Received: 5 April 2021 | Revised: 3 June 2021 | Accepted: 8 July 2021
DOI: 10.1002/pip.3453

RESEARCH ARTICLE

PROGRESS IN PHOTOVOLTAICS WILEY

Predicting diurnal outdoor performance and degradation of organic photovoltaics via machine learning; relating degradation to outdoor stress conditions

Tudur Wyn David[1] | Gabriela Amorim Soares[2] | Noel Bristow[1] | Diego Bagnis[2] | Jeff Kettle[3]

[1]School of Computer Science and Electronic Engineering, Bangor University, Bangor, UK

[2]CSEM Brasil, Belo Horizonte, Minas Gerais, Brazil

[3]James Watt School of Engineering, University of Glasgow, Glasgow, UK

Correspondence
Jeff Kettle, James Watt School of Engineering, University of Glasgow, Glasgow G12 8QQ, Scotland.
Email: jeff.kettle@glasgow.ac.uk

Funding information
Solar Photovoltaic Academic Research Consortium II (SPARC II) project

Abstract

Accurate prediction of the future performance and remaining useful lifetime of next-generation solar cells such as organic photovoltaics (OPVs) is necessary to drive better designs of materials and ensure reliable system operation. Degradation is multifactorial and difficult to model deterministically; however, with the advent of machine learning, data from outdoor performance monitoring can be used for understanding the relative impact of stress factors and could provide a powerful method to interpret large quantities of outdoor data automatically. Here, we propose the use of artificial neural networks and regression models for forecasting OPV module performance and their degradation as a function of climatic conditions. We demonstrate their predictive capability for short-term energy forecasting of OPV modules, showing that energy yield can be predicted if climatic conditions are known. In addition, the model has been extended so that the impact of climatic conditions on degradation can be predicted. The combined model has been validated on unseen OPV module data and is able to predict energy yield to within 4% accuracy.

KEYWORDS
AI, degradation, ISOS, machine learning, OPV, organic photovoltaics, organic solar cells, outdoor testing

1 | INTRODUCTION

The ability to better understand the degradation of solar cells would be a game changer for users of the technology as they will be able to predict the useful lifetime in different climates and forecast energy yield from arrays. It will also support academic researchers working in the design of new materials and device improvements that provide long-term stability. In recent years, a succession of incremental, yet impressive, improvements have been reported by the organic photovoltaic (OPV) community, and recent reports show that module efficiencies as high as 11.7% with an area of 204 cm^2 are achievable.[1] Additionally, data from indoor tests have shown that OPV stability has also been improved, with some reports indicating that several thousand hours of lifetime are achievable during accelerated life testing (ALT).[2] ALT has been used for identifying optimal material sets,[3] benchmarking module stability[4] and also information on failure mechanisms.[5] However, it can only be used to estimate the operational lifetime when well-designed multistress and qualitative life tests are applied[6,7] and should be complemented with outdoor testing. Outdoor performance provides improved guidance, as the stress factors and levels outdoors are uncontrolled and time-varying. These stress factors often possess synergistic effects, which can accelerate or decelerate degradation rates. This is due to various interactions between stress factors, which results in competing degradation mechanisms. For example, it was shown that when an OPV is stressed simultaneously by both temperature and humidity, a greater degradation is observed than when each factor is increased individually.[8] Although outdoor monitoring provides the best way to measure long-term stability under outdoor conditions,[8–10] it does require patience as tests require several months, or years, of testing to achieve conclusive results.

One of the earliest reports on outdoor testing of OPV modules investigated the long-term performance for three different active layer materials (MEH-PPV:PCBM, P3HT:PCBM and P3CT-C_{60}). It was found that the MEH-PPV:PCBM modules displayed the fastest degradation followed by P3HT:PCBM and P3CT-C_{60}, which was the most stable.[11] The outdoor performance of OPVs has additionally been characterised by investigating their temperature and irradiance dependence. Early studies benchmarked the performance against c-Si modules, and OPVs were found to display lower performances in the morning and evening; this was attributed to the lower open circuit voltage (V_{OC}) and fill factor (FF) as a result of and energy barrier in the zinc oxide (ZnO) transport layer,[12] which was removed in newer roll-to-roll modules.[13] Furthermore, the performance parameters of OPV modules have been studied as a function of irradiance, and OPV modules were found to display distinct characteristics in comparison with crystalline silicon technologies, such as better low light performance, attributed to the differences in charge transfer mechanisms between the technologies.[13] By measuring the module temperature and calculating the deviance from the ambient temperature, the Ross coefficient was determined for OPV modules.[12]

More commonly, machine learning (ML) has been applied in conventional technologies; for example, the photovoltaic (PV) power output from a solar plant using silicon-based technologies can be forecasted based on the meteorological variables from a numerical weather forecast.[14] The method employed utilises a quantile regression forests ML algorithm to forecast the AC power. This provides a means of predicting the hourly AC power production 1 day ahead. The model treats the PV plant as a black box, making it a non-parametric PV model. The short-term output power for three grid-connected PV systems has been forecasted using an extreme learning machine (ELM). One hour ahead and day ahead, forecasting was achieved using the ELM. The model was trained and tested on the power output from the PV installation as well as other climatic conditions and accurate predictions for the PV power output was achieved over 3 months. The ELM models displayed R^2 values of 0.8938–0.9446 when predicting the power outputs of the different solar power plants.[15]

With advances in data analytics and ML, it is feasible that the data from outdoor monitoring could provide greater insight into the degradation mechanisms occurring within PVs and predict the useful lifetime of them. Developing time series methods in conjunction with 'real-time' change point detection could monitor failures from solar PV installations in real time. This would allow for rapid intervention to mitigate major module failures and allow for instantaneous failure analysis to be conducted when major faults occur. Forecasting using artificial neural networks (ANNs) could be applied for future prediction of energy yield using weather data. However, this approach could be extended further; ML could be used to predict the degradation of OPV modules.

In this work, OPV modules have been subjected to outdoor testing conditions over the course of 6 months, and the performance and weather conditions have been simultaneously monitored. Subsequently, a ML methodology has been employed in order to predict future performances and stabilities based on weather conditions. The acquired dataset consists of 20 attributes comprising of the date, performance parameters and weather conditions. These attributes have been used to train a type of ANN known as a multilayer perceptron (MLP) model.[16] By training such algorithms, the performance of OPV modules can be predicted based on weather conditions, and subsequently the degradation rates can be determined based on both the instantaneous and 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.

2 | EXPERIMENTAL AND COMPUTATIONAL METHODS

2.1 | Experimental methods

Ten OPV modules (21.6 cm^2), supplied by CSEM S. A, were tested under outdoor conditions according to the standard ISOS-O-2.[17] The samples were prepared using an inverted architecture and processed using a single roll-to-roll fabrication procedure using a Smart Coater SC09 from Coatema Coating Machinery GmbH, modified by CSEM Brasil. Samples were fabricated on a flexible substrate, which were sputtered with indium tin oxide/metal/indium tin oxide, supplied by Oike and used non-chlorinated solvents. All device layers were fabricated under ambient air conditions. An amine-based polymer was used as the electron transport layer and polyethylenedioxythiophene: polystyrenesulfonate (PEDOT:PSS) were used as the hole transport layer. Active layer formulations were acquired from Merck, which were fullerene-derivative based. Six coated strips were serially connected by a top silver electrode, 80% rich in silver and deposited using a flatbed semi-automatic screen printer. The samples were encapsulated using a multilayer PET-based barrier film with a water vapour trasnmission rate of the order of 10^{-3} g cm^{-2} day^{-1} from Mitsubishi. A Delo epoxy-based UV curable adhesive with barrier properties of 6 g cm^{-2} day^{-1} was applied using a R2R lamination machine, built in-house at CSEM Brasil, which uses a nip pressure to reach a thin and homogenous layer of glue of approximately 40 μm. The current–voltage (I-V) characteristics are obtained in situ, at 5 min intervals, using a Botest Systems GmbH source measure unit (SMU) located indoors, which resolution of 0.1 nA/1 mV. Low resistance feedthrough cables are soldered onto the modules and fed into the SMU. The modules are orientated at an inclination angle of 35° from the vertical and facing south. The weather conditions and irradiance are monitored using a commercially bought Davis Vantage Pro2 weather station (resolution temperature: 0.1°C, irradiance/UV dose 0.1 W/m^2, dew point 0.1°C, pressure 0.01 ATM) and EgniTEC PV monitoring system (resolution 0.01 mA/0.01 V) and a calibrated pyronometer that was orientated in plane with the OPV modules. A schematic illustration of the setup is shown in Figure 1.

The system, described above, allows the OPV module performance parameters to be gathered along with the weather conditions and irradiance levels. The acquired dataset consists of 18 attributes constituting the date, performance parameters and weather conditions, as well as derived quantities such as the module temperature. The performance parameters are derived from I-V scanning of the modules under outdoor illumination. The computed parameters are the V_{OC}, the short circuit current (I_{SC}), the FF, the maximum power point current and voltage (I_{MPP} and V_{MPP}, respectively) and the power conversion efficiency (PCE). To summarise, the attributes collected include the following: *Date Time of Day*, I_{SC}, V_{OC}, *FF PCE*, *Irradiance*, *Outside Temperature*, *Module Temperature*, *Relative Humidity*, *Dew Point*, *Wind Speed*, *Atmospheric Pressure*, *Rain Fall*, *UV Index* and *UV Dose*.

2.2 | Computational methods

ML coding has been undertaken using the Python programming language and has been implemented using *Scikitlearn* (https://scikit-learn.org/stable/). 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 weather conditions and combined with an outdoor degradation model. In addition to the data gathered from the outdoor monitoring system, the module temperature is derived and included as an attribute in the dataset. The module temperature can be determined from

$$T_{DELTA} = T_{Module} - T_{Ambient} = kG, \qquad (1)$$

where T_{DELTA} is the temperature difference between the module temperature (T_{Module}) and the ambient air temperature ($T_{Ambient}$). G is the irradiance, and k is the gradient, known as the Ross coefficient.[13] The Ross coefficient is influenced by wind speed,[18] as given by

$$k = a + be^{-cv}, \qquad (2)$$

where v is the wind speed and a, b and c are fitting coefficients derived from the exponential variation of the Ross coefficient as a function of wind speed as given in Bristow and Kettle[13] for OPV. The fitting coefficients are as follows: $a = 0.01007$, $b = 0.01947$ and $c = 0.4765$. This method for determining the module temperature was adopted due to the good similarity of the coefficients, a, b and c, to the previously reported values in Maturi et al.[18] as well as the good fitting between the variation in Ross coefficient variation and the fitted exponential curve. The fitting led to a root mean square error (RMSE) of 0.0001. The module temperature is preferred over the ambient temperature because this gives a better indication of the heating effect of the OPV module.

2.2.1 | Principal component analysis

Principal component analysis (PCA) is a dimensionality reduction technique for analysing large datasets.[19] PCA is an unsupervised ML 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 because the number of variables to model becomes

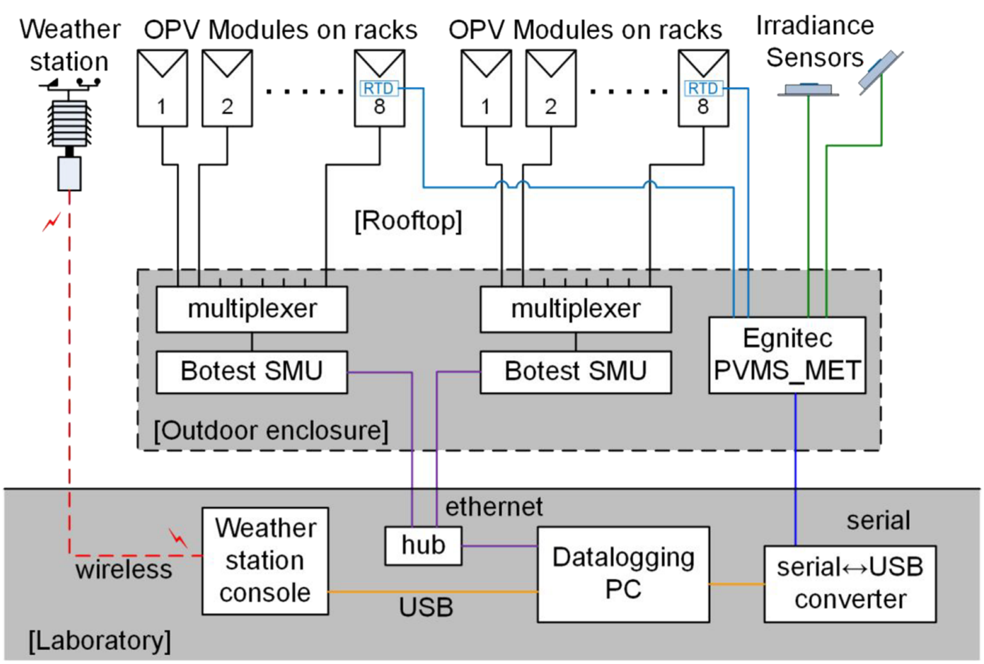

FIGURE 1 Schematic of the outdoor monitoring system installed on the roof of the School of Electronic Engineering, Bangor, North Wales[13] [Colour figure can be viewed at wileyonlinelibrary.com]

increasingly large. This makes interpretation of the dataset increasingly difficult.[20] 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.[21] The dimensionality of the dataset can be reduced by removing redundant dimensions, keeping only the most important ones.

2.2.2 | Multilayer perceptron

The MLP is a feedforward ANN 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.[21] Each node utilises 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.[16]

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 utilised 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 a certain set of training data, based on the diurnal weather conditions. This model is subsequently applied to unseen weather conditions in the future in order to predict the expected performance of the OPV modules. This procedure is employed within this investigation to predict the diurnal variation in the solar module performance parameters: V_{OC}, I_{SC}, FF, I_{MPP} and V_{MPP} and PCE. Subsequently, this allows the daily energy yield, Y, to be calculated using

$$Y = \sum_t I_{MPP} V_{MPP} t, \tag{3}$$

where t is the time delay between maximum power point measurements and the summation is over the entire period of illumination for the modules. In addition to the MLP model, the support vector regression (SVR) algorithm was tested along with the random forest algorithm. These algorithms, however, proved to be far less effective at accurately predicting the time-varying OPV performance parameters and yielded correlation coefficients (CCs) of less than 0.5.

2.2.3 | Multivariate linear regression

In order to predict the degradation rate of the OPV modules, a multivariate linear regression (MLR) approach is employed. The MLR method is achieved using a least-squares polynomial fitting model to represent the degradation rates of the modules in terms of the weather conditions, which they have been exposed to. This can be expressed as

$$y_i = f_0 + \sum_{k=1}^{z} f_k x_{nk} + \varepsilon_n, \tag{4}$$

where y_i is the predicted degradation rate, f_0 is a constant fitting parameter, f_k are the fitting parameters for each coefficient, $x_n k$ is the n th level of the k th predictor variable, and ε_n is the standard variance error.[22] The standard error of the predicted value is given by

$$\sigma_{\text{est}} = \sqrt{\frac{\sum \left(Y_{\text{real}} - Y_{\text{predicted}}\right)^2}{N}}, \tag{5}$$

where $\sigma_e st$ is the standard error of the predicted value, $Y_r eal$ is the actual value of the feature and $Y_p redicted$ is the predicted value, and N is the number of samples.[22] The standard error should be minimised during the MLR procedure in order to increase the accuracy of the prediction. Several conditions must be fulfilled in order to perform MLR regression on a dataset: (1) the dependent variable should be normally distributed with a constant variance, for each instance of the independent variable; (2) the dependent and independent variables should scale linearly with respect to each other; and (3) all observations should be independent (independence, linearity, normality and homoscedasticity).

Subsequently, MLR was used to relate the OPV module performance as a function of time to the instantaneous and cumulative weather conditions. The performance, PCE, is initially predicted at an irradiance of 250 Wm^{-2}. This irradiance level was selected in order to maximise the available data and is reasonably close to the mean irradiance in day light hours. This was because the dataset spanned both summer and winter months and, by selecting a lower irradiance, a comparison over the two seasons with a standardised irradiance could be made, enabling a larger dataset and a robust model. In addition to MLR, the RF algorithm was tested for degradation prediction. However, despite RF being effective at predicting the short-term degradation, the long-term performance could not be accurately predicted. In this work, it has been assumed that the degradation is constant as a function of irradiance. More accurate models could be formed that calculate the degradation rates at different irradiance levels. Nevertheless, further experimental work is needed to study how the degradation varies as a function of irradiance.

3 | RESULTS AND DISCUSSION

3.1 | PCA of performance and degradation rate

Initially, a qualitative examination of the performance and degradation rates of the modules over the course of the 6-month test is presented using PCA. PCA allows for an enhanced understanding of the variation of the performance and degradation due to the different weather conditions considered. Figure 2a–e shows the score plots for the PCE, I_{SC},

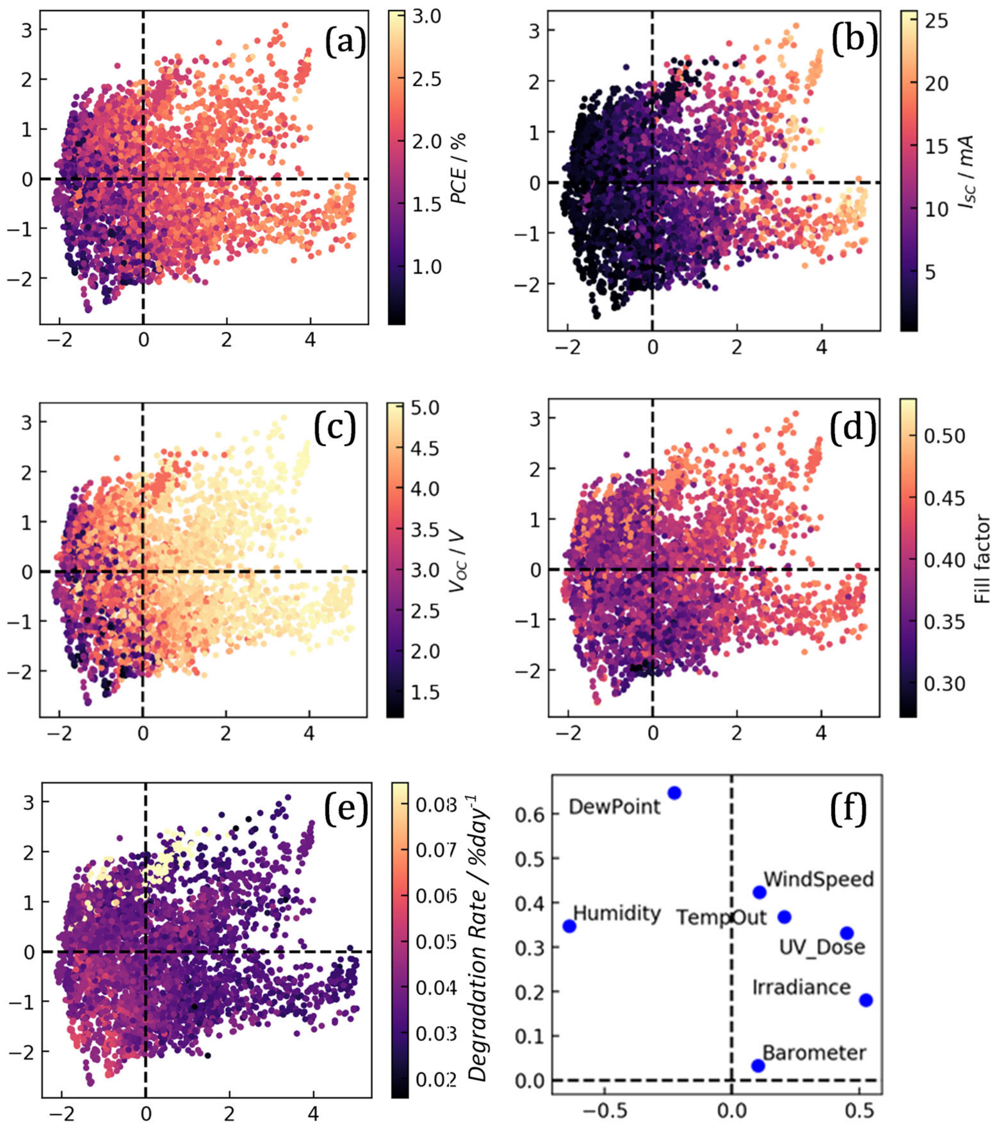

FIGURE 2 Score plot for (a) power conversion efficiency (PCE), (b) short circuit current (I_{SC}), (c) open circuit voltage (V_{OC}), (e) fill factor (FF) and (f) corresponding loadings plot for all modules under test for period of 6 months in outdoor conditions [Colour figure can be viewed at wileyonlinelibrary.com]

V_{OC}, FF and degradation rate respectively, and Figure 2f shows the corresponding loadings plot. In Figure 1a–e, the *x*-axis represents the scores of the first principal component, and the *y*-axis represents the scores of the second principal component.

The PCA allows the factors governing different features of the data to be visually represented. For example, attributes in the loadings plot that spatially align with score values correlate positively and vice versa. The score plot in Figure 2a shows that high PCE values reside on the right-hand side of the plot whereas low PCE values reside on the left. By comparing with Figure 1f, it can be seen that the high PCE values correlate positively with high irradiance, UV and outside temperature, as expected. Similarly, low PCE values correlate positively with high humidity and dew point (typically when it is raining, or early morning). Similar results can be seen for I_{SC}, V_{OC} and FF. For the degradation rate, shown in Figure 2e, the distribution is far more homogeneous. However, notable regions can be observed where high degradation rates correspond to high temperature and wind speed and to a lesser extent irradiance and UV index. The increased degradation due to elevated temperatures has been reported previously for OPV. For example, a study has been performed to develop a quantitative life test model for OPV where the lifetime was modelled as a function of temperature and a linear decay in T_{80} lifetime was found for increasing temperature.[7] Even though the direct effect of wind speed has not been reported in the literature, the effect of increased mechanical stress on the degradation has been identified where modules experiencing higher mechanical stress degrade more significantly.[23] The reduced degradation of OPVs due to lower temperature, irradiance and UV index has been reported by comparing the degradation of OPVs during summer and winter months by Bristow et al. where lower degradation was observed during winter months when temperature, irradiance and UV were lower.[13] The PCA serves a purpose for qualitative examination as it allows for identification of which factors should be included in the prediction and degradation models. It appears as though irradiance and UV are closely related, as are wind speed and temperature. Dew point and relative humidity have quite noticeable differences, probably indicating they are responsible for different failure modes in the solar cell. Dew point leads to the formation of liquid water on the surface or edges of modules, whereas during periods of high humidity, the water is gaseous in form. Based on the analysis performed, several attributes were removed for subsequent analysis. These included outside temperature, rain fall and UV dose.

3.2 | Diurnal performance prediction using MLP

In order to predict the diurnal variation in the performance parameters, the training of the MLP algorithm must be undertaken. This is performed by varying the training set duration, that is, the number of weeks of data that the model was trained on. Figure 3 illustrates the variation in the (a) CC and (b) normalised RMSE when predicting the PCE of the modules based on different durations of training and testing. The computed CC and RMSE correspond to the testing set prediction. The CC value can be between −1 and 1, where 1 represents a perfect positive correlation, 0 represents no correlation, and −1 represents a perfect negative correlation. The normalised RMSE value can be between 0 and 1 with 0 representing no error and 1 representing the maximum error in the test set prediction. This can be used to select the optimum training set duration. The colour map shows the variation in CC according to different durations of training and testing, and it is seen that, in order to achieve a high CC (>0.8), a minimum of approximately 2 weeks of training data is required. Only marginal improvements in diurnal PCE prediction are achieved beyond the 2-week training set. It should be noted that the CC colour map suggests that more training data is required for a lower amount of testing data. The CC value corresponds to the testing set prediction; with a smaller testing set, sporadic variations in weather conditions could influence the CC value more significantly due to unmodelled conditions. However, as the testing set increases, these sporadic changes could play a less significant role in relation to the already modelled conditions, and a higher CC value will be achieved for the larger testing set. The same can be seen when considering the variation in RMSE where a normalised value of less than 0.5 is achieved with 2 weeks of training data. When training the ML models, only the first 2 weeks of the data is used for training with the remaining data used for testing; the test data are not included in the training set.

Using the 2-week training set, the MLP algorithm has been used to predict the diurnal cycle in PCE, I_{SC}, V_{OC} and FF for an 'unseen' day based only on climatic conditions. The results are shown in Figure 4a–d, which shows the predicted and actual diurnal cycle in the performance parameters. The close correlation gives confidence and illustrates that the diurnal OPV module performance can be predicted based only on the weather conditions of that day. Subsequently, the diurnal power output is computed and is used to calculate the total energy yield over the course of each day using Equation 3 (see Section 4). Figure 4e illustrates the diurnal power output over the course of the same day, and Figure 4f shows the diurnal power output for the closest overcast day for comparison.

Table 1 quantifies the actual and predicted daily yields for the days shown in Figure 4. From inspection of the values obtained in Table 1, it is observed that the percentage error in the prediction made on a cloudy day (26/8/2018) is significantly larger (+12%) than the prediction made on a sunny day (5/9/2018) (−2%). This is not surprising as effects such as cloud lensing are difficult to predict. The cloudy day tested does display cloud lensing since there are spikes in the power output, indicating that there are spikes in irradiance. This implies that the day is not 'perfectly' overcast. This will introduce sporadic changes in irradiance levels due to irregular cloud coverage, which increases modelling complexity. The energy yield over the course of 6 months is also reported in Table 1, and in this case, a significant error of +20.9% is introduced. The primary reason for this error is related to the effect of degradation over the course of the test; Section 3.3 reports on how this can be reduced by the introduction of a degradation model.

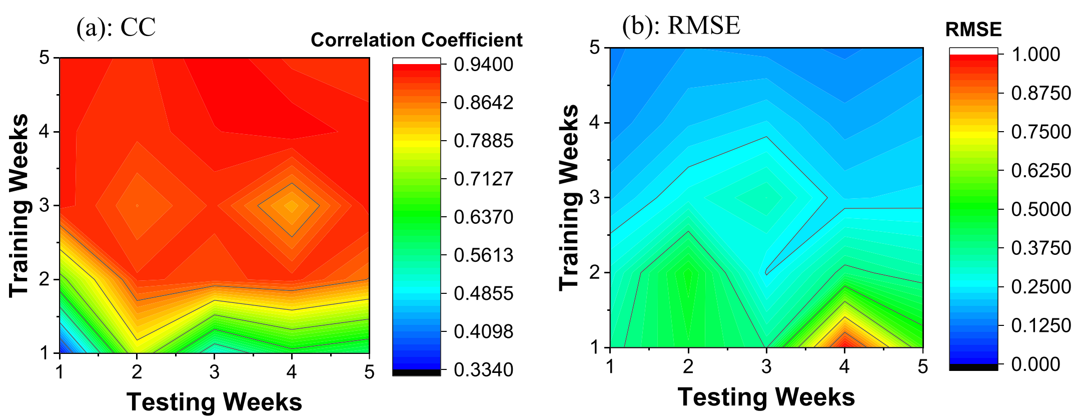

FIGURE 3 Two-dimensional plot showing the variation in (a) correlation coefficient and (b) normalised root mean square error (RMSE) when predicting the power conversion efficiency (PCE) based on various durations of training and testing [Colour figure can be viewed at wileyonlinelibrary.com]

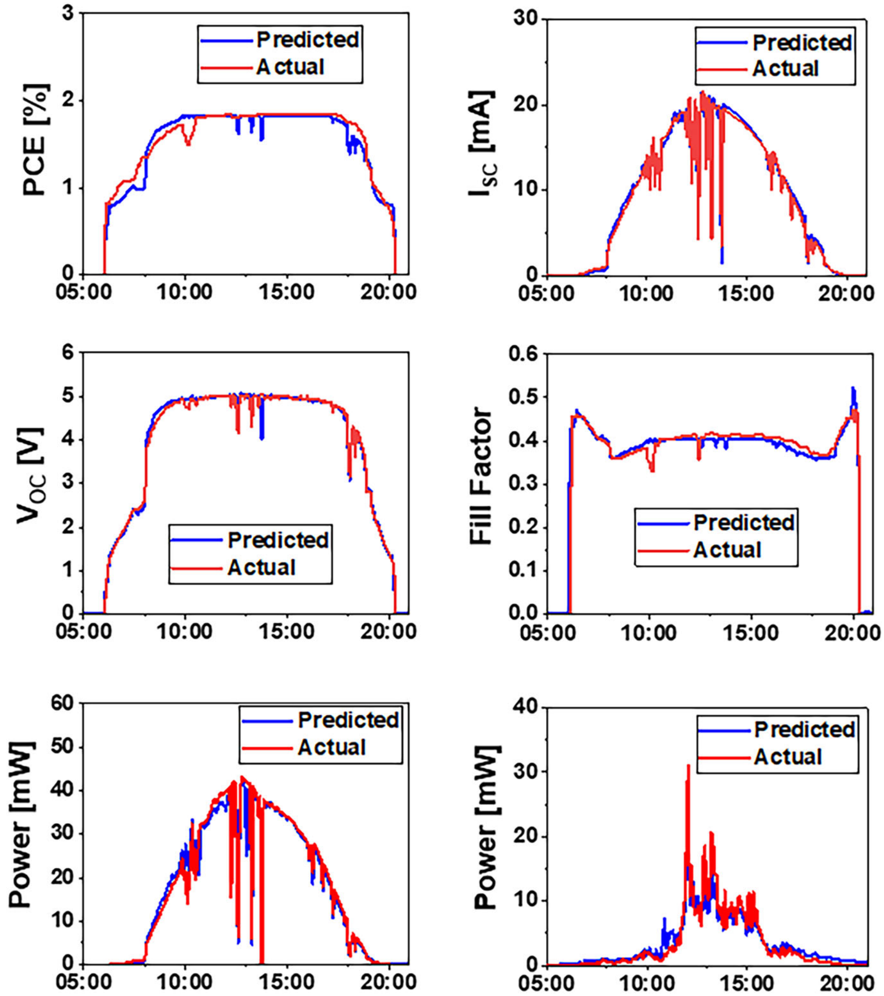

FIGURE 4 Comparison of actual and predicted diurnal variation in (a) power conversion efficiency (PCE), (b) short circuit current (I_{SC}), (c) open circuit voltage (V_{OC}) and (d) fill factor, (e) power (all tested on a sunny day after training period) and (f) power on cloudy day [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Actual and predicted daily energy yields for a cloudy day (26/8/2018) and a sunny day (5/9/2018) as well as the total energy yield for the entire test duration for summer (12/6/2018–30/11/2018) and winter (26/1/2018–26/4/2018)

Date	Actual energy yield	Predicted energy yield	Error
26/8/2018 (cloudy day)	53.2 Wh	60.8 Wh	+12.5%
5/9/2018 (sunny day)	276.6 Wh	270.7 Wh	−2.2%
12/6/2018–30/11/2018 (without degradation model)		28.9 kWh	+20.9%
12/6/2018–30/11/2018 (with degradation model)	23.9 kWh	24.9 kWh	+4.1%
26/1/2018–26/4/2018 (with degradation model)	13.2 kWh	13.7 kWh	+3.8%

3.3 | Degradation prediction of OPV modules

In order to predict the energy yield of OPV modules in outdoor conditions, the module degradation clearly needs to be accounted for. Modelling outdoor degradation is very complex, but it is significant to note that the degradation state of an OPV module at any point is time is dependent on the historical exposure of the modules to different stress factors. This is generally true for all PV systems. As a result, a 'cumulative dose' of the various weather conditions has been used as a factor for MLR analysis in order to predict the degradation of the OPV modules. MLR has been used for degradation prediction due to its suitability in time-series analysis. To illustrate this process, Figure 5 shows the instantaneous variation of each of the different weather conditions as well as the cumulative value of each condition with the full range of data; the thick blue line represents the rolling average value of each weather condition. Because the test progresses from summer into late autumn, different stress factors show different rates of change over the measurement period.

Figure 6a illustrates the prediction results where the median line for the PCE at 250 Wm^{-2} is shown along with the full range of data,

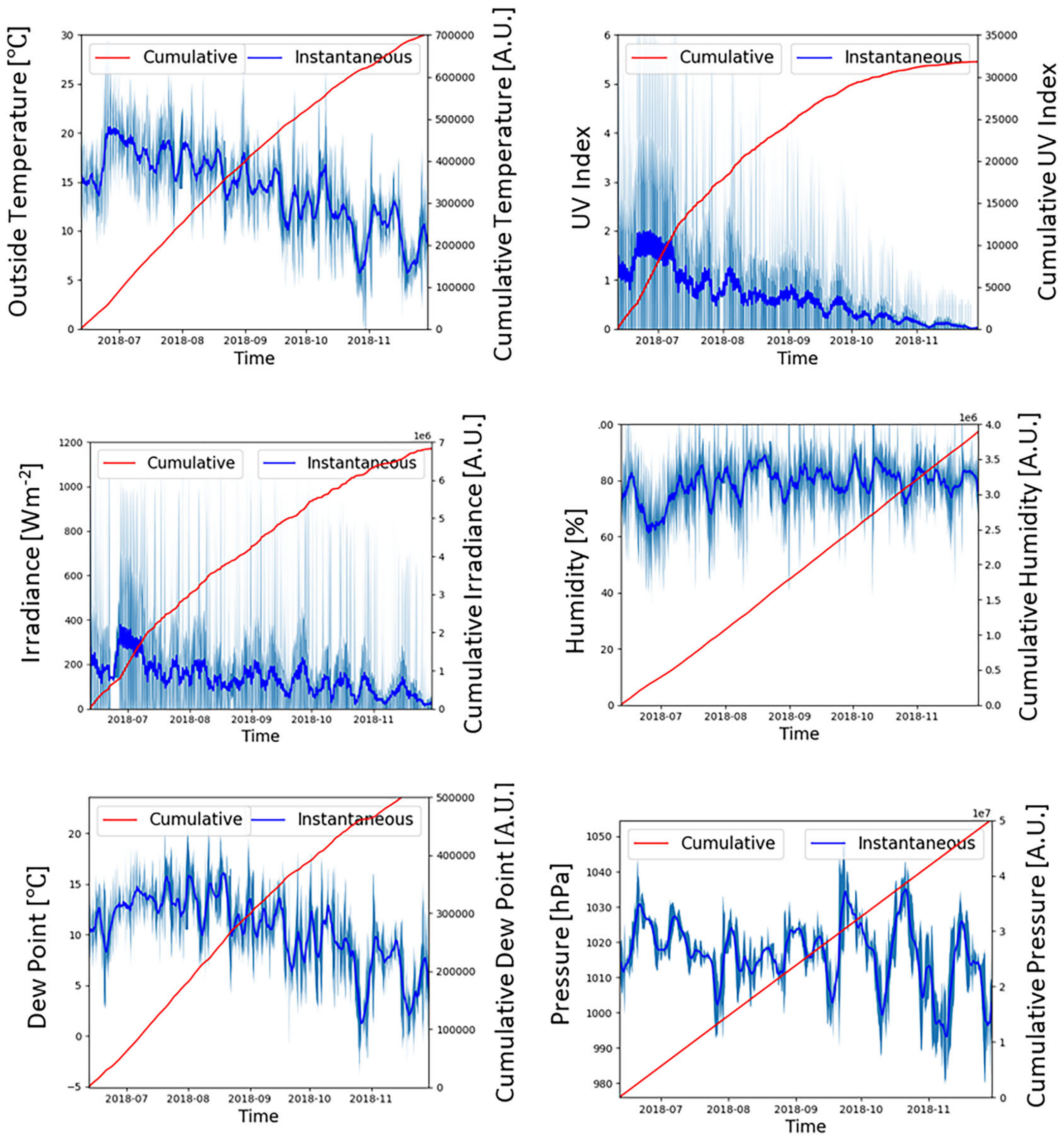

FIGURE 5 Variation in instantaneous and cumulative weather conditions over the course of 6 months [Colour figure can be viewed at wileyonlinelibrary.com]

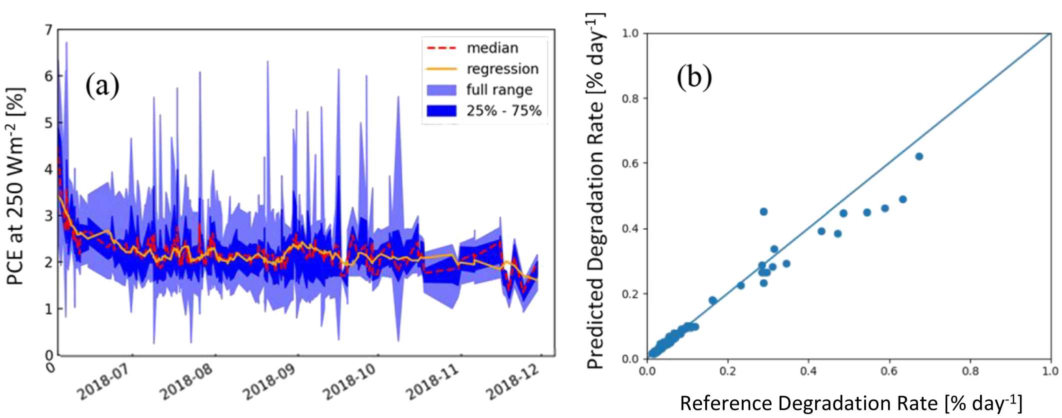

FIGURE 6 Forecasted degradation of organic photovoltaic (OPV) modules at 250 Wm^{-2}, showing the regression prediction, the rolling median of all the data, the full range of data and the interquartile range [Colour figure can be viewed at wileyonlinelibrary.com]

the interquartile range and the regression prediction. Both the instantaneous and cumulative weather conditions have been used in this analysis because the performance of the OPV module at a point in time is related to the weather conditions at that point in time, and the preceding weather conditions (which resulted in the degradation). Figure 6b shows the regression plot, illustrating the predicted degradation rates as a function of the reference degradation rate. A good fitting can be seen for low degradation rates; however, for higher degradation rates (which occur during phase), greater deviation in the regression can be seen where the model underpredicts the degradation. Therefore, this could suggest that the model does not fully account for the factors that govern the burn-in. The burn-in is an internal process occurring within the module's active layer and is related to the physical properties of the materials and process conditions.[24] Therefore, it is not surprising that the model fit is poorer during the burn-in phase as this is difficult to model based on the climatic conditions.

Determining the role and impact that the different weather conditions have upon OPV module degradation is highly desirable. By acquiring this information, better conclusions can be made regarding which factors need to be addressed so that the lifetime of OPV modules can be extended. As a result, significance testing was conducted in order to quantitatively determine the most significant factors. The null hypothesis (H_0) is defined stating that there is no relationship between the degradation rate and each of the instantaneous and cumulative factors. This will allow for a quantitative assessment of which factors govern the degradation rate significantly. A Pareto chart is shown in Figure 7a, which ranks the factors that have the biggest impact upon OPV module degradation. This shows the 1 *p*-value quantity, with the significance level set at 5%; when 1 *p* value is greater than 0.95, the attribute is statistically significant (shown by the vertical line).

The results of the significance analysis demonstrate that the statistically significant attributes are all cumulative, with the most significant attributes being air pressure, time, irradiance, UV index and module temperature, in rank order. Atmospheric pressure is found to be the most significant attribute. This can be understood from consideration of the effects of high pressure in relation to general weather conditions; high pressure is generally associated with long periods of sunny weather in the United Kingdom, which correspond to higher levels of irradiance, UV dose and absolute humidity, whereas low pressure is associated with poorer weather conditions with more cloud cover and rain. Therefore, although air pressure itself is not the cause degradation itself, it will be a governing factor for other weather

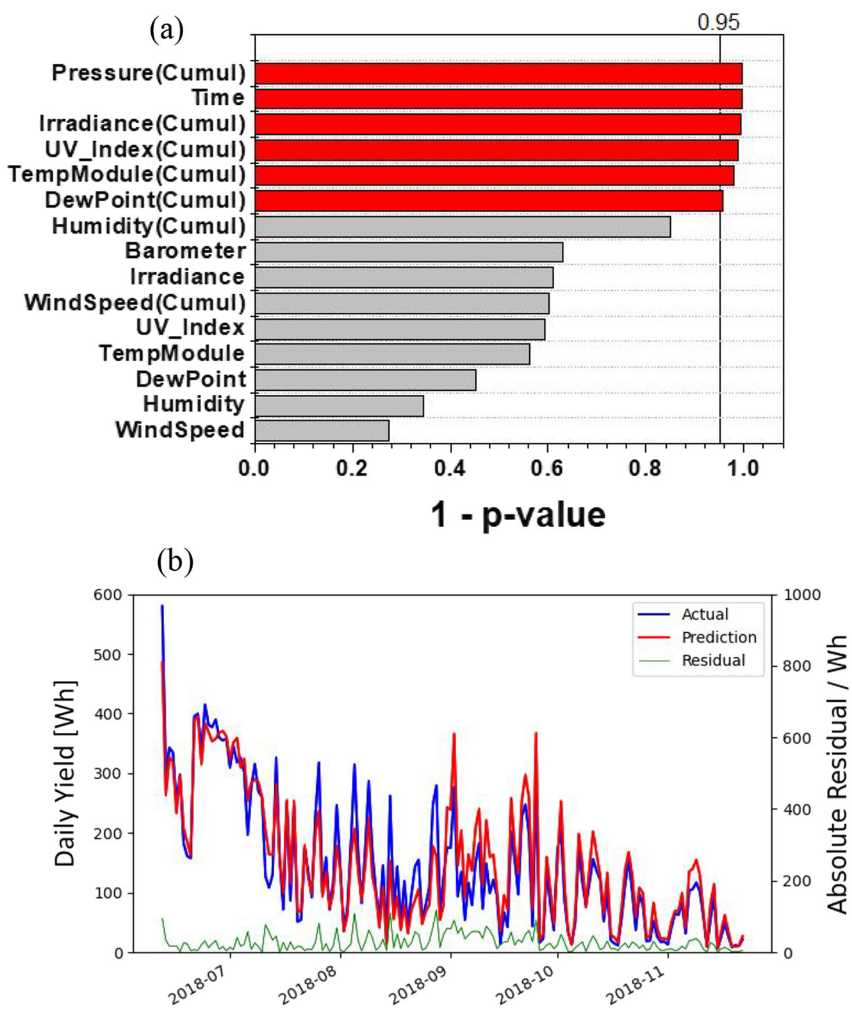

FIGURE 7 (a) Pareto chart showing the significance of each instantaneous and cumulative weather condition, given a 5% confidence interval. For clarity, the statistically significant attributes are shaded in red and the non-significant attributes shaded in grey. The significance level being tested has been represented by the vertical straight line. (b) Variation in actual and predicted daily energy yield over the course of the 5-month test. The predicted daily energy yield has been acquired by considering both the prediction made by a multilayer perceptron (MLP) model as well as consideration of the degradation rate of the modules. The green line represents the residual between actual and predicted daily energy yields [Colour figure can be viewed at wileyonlinelibrary.com]

conditions due to the strong correlation between the atmospheric pressure and the prevailing conditions, thus leading to greater degradation. In addition, module temperature is found to be a significant factor, and this is to be expected due to the instability of OPV modules at high temperatures, with module temperatures reaching over 60°C in summertime. Dew point is also found to be significant and is related to condensation, which can be very detrimental for OPV modules due to increased water infiltration and corrosion of contacts.

Based upon the degradation model, a method of improving the daily energy yield prediction approach can be conducted. The degradation rate analysis is included into the MLP diurnal prediction so that energy yield over several months of outdoor operation can be predicted. Figure 7b illustrates the actual daily energy yield variation over the course of the 5-month test as well as the predicted degradation, where the MLP model prediction has been combined with the prediction based on the degradation rate of the modules. A reasonable fitting is achieved via this method, although some discrepancy is still present. Nevertheless, the energy yield can be calculated from this method and compared with the original predicted energy, which did not feature the degradation model. Table 1 shows the comparison, and it is clear that the error in energy yield has been reduced substantially to only +4.1% (from +20.9%).

It is always important in data analytics to test the derived model on unseen data. As a result, we have applied the model to unseen data obtained on fresh modules (from the same manufacturing batch) tested in January to May 2018, as shown in the last row of Table 1. Using the climatic conditions as the input to the model, it is possible to estimate the energy yield from the model. In this case, the error in the energy yield prediction is 3.8%. The closeness of the predicted to the actual energy yield shows that this approach can be used for predicting outdoor performance of OPV modules accounting for the degradation. The accuracy of the model in predicting the degradation using the unseen data provides confidence in the applicability of the method when considering different seasonal effects. However, the 6-month period tested will inevitably not represent all possible weather conditions, and further tests should be conducted by applying the method to different climates and using longer testing periods such that a generalised model can be obtained.

4 | CONCLUSION

ML approaches have been applied to analyse the effects of different weather conditions on the performance and stability of OPV modules. This has been achieved using PCA, MLP and MLR algorithms. The PCA allowed for qualitative understanding of how the different weather conditions govern the performance. An MLP algorithm has been used to predict the diurnal variation in the OPV module performance parameters, where the weather conditions have been used as the attributes for prediction.

As a means of decreasing this relatively large error of +20.9%, the degradation of the modules is considered as a further attribute affecting the daily energy yield prediction during long-term forecasting. In order to model the module degradation, the cumulative dose of each weather condition has been used as the input for an MLR model. The degradation of the modules was subsequently predicted for a period of 6 months with high accuracy, and the variation in degradation rate over the course of 6 months also predicted. Using the combined model, energy yield was predicted to within 4.1% error. Further tests were conducted on unseen data and showed only a 3.7% error. Hypothesis testing is also conducted in order to assess the relative significance of the different weather conditions; all significant factors are found to be cumulative effects with atmospheric pressure, time and irradiance being the three most significant.

Therefore, a methodology has been introduced that allows the daily energy yield of OPV modules to be predicted 6 months into the future with very high accuracy. Even though the results of this analysis may not be true in general for all materials at all locations, the method presented is applicable to a wide range of technologies. If this method were to be applied to different climates using a range of technologies, a comprehensive understanding of solar cell prediction can be attained. This underlying methodology should be applicable to different OPV modules and also indoor degradation tests including multistress approaches. The ML algorithm is capable of learning new structured datasets based upon different OPV data producing accurate degradation predictions, given an initial period of testing. Nevertheless, it is worth stressing that a 'one ML outdoor degradation model fits all OPV types' is not yet available. To achieve this, this method must be applied to more datasets, corresponding to different technologies and different climates, so a generalised model could be achieved. This would allow the model to account for the different variables and conditions; this would enable more accurate techno-economic forecasts, improved OPV module design and prediction of stability around the world.

ACKNOWLEDGEMENTS

The authors would like to thank Sören Selbach for his help with the required Python programming and Dr. Vasil Stoichkov of Waterford IT for the setup of the modules tested and the data collection. JK, NB and TWD have been supported by the Solar Photovoltaic Academic Research Consortium II (SPARC II) project, gratefully funded by WEFO.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Diego Bagnis https://orcid.org/0000-0001-6933-6352
Jeff Kettle https://orcid.org/0000-0002-1245-5286

REFERENCES

1. Distler A, Brabec CJ, Egelhaaf HJ. Organic photovoltaic modules with new world record efficiencies. *Prog Photovolt: Res Appl*. 2020;29(1): 24-31. https://doi.org/10.1002/pip.3336

2. Du X, Heumueller T, Gruber W, et al. Efficient polymer solar cells based on non-fullerene acceptors with potential device lifetime approaching 10 years. *Aust Dent J*. Jan. 2019;3(1):215-226. https://doi.org/10.1016/j.joule.2018.09.001
3. Lloyd MT, Peters CH, Garcia A, et al. Influence of the hole-transport layer on the initial behavior and lifetime of inverted organic photovoltaics. *Sol Energy Mater Sol Cells*. May 2011;95(5):1382-1388. https://doi.org/10.1016/j.solmat.2010.12.036
4. Tanenbaum DM, Hermenau M, Voroshazi E, et al. The ISOS-3 inter-laboratory collaboration focused on the stability of a variety of organic photovoltaic devices. *RSC Adv*. Feb. 2012;2(3):882-893. https://doi.org/10.1039/c1ra00686j
5. Kettle J, Bristow N, Gethin DT, et al. Printable luminescent down shifter for enhancing efficiency and stability of organic photovoltaics. *Sol Energy Mater Sol Cells*. Jan. 2016;144:481-487. https://doi.org/10.1016/j.solmat.2015.09.037
6. Stoichkov V, Kumar D, Tyagi P, Kettle J. Multistress testing of OPV modules for accurate predictive aging and reliability predictions. *IEEE J Photovolt*. Jul. 2018;8(4):1058-1065. https://doi.org/10.1109/JPHOTOV.2018.2838438
7. Kettle J, Stoichkov V, Kumar D, Corazza M, Gevorgyan SA, Krebs FC. Using ISOS consensus test protocols for development of quantitative life test models in ageing of organic solar cells. *Sol Energy Mater Sol Cells*. Aug. 2017;167:53-59. https://doi.org/10.1016/j.solmat.2017.04.005
8. Gevorgyan SA, Heckler IM, Bundgaard E, et al. Improving, characterizing and predicting the lifetime of organic photovoltaics. *J Phys D Appl Phys*. 2017;50(10):103001. https://doi.org/10.1088/1361-6463/50/10/103001
9. Zhou Z, Carbajales-Dale M. Assessing the photovoltaic technology landscape: efficiency and energy return on investment (EROI). *Energ Environ Sci*. Mar. 2018;11(3):603-608. https://doi.org/10.1039/c7ee01806a
10. Behera MK, Majumder I, Nayak N. Solar photovoltaic power forecasting using optimized modified extreme learning machine technique. *Eng Sci Technol Int J*. Jun. 2018;21(3):428-438. https://doi.org/10.1016/j.jestch.2018.04.013
11. Katz EA, Gevorgyan S, Krebs FC. Out-door testing and long-term stability of plastic solar cells. *Eur Phys J Appl Phys*. 2007;36(3):307-311. https://doi.org/10.1051/epjap:2006159
12. Bristow N, Kettle J. Outdoor organic photovoltaic module characteristics: benchmarking against other PV technologies for performance, calculation of Ross coefficient and outdoor stability monitoring. *Sol Energy Mater Sol Cells*. Feb. 2018;175:52-59. https://doi.org/10.1016/j.solmat.2017.10.008
13. Bristow N, Kettle J. Outdoor performance of organic photovoltaics: diurnal analysis, dependence on temperature, irradiance, and degradation. *J Renew Sustain Energy*. Jan. 2015;7(1):013111. https://doi.org/10.1063/1.4906915
14. Almeida MP, Perpiñán O, Narvarte L. PV power forecast using a nonparametric PV model. *Solar Energy*. May 2015;115:354-368. https://doi.org/10.1016/j.solener.2015.03.006
15. Hossain M, Mekhilef S, Danesh M, Olatomiwa L, Shamshirband S. Application of extreme learning machine for short term output power forecasting of three grid-connected PV systems. *J Clean Prod*. Nov. 2017;167:395-405. https://doi.org/10.1016/j.jclepro.2017.08.081
16. Ramchoun H, Amine M, Idrissi J. Multilayer perceptron: architecture optimization and training multi-criteria learning and nonlinear optimization view project. *Int J Interact Multim Artif Intell*. 2016;4(1):1-26. https://doi.org/10.9781/ijimai.2016.415
17. Zhang Y, Samuel IDW, Wang T, Lidzey DG. Current status of outdoor lifetime testing of organic photovoltaics. *Adv Sci*. Aug. 2018;5(8):1800434. https://doi.org/10.1002/advs.201800434
18. Maturi L, Belluardo G, Moser D, del Buono. BiPV system performance and efficiency drops: overview on PV module temperature conditions of different module types. *Energy Proc*. 2014;48:1311-1319. https://doi.org/10.1016/j.egypro.2014.02.148
19. Pearson K. LIII. On lines and planes of closest fit to systems of points in space. *London Edinburgh Philos Mag J Sci*. Nov. 1901;2(11):559-572. https://doi.org/10.1080/14786440109462720
20. Bayne CK, Haswell SJ. *Practical Guide to Chemometrics*. Vol. 37, no. 2; 1995.
21. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev*. Nov. 1958;65(6):386-408. https://doi.org/10.1037/h0042519
22. Esbensen KH, Swarbrick B. *Multivariate Data Analysis: An Introduction to Multivariate Analysis, Process Analytical Technology and Quality by Design*. 6thed; 2018 Accessed: Aug. 05, 2020. [Online].
23. Magadley E, Teitel M, Peretz MF, Kacira M, Yehia I. Outdoor behaviour of organic photovoltaics on a greenhouse roof. *Sustain Energy Technol Assess*. Feb. 2020;37:100641. https://doi.org/10.1016/j.seta.2020.100641
24. Peters CH, Sachs-Quintana IT, Mateker WR, et al. The mechanism of burn-in loss in a high efficiency polymer solar cell. *Adv Mater*. Feb. 2012;24(5):663-668. https://doi.org/10.1002/adma.201103010

How to cite this article: David TW, Soares GA, Bristow N, Bagnis D, Kettle J. Predicting diurnal outdoor performance and degradation of organic photovoltaics via machine learning; relating degradation to outdoor stress conditions. *Prog Photovolt Res Appl*. 2021;1-11. https://doi.org/10.1002/pip.3453