Forecasting OPV outdoor performance and degradation rates using machine learning

Tudur David
School of Electronic Engineering
Bangor, United Kingdom
eepa2a@bangor.ac.uk

Gabriela Amorim
CSEM
Belo Horizonte, Brasil
gabriela.amorim@csembrasil.com.br

Diego Bagnis
CSEM
Belo Horizonte, Brasil
diego.bagnis@csembrasil.com.br

Noel Bristow
School of Electronic Engineering
Bangor, United Kingdom
n.d.bristow@bangor.ac.uk

Sören Selbach
School of Electronic Engineering
Bangor, United Kingdom
sselbach@uni-osnabrueck.de

Jeff Kettle
School of Electronic Engineering
Bangor, United Kingdom
j.kettle@bangor.ac.uk

***Abstract*— Predicting the potential diurnal performance and degradation of organic photovoltaics (OPV) in outdoor conditions is of key interests for users and industrialists. Therefore, machine learning methods are herein employed in order to model and predict the diurnal variation in performance parameters. Subsequently, this allows the expected power output of the modules to be determined. Accurate modelling of the diurnal performance is achieved via a multilayer perceptron algorithm, trained using only the climatic conditions. Furthermore, the degradation rate of the OPV modules is predicted using a separate multivariate regression model. This allows for the main factors that influence the degradation to be found, which in rank order are the 1) Irradiance, 2) Module Temperature 3) Dew point, 4) UV dose, 5) humidity, 6) time, 7) wind speed, in rank order. Using the regression model for degradation, improved understanding of the sources of outdoor degradation is possible.**

Keywords—Outdoor monitoring, machine learning, performance prediction, energy yield, lifetime prediction

I. Introduction

As the global generation of renewable energy increases, the need to predict the expected power output from a photovoltaic (PV) source intensifies, to assist consumers and grid. The performance and stability of organic photovoltaic (OPV) devices and modules depends on a wide range of environmental factors and conditions [1]. There is an ever-increasing need to realize efficient and stable OPV technologies for power generation on a wide scale. However, due to the relatively short period, which OPV and other emerging technologies have been under development, their long-term stability at different stresses is not fully understood [2]. Several stress factors may influence the performance and subsequent power output [3], so outdoor monitoring provides a platform under which the combined effect of a multitude of different stresses can be tested [4, 5, 6, 7].

In this work, OPV modules have been subjected to outdoor testing conditions over the course of 6 months and the performance and climatic conditions have been simultaneously monitored. This results in a dataset with approximately 50,000 instances per OPV module under test. Subsequently, a machine learning (ML) methodology has been employed in order to predict and forecast future performances and stabilities based on climatic conditions. The acquired dataset consists of 20 attributes constituting the date, performance parameters and climatic conditions, as well as derived quantities such as the module temperature. These attributes have been used for training a variety of ML algorithms, such as those based upon RandomForest [8] and multilayer perceptron (MLP) [9] approaches. By training such algorithms, the performance of OPV modules can be predicted based on climatic conditions and consequently the degradation rates can be determined based on both the cumulative climatic stress factors. To the best of our knowledge, this is the first example where cumulative stress effects have been used to predict the degradation of OPV modules in real-world scenarios. The analysis presented forthwith is organized into two main topics: diurnal predictions of performance parameters and daily yield and degradation modelling via linear regression.

II. Methodology

Ten small area 21.6 cm^2 OPV modules have been tested under outdoor conditions and their current-voltage (I-V) characteristics are measured with a source measure unit (SMU) at 5 minute intervals using a setup reported in [7,10,11]. The modules are orientated southwards and at an inclination angle of 35º from the vertical. Low resistance cables were soldered onto the module contacts and fed into the SMU. The climatic conditions and irradiance levels are monitored using a

commercially bought Davis Vantage Pro2 weather station and EgniTEC PV monitoring system. A schematic of the outdoor monitoring system is shown in Fig.1. ML coding is developed using the Python programming language. Several core python coding packages are implemented, namely Pandas, Numpy, Matplotlib, Datetime, Scipy and Scikitlearn.

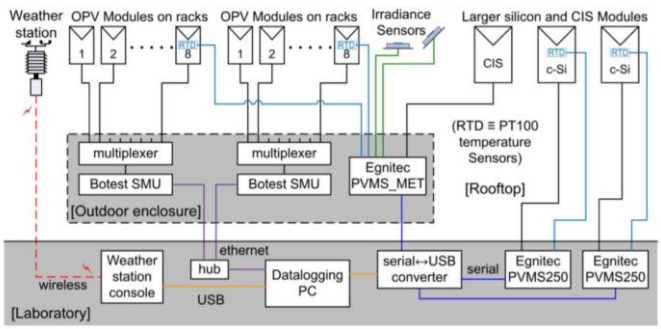

Fig. 1. Schematic of the outdoor monitoring system installed on the roof the School of Electronic Engineering, Bangor, North Wales [7].

The developed python program allows the acquired data to be compiled into a single dataset. All ML algorithms can be implemented through the use Scikitlearn. For training purposes, the ML algorithms are applied to an approximately stable period of 15 days, where the modules displayed very little degradation. The learnt models can then be applied to previously unseen data and the performance parameters predicted based on the climatic conditions: Irradiance, ambient temperature, module temperature, humidity, dew point, wind speed, UV index and UV dose. Subsequently the algorithms are trained on each climatic attribute individually and with all attributes, other than irradiance, UV index and UV dose. This is performed since knowledge of the irradiance levels and UV levels will not always be accessible for commercial buyers and industries. Throughout this study, the MLP algorithm is employed.

The MLP is a feedforward artificial neural network proposed by Rosenblatt in 1950 [12] and consists of an input layer, one or more hidden layers and an output layer. Each layer consists of several nodes, each connected to every node in the subsequent layer, forming a fully interconnected neural network, where a node can be considered as a neuron. Each node utilizes a nonlinear activation function to apply a weight to the input from a previous node. The activation function, usually takes the form of a sigmoid function [12],

$$f(x) = \frac{1}{1+e^{-x}}. \quad (1)$$

After the optimum weights have been determined *via* learning through the hidden layers, the vector of output weights is extracted from the output layer and utilized to make predictions based on the input values and produce a predictive model. In the case of forecasting the diurnal performance of OPV modules, the MLP model is first trained *via* supervised learning on certain set of training day data, based on the climatic conditions. This same model subsequently applied to unseen climatic conditions, corresponding to days in the future from the training set, in order to forecast the expected performance of the OPV modules, based on various climatic conditions. This procedure is employed within this investigation to predict the diurnal variation in the solar module performance parameters: open – circuit voltage (V_{OC}), short circuit current (I_{SC}), fill factor (FF), maximum power point current and voltage (I_{MPP} and V_{MPP}) and power conversion efficiency (PCE). Subsequently, this allows the daily yield, Y, to be calculated *via*,

$$Y = \sum_t I_{MPP} V_{MPP} t, \quad (2)$$

where t is the time delay between current – voltage scans and the summation is over the entire period of illumination for the modules.

In addition to the application of the MLP ML algorithm, principal componant analysis (PCA) is employed in order to acquire a qualitative understanding of the factors goverening the performance of OPVs in Outdoor conditions. PCA is a dimensionality reduction technique for analysing large datasets developed by Karl Pearson in 1901 [13]. PCA is an unsupervised machine learning algorithm used for explorative and qualitative assessment of datasets. Large datasets will possess a large number of dimensions, defined by the size of the dataset; a dataset consisting of n rows and m columns represents a data matrix having $n \times m$ dimension. This quantity is also referred to as the "dimensionality" of the dataset. In its entirety, the full dimensionality of the dataset fully describes the data. However, as the dimensionality increases, so does the difficulty in visualising and analysing the data since the number of variables to model becomes increasingly large. This makes interpretation of the dataset increasingly difficult [14].

Each dimension of the dataset will describe a certain proportion of the information contained within, quantified by the variance accounted for by each dimension. By visualising the dataset as an N dimensional ellipsoid, each axis of the ellipsoid is described by one of the dimensions. The larger the axis length, the greater the explained variance. The goal of PCA is to compute a new set of variables corresponding to a new set of dimensions. The dimensionality of the dataset can be reduced by removing redundant dimensions, keeping only the most important ones.

III. Computational Results and Discussion

In the first instance, principal component analysis (PCA) is employed in order to gain qualitative insight and determine the primary factors governing the four OPV performance metrics. Fig. 2 display the score and loadings plots for the PCA for PCE.

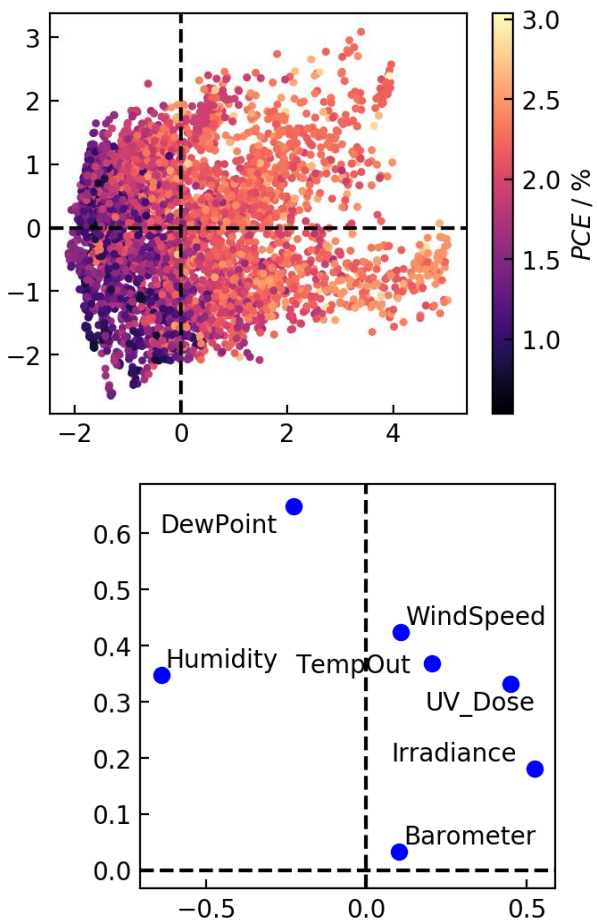

Fig. 2. Score and loadings plots for PCE when using PCA.

The score plot shown in Fig.2 illustrates the distribution of PCE values in relation to the loadings. By aligning the two plots, a comparison can be made where score values which align horizontally or vertically, correlate more strongly. Therefore, we can see that score values on the right-hand side vertically align with factors such as Irradiance, UV Dose, Outside Temperature. This means that these factors principally govern the higher PCE values observed during the study, as expected. In contrast, the lower PCE values, concentrated on the left-hand side are principally governed by the Humidity (i.e. when precipitation is high). Whilst these results are not overtly surprising, it does provide confidence in the dataset.

In addition, the same analysis can be conducted using the I_{SC}, V_{OC} and FF. Fig. 3 illustrates the score plots for each of the other performance parameters respectively.

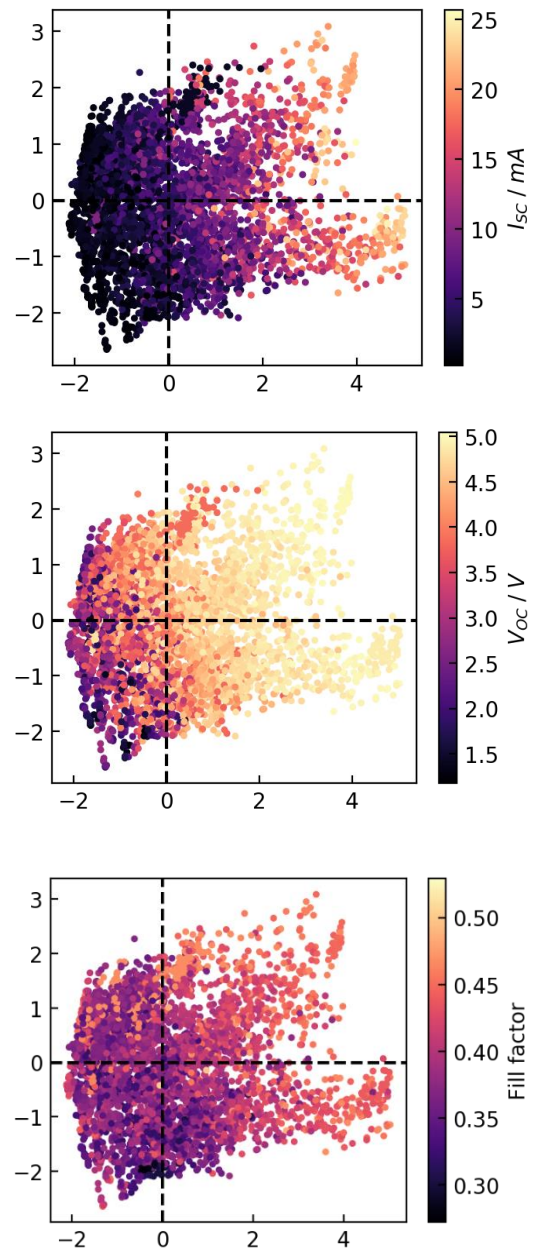

Fig. 3. Score plots for I_{SC}, V_{OC} and FF.

The main feature that can be identified from the plots shown in Fig. 3 is the difference in variability in the PCE gradation for each performance parameter. The greatest color gradation is observed for the I_{SC} which displays a significant variation from low I_{SC} values (<5mA) to high I_{SC} (>20mA) from left to right. This indicates a strong dependence on the I_{SC} with varying climatic conditions, similar to the PCE; high I_{SC} relates to high irradiance and low humidity, low I_{SC} relates to low

irradiance and high humidity, corresponding most probably to cloudy and damp conditions. A similar trend can be observed for the V_{OC}, albeit to a lesser extent than for I_{SC}, illustrating the weaker dependence of V_{OC} at the higher irradiance levels. However, this trend is not observed as strongly for the FF, where a much more uniform distribution of values is observed across all climatic conditions.

Furthermore, the correlation between different stress factors can be identified through the use of a correlation matrix which shows which pairs of factors correlate positively and which factors correlate negatively. This gives a holistic and qualitative understanding of how the climatic conditions relate to the performance operation of the modules. Fig. 4 shows the correlation matrix between all pairs of factors.

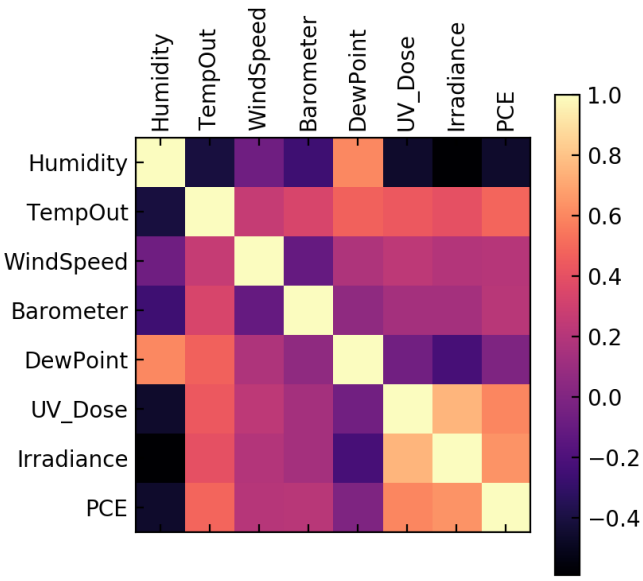

Fig. 4. Correlation matix for all climatic conditions governing module and PCE.

The correlation matrix shown in Fig. 4 demonstrates how different conditions relate to each other and how they relate to the PCE of the modules. The colour bar, on the right hand side, illustrates the strength of the correlation coefficient. For example, the PCE is most strongly governed by the Irradiance, UV dose and temperature. Most other strong correlations is as expected, such as the correlation of humidity and dew point, irradiance and UV dose and Irradiance and temperature.

The PCA provides quantitative analysis to understand how the climatic conditions affect the OPV performance. However, to develop models that predict OPV output a different approach is required. An MLP algorithm is employed in order to predict the diurnal variation in the OPV performance parameters (I_{SC}, V_{OC}, I_{MPP}, V_{MPP}, and FF). This is achieved by training the model on 14 days of diurnal data, where the modules are past the burn – in period and are stable in terms of performance. This means that the model is not adversely affected by the degradation of the modules. Fig. 5 illustrates the actual and predicted diurnal variations in each of the performance parameters over the course of the 14 day training period. Both the actual and predicted diurnal variations are shown for the training period and all are seen to fit the actual data with high accuracy, although the FF is seen to fit to a lesser extent, possibly due to the difficulty in modelling the low light intensity characteristics of the FF, where spikes are observed. This can introduce confusion in the algorithm where modelling the FF during night is difficult due to the very low light intensity. Regardless, the poor fitting of FF at low light conditions doesn't overlty affect the model as PV output at low light conditions is low anyway.

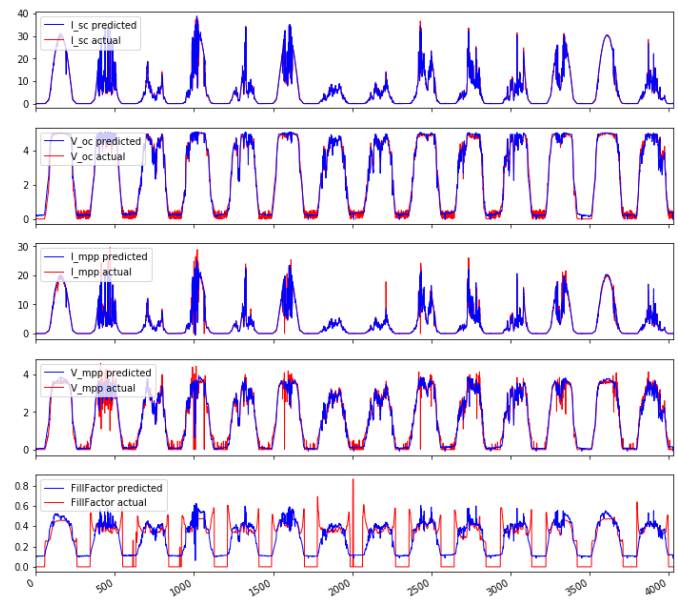

Fig. 5. Variation in each performance parameter over the course of 14 day period where modules display relative stability and little degradation.

After the model has been developed, based on the 14 day training period, the model is tested on the subsequent, 15th day. This constitutes an unseen testing day and the diurnal variation in the performance parameters is obtained purely on the variation in the climatic conditions and the irradiance. Fig. 6 Illustrates the diurnal variation in each parameter on the 15th testing day. Two different tests are displayed; one where the model is trained on 1) all climatic features and 2) only when the irradiance is used for training.

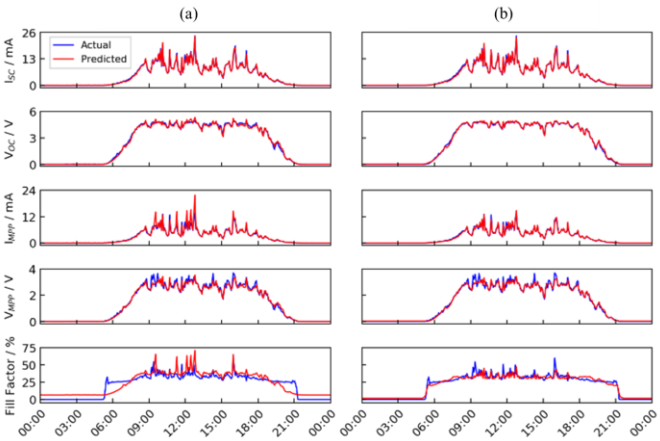

Fig. 6. Variation in each performance parameter over the course 15th testing day when trained on (a) all features, and (b) only irradiance.

From the plots shown in Fig. 6, it can be seen that for both training protocols, the predicted variation fits the actual data well for all performance parameters, although a small error when testing for the FF is observed at low irradiance levels. As this only occurs for the MLP applied when all climatic conditions are applied to the mode, this could be related to confusion created by training when using a large number of factors.

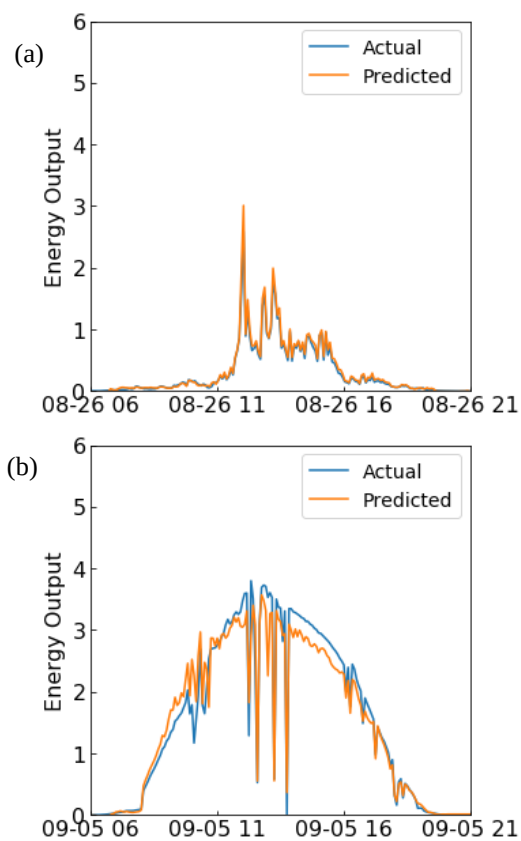

Fig. 7. (a) Actual and predicted diurnal variation in ernergy output on a cloudy day (26/8/2018). (b) Actual and predicted diurnal variation in ernergy output on a sunny day (5/9/2018).

Based upon the MLP, a quantification of the fitting can be made. In the first instance, the model is trained on a 10 day period (15/8/2018 to 24/8/2018), and the performance parameters are subsequently predicted based on the climatic conditions on the subsequent day. The model for predicting the yield displays an R^2 value of 0.93. The model is applied to two different days representing a cloudy day (26/8/2018) and a sunny day (5/9/2018). The actual and predicted variation in energy output can be seen in Fig. 7 (a) and (b). By integrating the curves with the respect to time, the daily yield for each day can be calculated. The actual and predicted daily yields for the two test days as well as the entire test period can be found in Table. 1, where the percentage error is also included.

Given the good overlap between predicted and actual diurnal performance, the calculation of energy yield can be extended for a further 2.5 months of data. The variation in actual and predicted daily yield is shown in Fig. 8 and demonstrates how the MLP model can be used to accurately predict the variation in daily yield based on 10 days of training for a period of 2 months. However, it is clear that the error between the actual and predicted values does increase over the course of the test. The increase in error could be attributed to two primary factors: the un-modelled degradation of the devices, and the change in the type of climatic conditions as the test proceeds into winter months; the model may be unable to predict the yield based on unseen types of weather.

Date	Actual Yield	Predicted Yield	% Error
26/8/2018	53.2Wh	60.8Wh	12.5
5/9/2018	276.6Wh	270.7Wh	2.2
25/8/2018–10/11/2018	7.93kWh	8.73kWh	10

Table. 1. Actual and predicted daily yields for a cloudy day (26/8/2018) and sunny day (5/9/2018). Total yield for test period also shown.

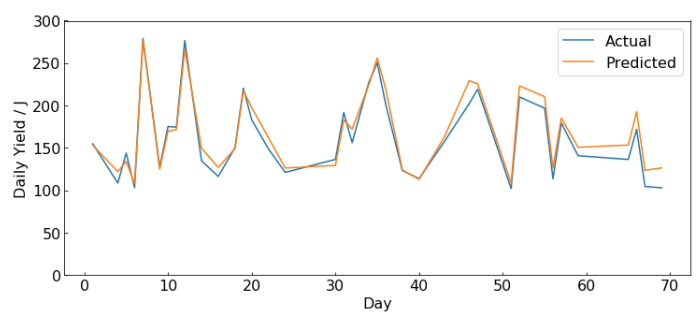

Fig. 8. Variation in daily yield for Sunny days over the course of 2.5 months.

To improve the energy yield forecasting, data analytics has also been applied to study the trends in PV degradation. As climatic conditions are time varying, we have based the study on cumulative climatic conditions. This allows the effect of prolonged exposure to different climatic conditions to be modelled; An example of the methodology used is shown in Fig. 9, where the diurnal irradiance as well as the cumulative dose of irradiation is shown (which is used for our model). Based upon using cumulative values for climatic conditions, a regression model has been used to relate the degradation rate to the climatic conditions and this is shown in Fig. 10. The acquired model for predicting the degradation rate yields an R^2 value of 0.93.

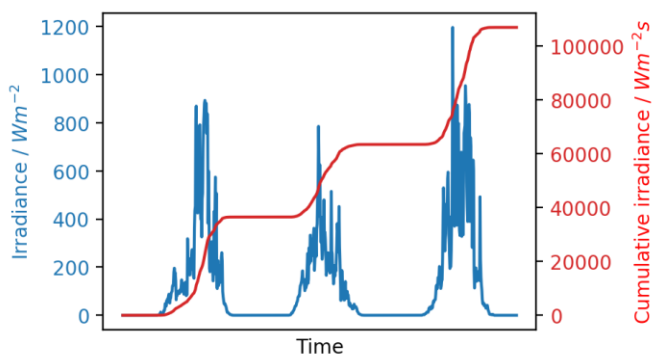

Fig. 9. Cumulative and diurnal variation in irradiance over the course of three days.

Based on the acquired MVA model, the t-values can be analysed to identify the most significant factors governing the stability of the modules. The regression t-values show that the factors that influence the degradation, in rank order are 1) Irradiance, 2) Module Temperature 3) Dew point, 4) UV dose, 5) humidity, 6) time, 7) wind speed. This information shows which environmental conditions are most detrimental for the OPV technology. For example, the prominance of irradiance as the most detrimental factor suggests that the light indued degradation of the active material is a governing factor in OPV stability.

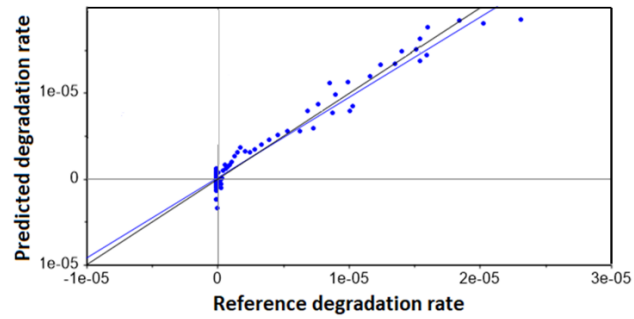

Fig. 10. MVA regression analysis for predicting the degradation rate of OPV modules based on the climatic conditions.

IV. Summary

Machine learning has been adopted as a means of forecasting the performance of OPV modules in outdoor conditions. The energy output of the modules has been calculated and a multilayer perceptron algorithm employed in order to predict the expected power output and can accurately model the variation in power for both sunny and cloudy days, highlighting the potential for this method to be employed in a wide variety of conditions and locations. This methodology provides invaluable information, since it allows users to estimate the expected power output of a module based on weather forecasts and make suitable contingencies. In addition, this method can provide a means of predicting the expected power output of new OPV devices, with relatively little testing data, and can thus provide a screening methodology for new emerging PV technologies. Furthermore, the degradation rate of the modules can be determined based on the cumulative climatic conditions. This allows the degradation state of the modules to be determined along with the diurnal performance based on the local climatic conditions. This, subsequently, allows for the long-term forecasting of the OPV performance based on future climatic conditions and how the OPV technology degrades. This method, therefore, provides rapid screening and feasibility assessment for new and emerging OPV technologies.

Acknowledgements

The Authors would like to thank the Solar Photovoltaic Academic Research Consortium II (SPARC II) project for supporting this work, gratefully funded by WEFO.

REFERENCES

[1] Zhang, Y., Samuel, I.D., Wang, T. and Lidzey, D.G., 2018. Current status of outdoor lifetime testing of organic photovoltaics. *Advanced Science*, *5*(8), p.1800434.

[2] Stoichkov, V., Kumar, D., Tyagi, P. and Kettle, J., 2018. Multistress Testing of OPV Modules for Accurate Predictive Aging and Reliability Predictions. *IEEE Journal of Photovoltaics*, *8*(4), pp.1058-1065.

[3] Kettle, J., Stoichkov, V., Kumar, D., Corazza, M., Gevorgyan, S.A. and Krebs, F.C., 2017. Using ISOS consensus test protocols for development of quantitative life test models in ageing of organic solar cells. *Solar Energy Materials and Solar Cells*, *167*, pp.53-59.

[4] Zhou, Z. and Carbajales-Dale, M., 2018. Assessing the photovoltaic technology landscape: efficiency and energy return on investment (EROI). *Energy & Environmental Science*, *11*(3), pp.603-608.

[5] Behera, M.K., Majumder, I. and Nayak, N., 2018. Solar photovoltaic power forecasting using optimized modified extreme learning machine technique. *Engineering Science and Technology, an International Journal*, *21*(3), pp.428-438.

[6] Gevorgyan, S.A., Heckler, I.M., Bundgaard, E., Corazza, M., Hösel, M., Søndergaard, R.R., dos Reis Benatto, G.A., Jørgensen, M. and Krebs, F.C., 2017. Improving, characterizing and predicting the lifetime of organic photovoltaics. *Journal of Physics D: Applied Physics*, *50*(10), p.103001.

[7] Bristow, N. and Kettle, J., 2018. Outdoor organic photovoltaic module characteristics: Benchmarking against other PV technologies for performance, calculation of Ross coefficient and outdoor stability monitoring. *Solar Energy Materials and Solar Cells*, *175*, pp.52-59.

[8] Breiman, L., 1999. Random forests. *UC Berkeley TR567*.

[9] Minsky, M. and Papert, S., 1969. An introduction to computational geometry. *Cambridge tiass., HIT*.

[10] Kettle, J., Bristow, N., Sweet, T.K., Jenkins, N., dos Reis Benatto, G.A., Jørgensen, M. and Krebs, F.C., 2015. Three dimensional corrugated organic photovoltaics for building integration; improving the efficiency, oblique angle and diffuse performance of solar cells. *Energy & Environmental Science*, *8*(11), pp.3266-3273.

[11] Bristow, N. and Kettle, J., 2015. Outdoor performance of organic photovoltaics: Diurnal analysis, dependence on temperature, irradiance, and degradation. *Journal of Renewable and Sustainable Energy*, *7*(1), p.013111.

[12] Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408.

[13] Pearson, K., 1901. Principal components analysis. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *6*(2), p.559.

[14] Ivosev, G., Burton, L. and Bonner, R., 2008. Dimensionality reduction and visualization in principal component analysis. *Analytical chemistry*, *80*(13), pp.4933-4944.