-detection approach, as it could only be applied to a small number of installations. This section

investigates the effect of removing/including different input variables, aiming to evaluate the success of this approach on sites with more limited datasets. RF Model 2 was used for all analysis in this section, meaning input data for both the prediction interval and previous interval are included. Table VII shows MRE values for all sites upon the removal of individual weather variables, with the best modelling case, which uses all weather variables, included for reference. Clearly irradiance is

the most important variable, with average MRE increasing from 2.70% to 3.92% when it is not included. Interestingly the individual removal of all other variables does not seem to have a noticeable impact on model performance, with negligible changes in the MRE. This is not entirely unexpected as irradiance has a first-order effect on the panel power generation, whilst temperature, wind and humidity act as a second-order modifier to the panel efficiency.

Table VII - MRE for removal of specific variables

Site	MRE (%)						
	All	No Irradiance	No Previous Generation	No Solar Position	No Wind	No Temperature	No Relative Humidity
A	2.88	4.63	3.07	3.10	2.87	2.93	2.84
B	2.29	3.28	2.52	2.41	2.26	2.35	2.30
C	1.74	2.99	1.79	1.96	1.67	1.73	1.68
D	3.88	4.76	4.16	3.85	3.81	3.90	3.92
Average	2.70	3.92	2.89	2.83	2.65	2.73	2.68

Table VIII - MRE for different training times

Site	MRE (%)					
	1 Month	3 Months	6 Months	1 Year	2 Years	3 Years
A	4.26	3.89	3.74	3.08	2.99	2.88
B	3.19	2.96	2.85	2.45	2.37	2.29
C	2.79	2.62	2.52	1.91	1.81	1.74
D	4.65	4.32	4.38	3.91	3.87	3.88
Average	3.72	3.45	3.37	2.84	2.76	2.70

Table IX - MRE values for specific variable selection

Sites	MRE (%)			
	All	Irradiance	Irr + Gen	Irr, Gen + Sun Pos
A	2.88	4.14	3.63	2.99
B	2.29	3.33	2.78	2.35
C	1.74	2.81	2.41	1.70
D	3.88	5.13	4.27	4.01
Average	2.70	3.85	3.28	2.76

Table IX shows MRE values for each site for the three scenarios outlined in the methodology. As shown, including the previous generation and solar position with irradiance results in significant improvement in MRE. When only these 3 variables are included, modelling accuracy is comparable to the best-case performance for all sites, with average MRE only increasing from 2.70% to 2.76%. Clearly, accurate modelling of PV generation is still possible when only these 3 variables are known, and the inclusion of all other weather data, although helpful, does not seem to be fundamental.

C. Training Time

Random Forest Model 2 was again used to generate results with varying training set size. As seen in Table VIII, model

performance and training time show a clear correlation from 1 month to 1 year, with average MRE decreasing from 3.72% to 2.84% across this period. This may be due to the larger training sets including more of the natural seasonal variations in weather across a year; which is especially notable in the UK. For example, the one-month training set only includes weather data from December where weather is generally more overcast, and so may struggle more predicting generation than in the clearer summer months compared to the larger training sets. This may also explain the limited improvement in MRE upon increasing the training size above 1 year, with both the 2-year and 3-year training sets showing limited improvement in MRE across all sites. Although 1 year of historical data is still a large dataset, it can be confidently concluded that it can be used as an effective training set if 3 years of data are not available.

Another significant finding is that this approach will outperform the existing commercial model when trained on only 1 month of data, managing an average MRE of 3.72% compared to 5.42%. Even with a very limited dataset available, this data-driven machine learning approach can outperform an off the shelf modelling solution.

D. Weather Station Distance

Within the initial 5 km radius there is a clear correlation between a closer distance of the weather station and improved

modelling accuracy across all sites. Site C for example has the smallest distance of 2.15 km, and achieves an MRE of just 1.74%, while site D only reaches 3.88% with a distance of 5.15 km. This clear dependence suggests this approach may benefit from a regional Numerical Weather Prediction (NWP) model as its source of weather data, such as highlighted in [8]. Such a model may be able to give weather data even closer to the site and improve modelling accuracy as a result.

The effect of increasing weather station distance on MRE and RMSE in the extended 75 km radius is shown in Fig. 4. The most significant increase in MRE occurs when the distance is increased from 4.72 km to 22.2 km, resulting in an increase in MRE from 2.99% to 4.16%. After this point the increase in weather station distance seems to have a lesser impact, with both MRE and RMSE increasing only slightly over this range. At such large distances the weather data used may be significantly different to that at the site, and the models may instead be relying on the previous generation and solar position inputs to make its predictions.

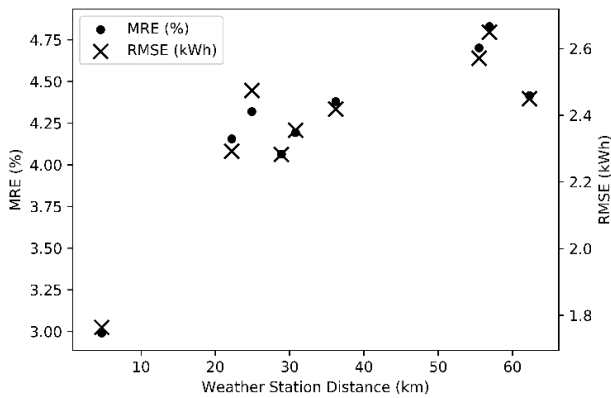


Fig. 4. MRE and RMSE as a function of weather station distance for Site A

V. CONCLUSIONS

This paper aimed to provide an in-depth overview of the dataset and modelling approach required for day-ahead, weather-based modelling of PV generation in relation to fault detection. The performance of RF, ANN and SVM algorithms were compared, with Model 1 and Model 2 inputs used for all. The effects of variable selection, training time and distance between weather stations and sites were also covered. The key learnings from this research are:

- RF Model 2 performed best, with an average MRE of 2.70%;
- All algorithms benefit from the Model 2 approach with weather and generation data from the previous time interval included;
- Irradiance, previous generation and solar position were found to be the most important variables to include;
- A 1-year training set represents the best trade off between dataset size and modelling accuracy;
- Modelling accuracy is strongly correlated to weather station distance, with distances less than 5 km preferred.

Overall this approach shows promise for developing a

remote data-driven, fault-detection solution for existing PV installations with a nearby weather station and historical generation records. Based on this context, the approach cannot be applied to new or planned sites without nearby weather data, making it a necessity to source reliable weather data. As further work, optimisation of the machine learning algorithms should be undertaken, as well as investigating the use of a regional NWP model for weather data. These improvements could both increase modelling accuracy and allow sites without nearby weather stations to use this approach.

VI. ACKNOWLEDGMENT

This research was supported by funds provided via the Imperial-Sainsbury's Supermarkets Ltd. Partnership.

VII. REFERENCES

- [1] S. Daliento, A. Chouder, P. Guerriero, A. Massi Pavan, A. Mellit, R. Moeini and P. Tricoli, "Monitoring, Diagnosis, and Power Forecasting for Photovoltaic Fields: A Review," *International Journal of Photoenergy*, vol. 2017, 13 pages, Jan 2017
- [2] A. Chouder and S. Silvestre, "Automatic supervision and fault detection of PV systems based on power losses analysis," *Energy Conversion and Management*, vol. 51, no. 10. Pp. 1929-1937, October 2010
- [3] A. Drews, A. de Keizer, H. Beyer, E. Lorenz, J. Betcke, W. van Sark, W. Heydenreich, E. Wiemken, S. Stettler, P. Toggweiler, S. Bofinger, M. Schneider, G. Heilscher and D. Heinemann, "Monitoring and remote failure detection of grid-connected PV systems based on satellite observations," *Solar Energy*, vol. 81, no. 4, pp. 548-564, April 2007
- [4] A. Al-Amoudi, L. Zhang, "Application of radial basis function networks for solar-array modelling and maximum power-point prediction," *IEE Proceedings - Generation, Transmission and Distribution*, vol. 147, no.5, pp. 310-316, September 2000
- [5] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F.J. Martinez-de-Pison and F. Antonanzas-Torres, "Review of photovoltaic forecasting," *Solar Energy*, vol. 136, pp. 78-111, October 2016.
- [6] J. Fonseca Jr, T. Oozeki, T. Takashima, G. Koshimizu, Y. Uchida and K. Ogimoto, "Use of support vector regression and numerically predicted cloudiness to forecast power output of a photovoltaic power plant in Kitakyushu, Japan," *Progress in Photovoltaics*, vol. 20, no. 7, pp. 874-882, November 2012
- [7] G. Graditi, S. Ferlito, G. Adinolfi, "Comparison of Photovoltaic plant power production prediction methods using large measured dataset," *Renewable Energy*, vol. 90, pp. 513-519, May 2016
- [8] M. Almeida, O. Perpinan, L. Narvarte, "PV power forecast using a nonparametric PV model," *Solar Energy*, vol. 115, pp. 354-368, May 2015
- [9] H. Yang, C. Huang, Y. Huang, Y. Pai "A Weather-Based Hybrid Method for 1-Day Ahead Hourly Forecasting of PV Power Output," *IEEE Trans. On Sustainable Energy*, vol.5, no.3, pp. 917-926, July 2014
- [10] Y. Zhao, B. Lehman, R. Ball, J. Mosesian and J. de Palma, "Outlier Detection Rules for Fault Detection in Solar Photovoltaic Arrays," in *2013 IEEE Applied Power and Electronics Conference*, pp. 2913-2920
- [11] Y. Zhao, B. Lehman, R. Ball, J. Mosesian and J. de Palma, "Decision tree-based fault detection and classification in solar photovoltaic arrays," in *2012 IEEE Applied Power Electronics Conference*, pp. 93-99
- [12] Syafaruddin, E. Karatepe, T. Hiyama, "Controlling of Artificial Neural Network for Fault Diagnosis of Photovoltaic Array," in *2011 International Conference on Intelligent System Applications to Power Systems*
- [13] D. Riley, J. Johnson, "Photovoltaic prognostics and health management using learning algorithms," in *2012 IEEE Photovoltaic Specialists Conference*, pp. 1535-1539
- [14] J. Shi, W. Lee, Y. Liu, Y. Yang, P. Wang, "Forecasting power output of photovoltaic system based on weather classification and support vector machine," in *2011 IEEE Industry Applications*
- [15] R. Noufi, "High Efficiency CdTe and CIGS Thin Film Solar Cells: Highlights of the Technologies Challenges," in *2006 IEEE Conf. on Photovoltaic Energy Conversion*