

**Bangor University**

## **DOCTOR OF PHILOSOPHY**

### **Stability and performance assessment of organic photovoltaics using data analytics and machine learning**

David, Tudur

*Award date:*  
2021

*Awarding institution:*  
Bangor University

[Link to publication](#)

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*STABILITY AND PERFORMANCE ASSESSMENT  
OF ORGANIC PHOTOVOLTAICS USING DATA  
ANALYTICS AND MACHINE LEARNING*



PRIFYSGOL  
**BANGOR**  
UNIVERSITY

**Tudur Wyn David**

School of Computer Science and Electronic Engineering  
Bangor University

A thesis submitted in partial fulfilment for the degree of  
*Doctor of Philosophy*

June 2021

## DECLARATION OF AUTHORSHIP

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless, as agreed by the University, for approved dual awards.

Signed:

Date:

### Statement 1

I hereby declare that this thesis is the results of my own investigations, except where otherwise stated. All other sources are acknowledged by bibliographic references. The OPV modules tested in Chapter 5 were supplied by CSEM Brasil and the outdoor testing results for Belo Horizonte were performed at CSEM Brasil by Gabriela Amorim.

Signed:

Date:

### Statement 2

I agree to deposit an electronic copy of my thesis in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorised for use by Bangor University and where necessary have gained the required permissions for the use of third party material.

Signed:

Date:

## ABSTRACT

This thesis principally addresses the stability issues of organic photovoltaics (OPVs) by adopting machine learning approaches to identify the causes of degradation and identify the factors which contribute most significantly.

To achieve this, multiple datasets of OPV stability and performance parameters have been collated and studied. The first of these datasets contains data sourced from the literature containing detailed device information, including the structure, materials and testing conditions. The sequential minimal optimisation regression (SMOreg) machine learning algorithm was applied to the dataset, allowing for the ten most beneficial, and ten most detrimental, attributes to be identified for OPVs tested under ISOS-L and ISOS-D protocols. In doing so, the optimum configuration of OPVs can be identified, although further experimental verification is needed to confirm the results.

Machine learning was also applied to an OPV dataset derived from modules tested outdoors. The aim of this work was to allow the performance and stability of the modules to be correlated with the weather conditions. Principal component analysis illustrated that the most influential weather conditions were UV, irradiance, temperature and wind speed. A novel approach was taken for degradation forecasting, whereby the cumulative contribution of each weather condition was used as the predictive attributes, such that the total “dose” of each condition was accounted for. The degradation was forecasted by combining two machine learning algorithms. The method allowed the energy yield of OPVs tested outdoors to be predicted within 5% of the actual value using unseen data.

Finally, the database of OPV structures, performance and stability was extended to include the embodied energy of each material and processing step. From the initial performance, the stability and the embodied energy, the ‘net energy output’ was calculated which assesses the predicted energy yield of the OPV and subtracts the embodied energy. The SMOreg algorithm was used to determine the attributes which most significantly govern the embodied and net energies. A genetic search clustering algorithm (GenClust++) was used to identify the optimum combinations of the materials to maximise the net energy. This provided a means of finding new and improved combinations of materials and structures which have a lower environmental footprint whilst maximising the energy generation potential.



## ACKNOWLEDGEMENTS

Firstly, I would like to thank Dr Jeff Kettle, my supervisor, for giving me this opportunity and his tireless support and guidance over the last four years as well as for the opportunities which he has provided me.

I would like to thank the Solar Photovoltaic Academic Research Consortium II (SPARCII) for providing financial support for my research from the Welsh European Funding Office (WEFO). I would also like to thank all my colleagues who are part of SPARCII.

I would like to thank my external collaborators: Jesper Jacobsson from Helmholtz-Zentrum Berlin, Gabriela Amorim and Diego Bagnis from CSEM Brasil, Helder Anizelli and Edson Laureto from Universidade Estadual de Londrina and Giselle Benatto from the Technical University of Denmark (DTU).

I would like to thank Dr Paul Sayers and Ben Assinder for their technical support and for ensuring the effective operation of the clean room and equipment at the School of Electronic Engineering. Furthermore, I would like to thank all my colleagues at Bangor for making this experience so enjoyable: Noel, Priyanka, Vasil, Dinesh, Luke, James, Jake, Nissar, Helder, Douglas, Sören, Aeron, Rakesh, Bing, Fahad, Amjad, David, Paul, Ben, Iwan, Kevin, Nawaf, Josh, Cameron and Bill.

Finally, I would like to thank my whole family for their support. In particular I would like to thank my parents and brother for their unwavering support and encouragement.

## PUBLICATIONS

David, T.W., Anizelli, H., Tyagi, P., Gray, C., Teahan, W. and Kettle, J., 2019. Using large datasets of organic photovoltaic performance data to elucidate trends in reliability between 2009 and 2019. *IEEE Journal of Photovoltaics*, 9(6), pp.1768-1773.

Tyagi, P., David, T.W., Stoichkov, V.D. and Kettle, J., 2019. Multivariate approach for studying the degradation of perovskite solar cells. *Solar Energy*, 193, pp.12-19.

Anizelli, H., David, T.W., Tyagi, P., Laureto, E. and Kettle, J., 2020. Enhancing the stability of perovskite solar cells through functionalisation of metal oxide transport layers with self-assembled monolayers. *Solar Energy*, 203, pp.157-163.

David, T., Amorim, G., Bagnis, D., Bristow, N., Selbach, S. and Kettle, J., 2020, June. Forecasting OPV outdoor performance, degradation rates and diurnal performances via machine learning. In *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)* (pp. 0412-0418). IEEE.

Huang, H., Coote, T., Bristow, N., David, T.W., Kettle, J. and Todeschini, G., 2020, September. Development of an improved computer model for organic photovoltaic cells. In *2020 9th International Conference on Renewable Energy Research and Application (ICRERA)* (pp. 78-82). IEEE.

Soares, G.A., David, T.W., Anizelli, H., Miranda, B., Rodrigues, J., Lopes, P., Martins, J., Cunha, T., Vilaça, R., Kettle, J. and Bagnis, D., 2020. Outdoor performance of organic photovoltaics at two different locations: A comparison of degradation and the effect of condensation. *Journal of Renewable and Sustainable Energy*, 12(6), p.063502.

David, T.W., Anizelli, H., Jacobsson, T.J., Gray, C., Teahan, W. and Kettle, J., 2020. Enhancing the stability of organic photovoltaics through machine learning. *Nano Energy*, 78, p.105342.

Todeschini, G., Huang, H., Bristow, N., David, T.W. and Kettle, J., 2020. A Novel Computational Model for Organic PV Cells and Modules. *International Journal of Smart Grid-ijSmartGrid*, 4(4), pp.157-163.

Tudur Wyn David, Gabriela Amorim Soares, Noel Bristow, Diego Bagnis, Jeff Kettle, Forecasting diurnal outdoor performance and degradation of OPVs via machine learning. *Progress in Photovoltaics (Under Review)*.

Tudur Wyn David, Gisele Alves dos Reis Benatto, and Jeff Kettle, Use of machine learning to quantify the trade-off between performance, stability and environmental impact in organic solar cells. *(Under Review)*

# Contents

Declaration of Authorship .....	2
Abstract.....	3
Acknowledgements .....	4
Publications.....	5
Chapter 1. Introduction .....	11
1.1 Global need and use of renewable energy .....	11
1.2 Outline and Structure of Thesis .....	14
Chapter 2. Literature Review and Scientific Background .....	16
2.1 Solar Radiation Spectrum and Current – Voltage Characteristics.....	16
2.1.1 Solar Radiation Spectrum .....	16
2.1.2 Current – Voltage characteristics of Organic Photovoltaic Solar Cells .....	18
2.2 Introduction to OPV and PSC development .....	21
2.3 OPV Operation .....	22
2.3.1 Charge generation and dissociation in OPVs.....	23
2.3.2 OPV Device Structures. ....	25
2.3.3 OPV architectures .....	27
2.3.4 High Performance and Stability OPV devices.....	28
2.4 Stability Tests.....	29
2.4.1 ISOS stability test protocols .....	29
2.4.2 Stability issues in OPVs .....	31
2.4.3 Outdoor monitoring of OPVs .....	41
2.5 Application of ML in OPV and PSC development .....	48
2.5.1 Use of Data Analytics for Material Studies .....	49
2.5.2 Application of ML in Forecasting .....	53
Chapter 3. Experimental and Computational Methods.....	58
3.1 Introduction .....	58
3.2 CSEM module preparation.....	58
3.3 Characterisation Techniques .....	61

3.3.1 I – V Characterisation .....	61
3.3.2 Absorption Spectroscopy.....	61
3.3.3 Scanning Electron Microscopy.....	61
3.3.4 Electrochemical Impedance Spectroscopy.....	62
3.3.5 Environmental Test Chambers For accelerated testing of PSCs.....	62
3.4 Data Analytical and Machine Learning Approaches .....	63
3.4.1 Principal Component Analysis .....	63
3.4.2 Multivariate Linear Regression .....	66
3.4.3 Sequential Minimal Optimisation Regression.....	68
3.4.4 Multilayer Perceptron .....	73
3.4.5 Random Forest.....	75
3.4.6 Hypothesis Testing and Statistical Significance .....	76
3.4.7 Software for Data Analysis and Machine Learning.....	77
3.5 Outdoor Measurement.....	77
3.5.1 Outdoor Measurement System .....	77
Chapter 4. Enhancing the Indoor Performance and Stability of Emerging Photovoltaics through Machine Learning.....	80
4.1 Introduction .....	80
4.2 Methodology.....	81
4.2.1 Overview of Data Acquisition and Ethics.....	81
4.2.2 Data Acquisition and Description of Data Format and Categories .....	82
4.3 Results and Discussion.....	85
4.3.1 Initial Data Exploration.....	85
4.3.2 Application of the Multivariate Linear Regression for understanding stability data.....	86
4.3.3 Algorithm optimisation.....	91
4.3.4 Application of the SMOreg algorithm for understanding stability data after dataset optimisation .....	98
4.3.5 Using the SMOreg algorithm for predicting the initial Stability, $E_0$ .....	107

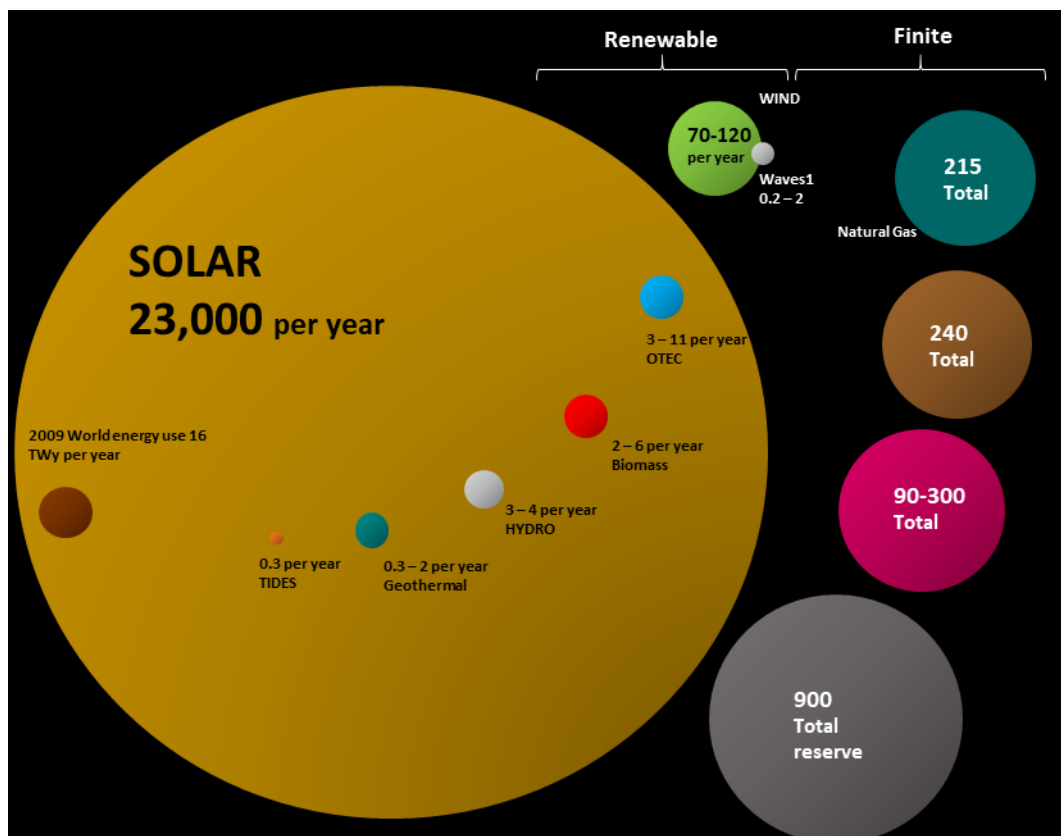
4.4 Challenges and Perspectives.....	111
4.5 Conclusion.....	112
Chapter 5. Outdoor testing and forecasting of Diurnal Performance and Degradation Rates.....	114
5.1 Introduction .....	114
5.2 Results and Discussion.....	115
5.2.1 Outdoor testing of OPV modules at different latitudes .....	115
5.2.2 Machine Learning Methodology .....	122
5.2.3 Principal Component Analysis of Performance and Degradation Rate....	124
5.2.4 Diurnal Performance Prediction using MLP .....	127
5.2.5 Degradation Forecasting of OPV Modules .....	131
5.2.6 Degradation Forecasting of Other Technologies – Perovskite Solar Cells .....	139
5.3 Conclusion.....	145
Chapter 6. Life – Cycle Assessment of Organic Photovoltaics with the application of Genetic Clustering Algorithms .....	146
6.1 Introduction .....	146
6.2 Methodology.....	147
6.2.1 Material Embodied Energy Determination .....	147
6.2.2 Genetic Algorithm Clustering .....	152
6.3 Results and Discussion.....	155
6.3.1 Initial Data Exploration.....	155
6.3.2 Classification using SMO algorithm.....	157
6.3.3 Application of SMOreg algorithm.....	159
6.3.4 Clustering using Genetic Algorithms .....	168
6.4 Conclusion.....	172
Chapter 7. Conclusions and Avenues for Future Research.....	173
7.1 Conclusions from Results Chapters.....	173
7.2 Future Work .....	175
References .....	178
Appendix A Enhancing the Indoor Performance and Stability of Emerging Photovoltaics through Machine Learning – Additional Information.....	192

Appendix B Life Cycle Assessment of Organic Photovoltaics with the application of Genetic Clustering Algorithms – Additional Information.....	193
Appendix C Additional Studies – HTL and Carbon-based PSCs.....	198
C.1 Methods: Inverted PSC fabrication.....	198
C.2 Methods: Carbon Based PSC fabrication .....	200
C.3 Results and Discussion – HTL Screening .....	201
C.4 Results and Discussion – Intercalated Carbon-Based Perovskites.....	205
C.4.1 Device Characterisation.....	205
C.4.2 Device Stability.....	211
Appendix D Additional Studies – Large area OPVs.....	221
D.1 Results and Discussion.....	221
D.1.1 Diurnal Performance and Stability .....	221

# Chapter 1.Introduction

## 1.1 Global need and use of renewable energy

Global warming has presented humanity with one of the most difficult challenges it has faced. Human driven climate change first came to the forefront of attention in 1988 when The Intergovernmental Panel on Climate Change (IPCC) was established in order to provide policymakers with the scientific assessments on the state of climate change in order to take mitigating steps to minimise the effects [1]. The last 20 to 30 years has seen one of the most notable change in attitudes, with the Paris agreement of 2015 being one of the most definitive global statements on the solidarity of nations to address the issue of global warming [2].

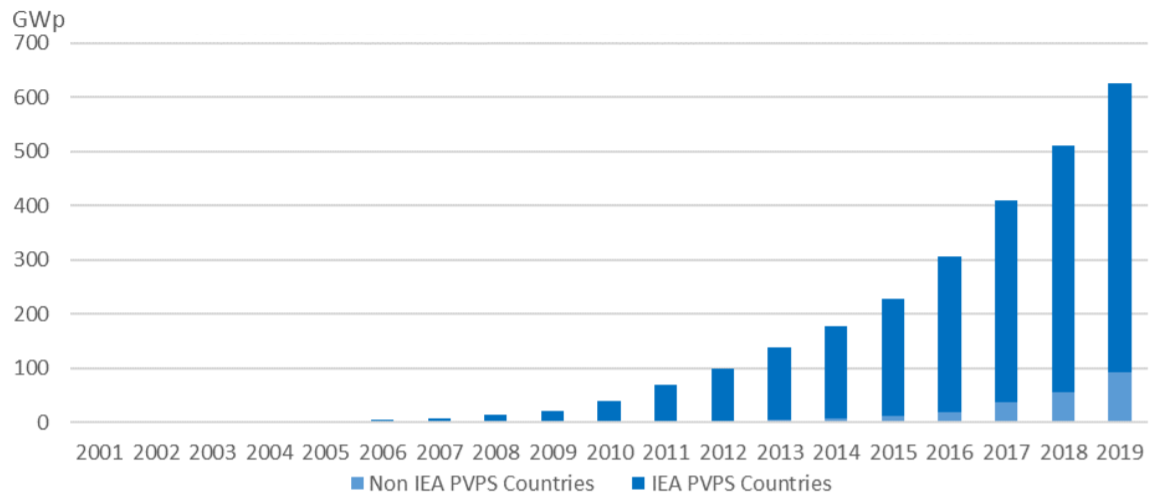


*Figure 1.1: Schematic illustration of the energy sources available to support the Earth and the current world energy consumption redrawn from [3].*

Renewable energies could provide a solution to the problem of global warming. For example, solar energy, wind energy, tidal and geothermal. In Figure 1.1, the different



sources of energy and the corresponding available energy potential measured in TW-yr. The volume of each sphere in Figure 1.1 represents the energy obtainable from each energy source per annum, in comparison to the economically exploitable reserves of fossil fuels and demonstrated that the solar energy reserves far exceeds the reserves from any other source.



Source: IEA PVPS

*Figure 1.2: Global Evolution of Cumulative PV Installations [4].*

A recent report in 2019 has further shown how solar energies are now being exploited for photovoltaic (PV) markets [4]. The cumulative energy production from PV shows an exponential growth, as illustrated in Figure 1.2. The report shows that the PV penetration into the energy sector exceeded 100 GW yearly installation for the third year in a row, further highlighting the promise and general move towards PV throughout the world, with a 12% annual growth in the total PV market.

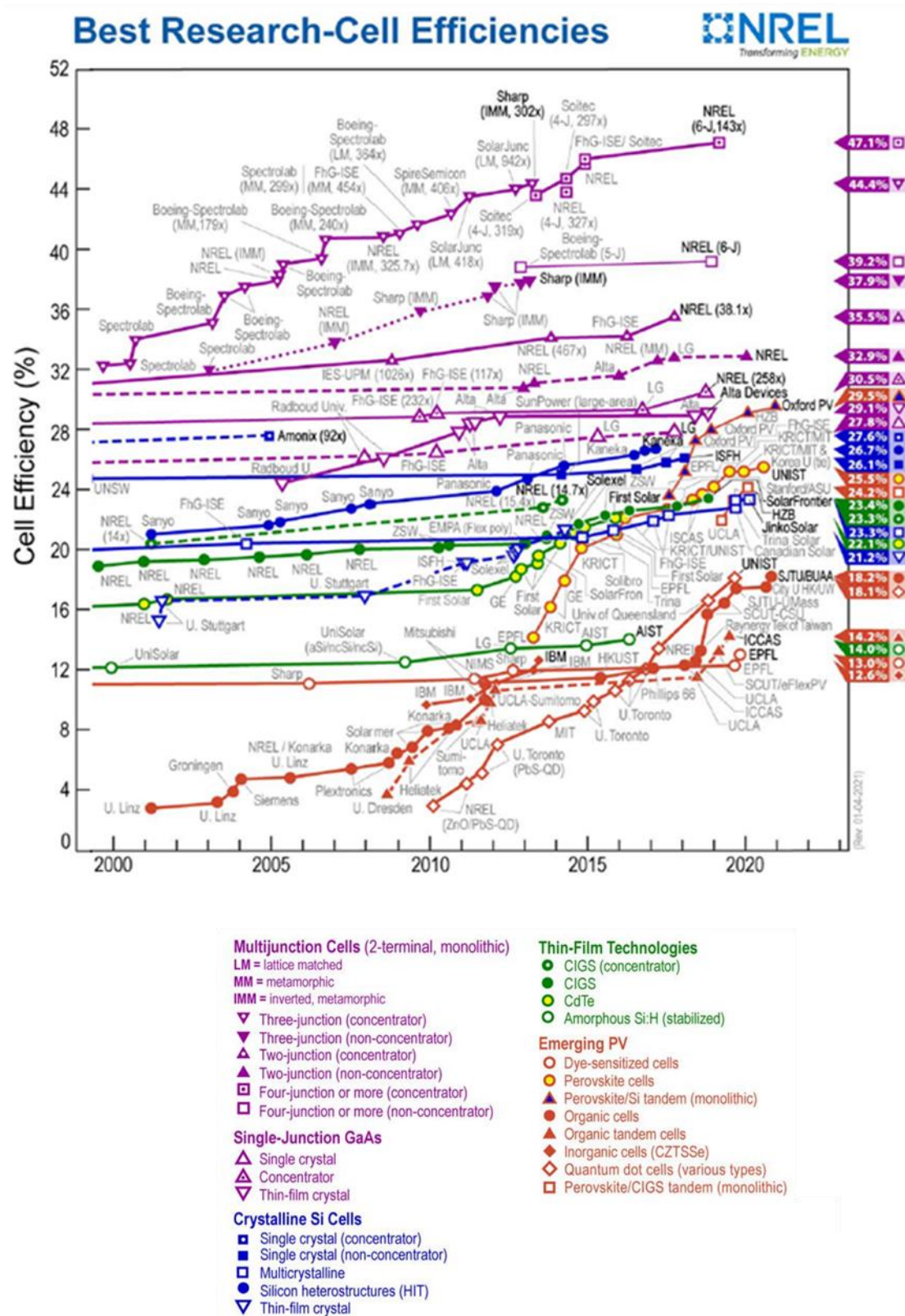


Figure 1.3: Evolution of solar cell efficiency evolution modified to show years 2000-2020 (NREL) [5].

Figure 1.3 illustrates the evolution in record efficiency for all current solar cell types. However, this figure does not include the record efficiencies for modules [5]. The growth is partly driven by technological advancements. The assessment of new technologies is required not only in terms of the efficiency of the technologies, but also in terms of their stability, since many of the emerging technologies, such as organic photovoltaics (OPV) and Perovskite solar cells (PSC) display poor lifetimes and rapid degradations due to different environmental conditions. OPVs and PSCs, however, have shown a faster rate of efficiency improvement in comparison to other technologies such as CIGS, CdTe and crystalline silicon devices. This provides motivation for addressing their stability issues, due to their potentially high efficiencies.

In this thesis, data analytical techniques and machine learning are employed to analyse datasets containing OPV performance and stability data. Machine learning has provided a means of identifying sources of instability in OPV materials from a broad and comprehensive set of sources. Additionally, machine learning forecasting techniques have been applied to outdoor test data. For the first time, a novel approach whereby the cumulative exposure of weather conditions on outdoor tested modules has been used to predict the energy yield of OPV modules to within 5% accuracy by combining multiple algorithms which account for changes in OPV performance due to degradation. Finally, machine learning has been employed to predict the environmental impact of OPVs. This allows the trade – off and balance between material energy costs, environmental impact and energy production to be analysed. A genetic clustering algorithm is used to analyse the net energy of OPV modules which allows the optimum device architecture for maximised net energy output to be derived and paves a pathway for significantly reducing the costs of experimentation and development time of OPVs.

## 1.2 Outline and Structure of Thesis

The subsequent content of the thesis will be ordered as described below:

**Chapter 2:** Literature Review provides an overview of the scientific background and the current methods of stability and performance assessment for OPVs and PSCs as well as the application of machine learning methods.

**Chapter 3:** Experimental and Computational methods employed in order to assess the performance and stability of OPVs and PSCs, both *via* indoor and outdoor testing, as well as *via* computational machine learning and data analytical methods.

**Chapter 4:** First of the three results chapters which reports on the application of Machine Learning (ML) and data analytics to assess the stability of OPVs as well as identifying the causes of degradation as well as identification of materials for stability enhancement.

**Chapter 5:** Second results chapter which reports on the outdoor testing of OPVs, employing similar I-V characterisation techniques as described in chapter 4 to assess and determine the causes of degradation of OPVs in outdoor conditions. In addition, data analytical and ML techniques are applied in order to forecast the lifetimes, and degradation rates of OPVs in outdoor conditions as well as identify the root cause of degradation in outdoor conditions.

**Chapter 6:** Final results chapter which reports on the use of machine learning to predict the embodied energies of OPV devices. A new parameter, 'net energy output' is calculated, which is formed by calculating the energy generated under AM1.5G and subtracting the embodied energy of the solar cell. Machine learning methods are used to find material combinations for maximising the net energy of OPVs and assessing the trade-off between PCE, stability and environmental impact.

**Chapter 7:** Concluding chapter, summarising the main results from the thesis and discusses possible directions for further work.

## Chapter 2. Literature Review and Scientific Background

### 2.1 Solar Radiation Spectrum and Current – Voltage Characteristics

#### 2.1.1 Solar Radiation Spectrum

The operation of solar cells is intrinsically dependent on the solar radiation incident on them and, therefore, understanding how the Earth's atmosphere affects this radiation is important. The Sun's spectrum can be approximated by a blackbody spectrum with a source temperature of 5760 K; when the radiation reaches the upper atmosphere of the Earth, the spectrum is relatively unaffected with a Solar constant of  $\sim 1.37 \text{ kWm}^{-2}$  [6]. However, due to the transmission of the Solar radiation through the Earth's atmosphere, the light is attenuated by scattering and absorption processes: Rayleigh scattering, Mie scattering, non – selective scattering, and absorption. Rayleigh scattering occurs when the solar radiation interacts with particles which are small compared to the wavelength of the radiation. Mie scattering occurs when the solar radiation interacts with particles which have approximately the same diameter as the wavelength of the radiation. Non – selective scattering occurs when the solar radiation interacts with particles which are significantly larger than the wavelength of the radiation and results in all wavelengths being scattered equally [7]. The distance through which the radiation must travel is dependent on the angle of incidence which the solar radiation enters the atmosphere; the path length through which the radiation travels is referred to as the air mass (AM), which subsequently determines the attenuation of the radiation. In the higher regions of the atmosphere the solar spectrum is defined as being AM0 and as AM1.5G at the surface, as defined by the American Society for Testing and Materials (ASTM). Figure 2.1 illustrates the AM0, AM1.5G and AM1.5D spectra, where G and D refer to global and direct respectively. Three factors contribute towards the total irradiation on a surface: the direct normal irradiation (DNI), defined as the amount of solar radiation received per unit area by a surface held perpendicularly to the incident radiation [8]; the diffuse radiation,

defined as the solar radiation received after all scattering effects [9]; and the ground reflected radiation, also known as the albedo.

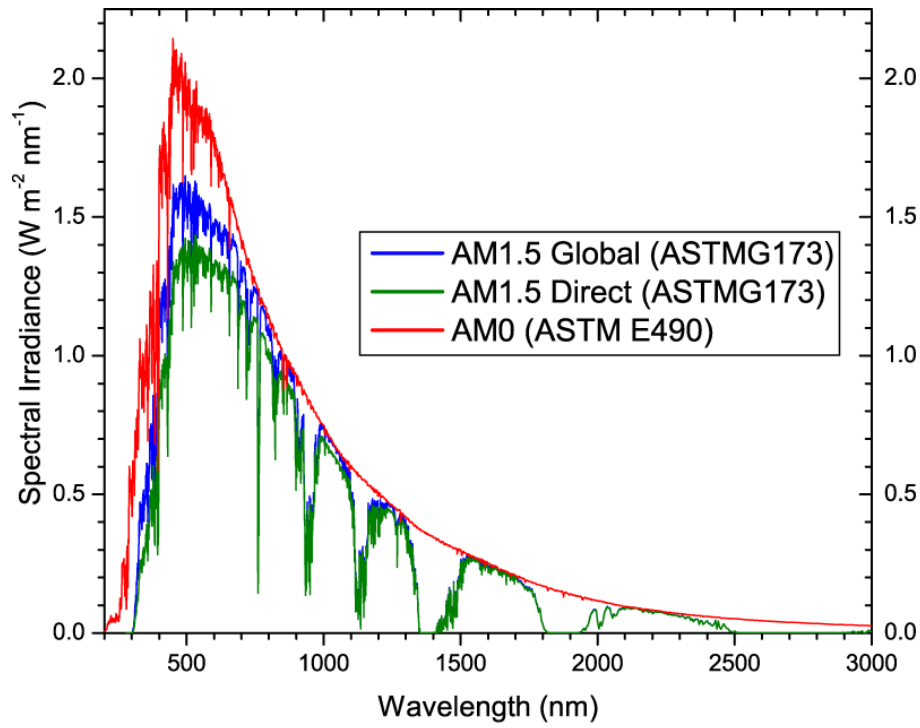


Figure 2.1: AM0, AM1.5 Global and AM1.5 Direct solar spectra [10].

In order to experimentally test solar cells in laboratory conditions, a consensus was required for the irradiance level to be used. The consensus spectrum (AM1.5 G173-03) was agreed between the ASTM and US government laboratories and defined as being  $963 \text{ Wm}^{-2}$ , which is re-standardised to  $1000 \text{ Wm}^{-2}$  [11], [12].

### 2.1.2 Silicon Based Photovoltaics

Silicon based photovoltaic cells are the most common type of solar cells currently available on the commercial market. The first demonstration of these devices, developed in Bell Laboratory, showed efficiencies of 4.5% in 1953 and subsequently 6% in 1954 [13]. Crystalline silicon (c-Si) solar cells currently dominate the PV market and their efficiencies are approximately 27% (with a theoretical limit of  $\sim 29\%$ ) [14]. The c-Si PV are manufactured using two types of semiconductors, namely *n*-type and *p*-type silicon.

The *n*-type and *p*-type silicon is produced by doping with impurity atoms such as boron or gallium [15]. The doping of silicon using boron results in *p*-type silicon since boron has one less electron than required to form bonds with silicon. Therefore, a hole is created in the doped silicon. The *n*-type silicon is produced by doping using atoms which have one more electron in the outer energy level than silicon. This atom can be, for example, phosphorus which has five electrons in its outer shell such that one electron is not involved in bonding and is free to be transported through the silicon structure, producing an *n*-type semiconductor. When the *n* and *p*-type doped silicon are brought into contact with each other a *pn* junction will be formed where there is an excess of electrons on the *n* side and an excess of holes on the *p* side. At the junction, the electrons in the *n*-type semiconductor will diffuse into the *p*-type semiconductor and fill the holes. Consequently, a depletion region is formed in the vicinity of the junction. When all the holes are filled by electrons, the *p*-type side of the depletion region will contain negatively charged ions and the *n*-type side will contain positively charged ions. This produces an electric field that prevents electrons in the *n*-type side from diffusing towards the holes in the *p*-type side. When incident light illuminates the silicon, electrons are ejected leading to the formation of holes. When this occurs in the depletion region, where the electric field exists, the electrons will be transported towards the *n*-type side and holes will be transported towards the *p*-type side. If the *n*-type side and *p*-type side are electrically connected to a load, the electrons will travel through from the *n*-type layer to the *p*-type side, crossing the depletion region and will be subsequently conducted through the external circuit [16].

### 2.1.3 Current – Voltage characteristics of Organic Photovoltaic Solar Cells

OPV solar cells convert electromagnetic radiation into electrical energy by absorption of photons in the active layer, creating electron – hole pairs known as excitons. The active layer is sandwiched between electrodes of different work functions, which transport the electrons and holes out of the active material. The active material is usually a combination of a *p*-type (acceptor) polymer and an *n*-type (donor) fullerene. One of the electrodes is typically transparent to allow illumination of the active material. Interfacial layers are also applied in between the electrodes and the active material in order to ensure charge – selective transport of the electrons and holes to the respective electrodes. Therefore,

the overall efficiency of the OPV depends on several fundamental processes, dictated by all the components in the OPV device and is given by,

$$\eta = \eta_{abs} \times \eta_{diss} \times \eta_{trans} \times \eta_{col}, \quad 2.1$$

where  $\eta_{abs}$  is the photon absorption efficiency,  $\eta_{diss}$  is the exciton dissociation efficiency,  $\eta_{trans}$  is the charge transport efficiency and  $\eta_{col}$  is the charge collection efficiency.

The electrical characterisation of an OPV solar cell is performed by measuring the current – voltage (IV) curve, as shown in Figure 2.2. From the characteristic curve, the following parameters can be derived: the short – circuit current ( $I_{sc}$ ), the open – circuit voltage ( $V_{oc}$ ), the maximum power point (MPP) and corresponding current and voltage ( $I_{MPP}$  and  $V_{MPP}$ ). These are indicated in Figure 2.2. The Fill Factor (FF) is defined as,

$$FF = \frac{I_{MPP} \times V_{MPP}}{I_{sc} \times V_{oc}}. \quad 2.2$$

Subsequently, the power conversion efficiency of the device or module can be calculated using,

$$\eta = \frac{I_{sc} \times V_{oc} \times FF}{P}, \quad 2.3$$

where  $P$  is the incident power. In addition, the irradiance can be defined as,

$$I_{Irr} = \frac{P}{A}, \quad 2.4$$

where  $A$  is the illuminated area and  $P$  is the incident light power. Therefore, defining the short – circuit current – density as  $J_{sc} = I_{sc}/A$ , it follows that,

$$\eta = \frac{J_{sc} \times V_{oc} \times FF}{I_{Irr}}. \quad 2.5$$



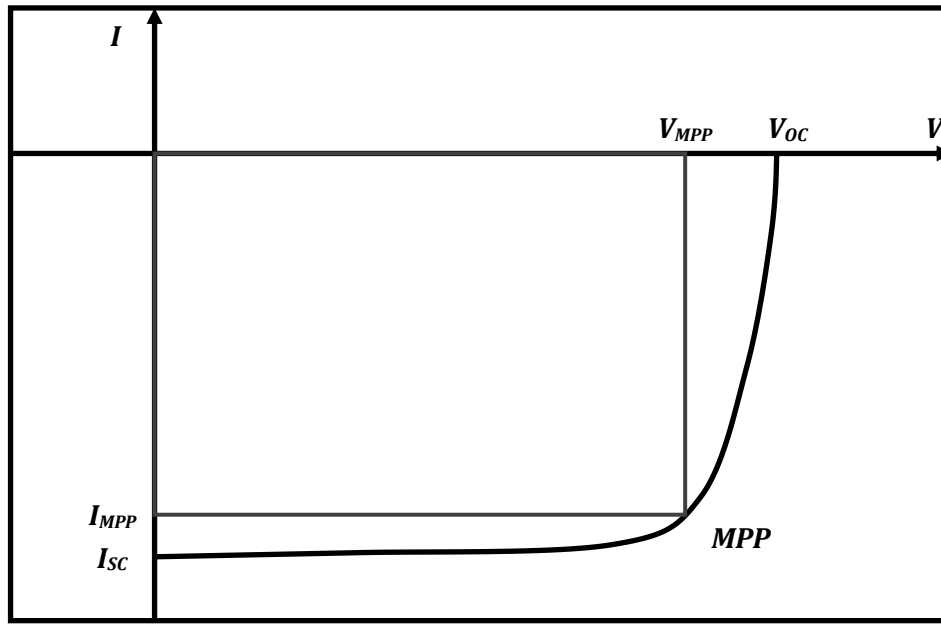


Figure 2.2: Current – voltage characteristics of solar cell.

Inspection of the IV curve allows the series resistance ( $R_s$ ) and shunt resistance ( $R_{sh}$ ) to be determined [17]. The gradient near  $V_{oc}$  is proportional to  $-1/R_s$  and the gradient near  $I_{sc}$  is proportional to  $-1/R_{sh}$ . In order to maximise the photocurrent ( $I_L$ ) extracted from the solar cell,  $R_{sh}$  must be large whilst  $R_s$  must be small. Poor active layer morphology and electrical short circuits through the cell lead to low  $R_{sh}$ . An imperfect interface between the active layer and the electrodes, resulting in a low  $R_{sh}$ , can lead to bimolecular charge recombination losses.  $R_s$  is dependent on the resistances of each of the constituent device layers, with high  $R_s$  values leading to a decrease in charge transportation due to low mobility in the semiconductor active material, interfacial mismatching between the constituent layers and low conductivity at the metallic contacts.

An ideal solar cell can be modelled by an equivalent circuit consisting of a current source,  $S$ , connected in parallel to a diode,  $D$ . In the absence of any illumination on the solar cell, the I-V characteristics will be the same as that of a diode. However, when the solar cell is illuminated, the current source produces a photocurrent,  $I_{ph}$ , causing the diode I-V characteristics to shift along the current axis;  $I_{ph}$  will be directly proportional to the incident irradiance on the device. However, in practice, a solar cell possesses the

parasitic resistances  $R_S$  and  $R_{SH}$ . The equivalent circuit for a solar cell with parasitic resistances modelled, is shown in Figure 2.3.

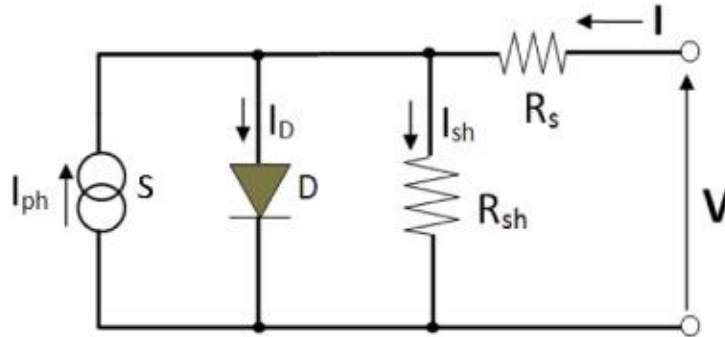


Figure 2.3: Equivalent circuit for a solar cell including parasitic resistances,  $R_S$  and  $R_{SH}$ .

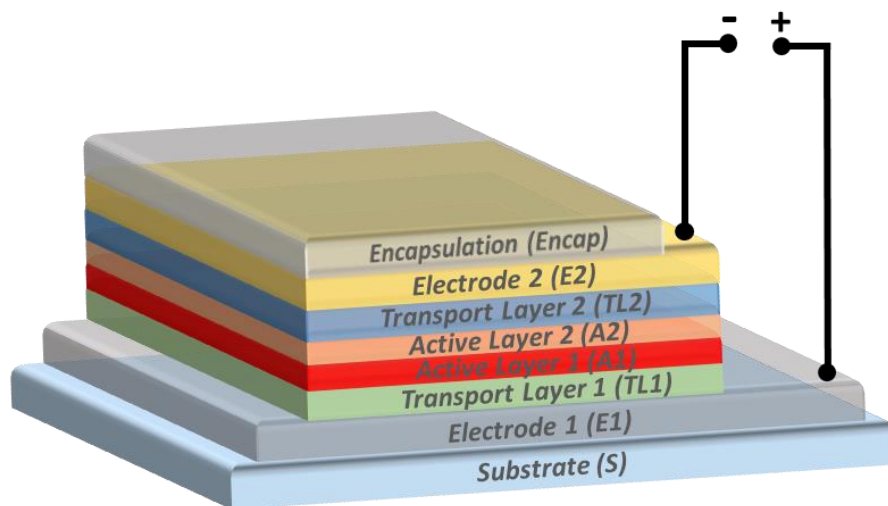
## 2.2 Introduction to OPV and PSC development

Solar photovoltaic (PV) development has been ongoing since the inception of crystalline silicon wafer technology in 1955 when the first silicon module was produced for outdoor use at Bell Laboratories [18]. This technology can be either monocrystalline, polycrystalline, amorphous, nanocrystalline or microcrystalline. By 2019 these technologies accounted for 95% of the total worldwide PV energy production [19]. However, silicon-based PV can be material intensive to manufacture with 1000 times more light absorbing material required, as compared to alternative technologies, and, in addition, ultra-high purity silicon is required for manufacture [20]. This instigated the search for alternative photovoltaic technologies which could be manufactured at a lower cost. Organic photovoltaic (OPV) devices have shown promise with their emergence in 1986 when Tang *et al.* developed the first two component device [21]. However, their efficiencies struggled to compete with the state-of-art performance seen in silicon and multi-junction cells. To overcome the limitations in the efficiency, bulk heterojunction devices were proposed in 1995 [22].

From 2010, two additional types of solar cell have emerged: quantum dot cells and Perovskite solar cells. Quantum dot cells have demonstrated encouraging improvement in efficiency, but Perovskite solar cells (PSC) have yielded a much better efficiency and a rate of growth in development unseen in any other forms of solar cell technology. The first PSC was reported by Snaith et al. in 2012 [23]. Already, the efficiencies of PSCs have reached a level greater than other OPV technologies in a very short space of 5 years.

## 2.3 OPV Operation

Whilst section 2.1 discussed the general operation of a solar cell, a more detailed discussion of OPV now follows as this technology is the focus of this thesis. The generic architecture for an OPV is illustrated in Figure 2.4. The structure normally consists of five layers, or thin films, namely: the back electrode (electrode 1), the lower transport layer (TL1), the active layer, the upper transport layer (TL2) and a top electrode (electrode 2). The OPV solar cell will convert photon energy into electrical energy within the active layer, which typically consists of a heterojunction of n – type (donor) and p – type (acceptor) materials. Electrode 1 is typically transparent such that light can reach the active layer. The OPV architectures will be discussed in greater detail in 2.3.3.



*Figure 2.4: The generic structure of an organic photovoltaic (OPV).*

### 2.3.1 Charge generation and dissociation in OPVs.

In order to understand the conduction properties of organic electronics, the process of charge generation and charge transfer must first be understood for organic semiconductor materials. Charge generation occurs within photoactive materials *via* absorption of light. When light is absorbed by a photoactive material, an electron – hole pair is created, also known as an exciton. The exciton is bound by the coulomb force interaction and, therefore, the electron and hole are not free charges. The coulomb force arises from the opposite charges of the electron and hole [24]. Organic semiconductor materials can be classified as being either polymeric or small molecules. The principal feature of any organic semiconductor, providing its charge transfer characteristics, is the conjugated backbone which consists of alternating single and double carbon – carbon bonds [25].

The single bond in a conjugated backbone is formed from a  $\sigma$  bond which is a localised  $sp^2$  state which results from the hybridisation of the 2s orbital and two of the carbon's 2p orbitals. This leads to a strong  $\sigma$  bond in plane with the conjugated backbone and orientated at  $120^\circ$  to each other. The double bonds are formed from the combination of a single  $\sigma$  bond and a single  $\pi$  bond. The  $\pi$  bond is formed from the remaining, unhybridized electron in the third p orbital, orientated perpendicularly to the  $sp^2$  plane. Neighbouring orbitals in these orbitals overlap sideways, thus allowing for delocalisation along the conjugated backbone. The excited states of the  $\sigma$  and  $\pi$  bonds are represented by  $\sigma^*$  and  $\pi^*$  respectively. Antibonding orbitals are where electron density surrounds the nuclei as opposed to bonding orbitals where the electron density is between the bonding nuclei. The  $\sigma - \sigma^*$  energy gap is significantly larger than the  $\pi - \pi^*$  energy gap. Subsequently, the  $\pi - \pi^*$  energy gap allows for the charge transport along the conjugated backbone. The  $\pi$  orbital represents the highest occupied molecular orbital (HOMO) whilst the empty states of the  $\pi^*$  orbital represents the lowest unoccupied molecular orbital (LUMO) [26]. The energy difference between the HOMO and LUMO represent the band gap of the semiconductor. If an electron in the HOMO level possesses sufficient energy from a thermal source or by the absorption of an incident photon, the electron in the HOMO will be promoted to the LUMO level, leaving a hole in the HOMO level.

Greater insight into the charge transport mechanism in solar cells can be achieved by considering the energy level diagram, consisting of the energy bands for the active material and electrodes, shown in Figure 2.5, for OPVs. The diagram shows exciton generation in the donor active material and how it is separated [27]. For an OPV, the exciton formation occurs within the donor material and the exciton can diffuse towards the donor – acceptor interface. Due to the potential difference at the interface, the exciton can be dissociated into free charges *via* the charge-transfer state, where the electron in the LUMO of the donor is transferred to the LUMO of the acceptor. The dissociated free electrons and holes are subsequently transported to the cathode and anode respectively due to the electric field created by the work function mismatch between the two electrodes. This dissociation can only occur when the energy difference between the donor electron and the LUMO acceptor energy level is greater than the exciton binding energy.

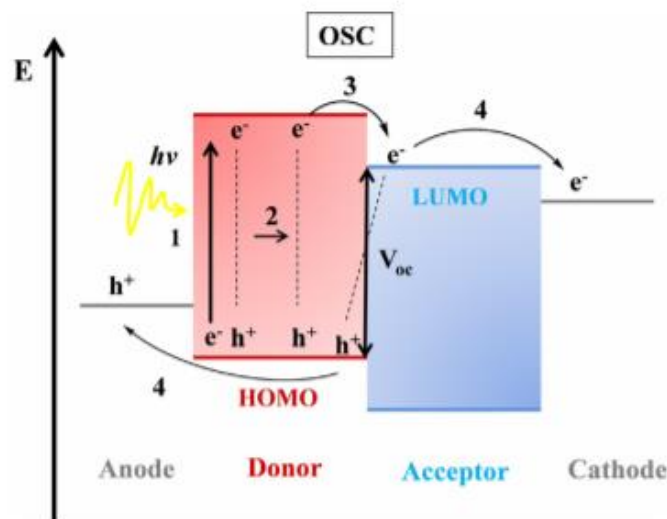


Figure 2.5: Energy band diagram showing exciton formation and dissociation in an OPV, followed by charge transport to the electrodes [27].

As the distance between the electron and hole increases, the attractive force between them decreases [28]. However, even after the dissociation of the exciton has occurred, and the distance between the electron and hole increases, there is still attraction *via* the charge transfer state [29]. This means that recombination of the exciton

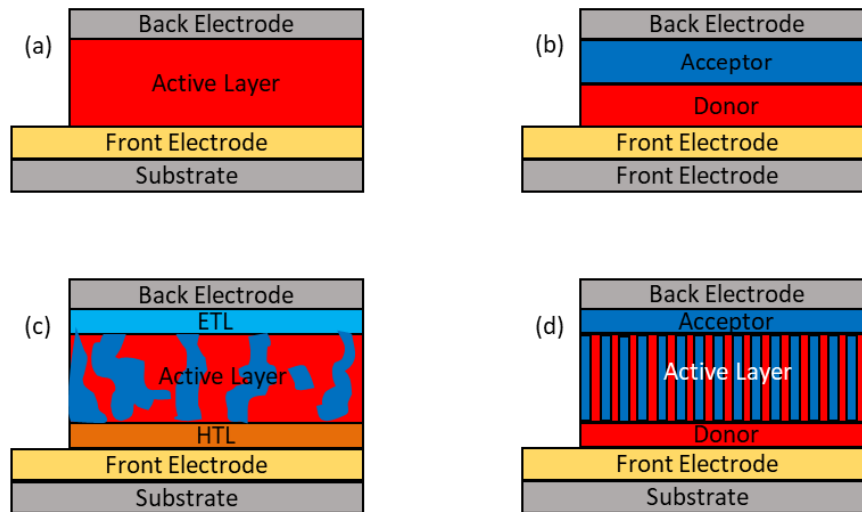
remains possible across the interface between the two materials [30]. Several recombination mechanisms exist [31]: geminate recombination by exciton decay, recombination through the charge transfer state, or non-geminate recombination of free carriers [32], [33]. Geminate and non – geminate recombination can be radiative, or non – radiative in nature [34].

### 2.3.2 OPV Device Structures.

Early OPV device structures employed a single active layer with a Schottky electrode, also known as a Schottky diode, as shown in Figure 2.6 (a). Exciton dissociation in this regime utilises the difference in energy levels between the active material and the electrodes and due to the short diffusion lengths displayed by excitons, this approach had limited efficiency ( $< 0.1\%$ ) [35]. Later a bilayer structured active layer was utilised, where the active layer was split into two layers, one being the donor and the second being the acceptor. This structure was first introduced by Tang in 1986 [36]. The interface of the bilayer between the donor and acceptor materials, produces a more efficient method of charge dissociation, so long as the excitons are generated at a location less than one diffusion length from the interface. The diffusion length is defined as the distance that the exciton will travel before recombination occurs. The requirement for two semiconductor materials is necessary due to the low dielectric constant of organic semiconductor materials, which leads to an increased exciton binding energy,  $E_b$ , given by  $E_b = R_0\mu/m_0\epsilon_r^2$  where  $R_0$  is the atomic Rydberg,  $\mu$  is the reduced effective mass,  $m_0$  is the free electron mass and  $\epsilon_r$  is the relative dielectric constant of the material, defined by  $\epsilon_r = \epsilon/\epsilon_0$ , where  $\epsilon$  is the permittivity of the substance and  $\epsilon_0$  is the permittivity of free space. In inorganic semiconductors  $E_b$  can be overcome by thermal energy at room temperature [37].

The bulk heterojunction active layer was introduced in order to overcome the short diffusion length issue observed in OPVs, which required small device thicknesses [38]. A bulk heterojunction active layer is formed by mixing the donor and acceptor materials prior to deposition. This results in a dispersed structure with an increased donor – acceptor interface [35]. In this regime, the formation of an exciton always occurs near to the donor – acceptor interface, and dissociation occurs more readily. A variant of the bulk heterojunction structure is the ordered heterojunction. This structure is formed

by nanoimprint lithography of the donor material, followed by deposition of the acceptor material [39], [40]. A schematic of the different active layer structures is shown in Figure 2.6.



*Figure 2.6: OPV device structures for (a) Schottky diode, (b) Bilayer device, (c) Bulk heterojunction and (d) Ordered heterojunction.*

It should also be noted that interlayers and transport layers are used for OPV devices and play an important role in improving the charge extraction [27]. This is particularly the case in bulk heterojunction active layer structures since both the acceptor and donor materials extend to both electrodes and, therefore, selective extraction of either the electrons or holes is required. In addition, interlayers in OPV devices help to ensure an Ohmic contact and facilitate adequate internal electric fields, good film morphology and carrier recombination rates [41]. The hole is transported out through the hole transport layer (HTL) and the electron through the electron transport layer (ETL). At the interface between the active material and each of the transport layers a charge separated state will form [42]. The electrons move from the HOMO to LUMO whilst the holes move from the HOMO to the LUMO. Finally, the electrons pass through the ETL to the anode and the holes will pass through the HTL to the cathode.

### 2.3.3 OPV architectures

The architecture of an OPV can be identified as two main types: “conventional” and “inverted” configurations. For a conventional configuration, Electrode 1 is the anode and Electrode 2 is the cathode. Therefore, in the conventional configuration, TL1 will be an HTL whilst TL2 will be an ETL. The normal configuration was the first to be introduced and typically used PEDOT:PSS as the HTL and Calcium (Ca) as the ETL and aluminium (Al) as Electrode 2 [43].

However, these conventional configuration devices were found to be highly unstable, primarily due to the acidic nature of PEDOT:PSS [44] and the ease of oxidation and reactivity with water of Ca [45], [46] and Al [47], [48]. However, more recent examples use nickel oxide [49], copper thiocyanate [50], molybdenum oxide and graphene oxide [51]. The inverted architecture was introduced in order to improve the stability of OPVs. Inverted configuration OPVs typically use Zinc oxide (ZnO) as the ETL for TL1 and molybdenum oxide ( $\text{MoO}_3$ ) as the HTL for TL2 [52]–[54]. More recently, PEDOT:PSS [55], nickel oxide ( $\text{NiO}_x$ ) [56] and vanadium pentoxide ( $\text{V}_2\text{O}_5$ ) [57] have been used as the HTL whilst titanium oxide ( $\text{TiO}_x$ ) [58] and tin oxide ( $\text{SnO}_2$ ) [59] have been used as the ETL. Silver (Ag) is typically used as Electrode 2 [60]–[62]. Figure 2.7 shows the normal and inverted architectures for typical OPV solar cells, employing a bulk heterojunction active layer.

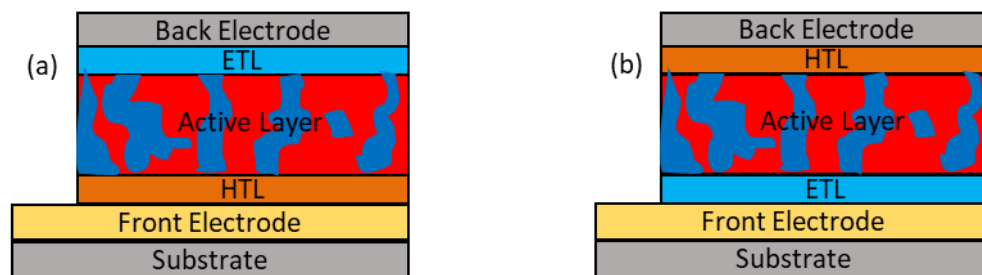


Figure 2.7: (a) Conventional and (b) inverted device architectures for OPV devices.



### 2.3.4 High Performance and Stability OPV devices

Extending the lifetime of OPVs is vital in order to realise their feasibility for commercialisation. Over the course of the past 20 years, reports have shown that changes to OPV material and design have increased both light and dark stability [63]. Methods for enhancing the stability of OPVs are varied and diverse with approaches ranging from different encapsulation techniques – for example [64]–[69], optimisation of the transport layers – for example [54], [70]–[75] and electrode variation – for example [76]–[80] to name but a few. OPV lifetimes are still far off the required stability in order to realise commercialisation, but there are a number of examples where high stabilities have been reported. Polymer solar cells displaying long operating lifetimes have been reported for devices constructed with two different photoactive materials: P3HT:PCBM and PCDTBT:PCBM [81]. Therein, they show how such devices can maintain approximately 80% of their initial efficiency after 4000 hours of continuous testing. The initial average efficiencies of the P3HT:PCBM and PCDTBT:PCBM devices were  $5.5 \pm 0.15\%$  and  $4 \pm 0.05\%$  respectively. These values represent the average efficiencies for eight samples of each device type. Non-fullerene acceptors, IDTBR and IDFBR were combined with both a scalable and affordable donor polymer, PBDTTT-EFT (PCE10) and devices were found to be highly efficient owing to changes in the microstructure which reduces charge recombination and increases photovoltage [82]. Burn – in free non-fullerene – based organic solar cells have been developed [83]. Devices possessing a P3HT:IDTBR active layer were tested in comparison to devices possessing a P3HT:PCBM active layer and 90% of the initial efficiency was maintained after 2000 hours of testing as opposed to approximately 65% respectively [83].

Electron filtering compound buffer layers (EF-CBLs) were employed by Burlingame *et al.* in order to improve the lifetime of their devices. These included 10 nm 1:1 BPhen:C<sub>60</sub>/5 nm BPhen (BP60), 10 nm 1:1 3TPYMB:C<sub>60</sub>/5 nm 3TPYMB (3T60), 10 nm 1:1 TPBi:C<sub>60</sub>/3 nm TPBi (TP60), 10 nm 1:1 TPBi:C<sub>70</sub>/3 nm TPBi (TP70) and 10 nm 1:1 BP4mPy:C<sub>60</sub>/3 nm BP4mPy (BPPy60). By incorporating EF-CBLs into planar – mixed heterojunction OPVs, improved morphological stability was found as opposed to devices employing neat Bathophenanthroline (BPhen) cathodic buffer layers [84].

OPV efficiency records are also improving yearly with reports of flexible module OPVs displaying efficiencies of 13.4% [85]. OPVs have also been fabricated which display efficiencies of over 16% when chlorinated acceptors are used, due to increased open – circuit voltages [86]. The current world record for OPV efficiency stands at 18.22% and was achieved by Liu *et al.* in 2020 [87]. This result was achieved using a ITO/PETDOT:PSS/D18:Y6/PDIN/Ag architecture, where D18 is a copolymer donor and Y6 is non-fullerene acceptor. D18 synthesis uses a fused – ring acceptor unit, dithieno [3',2':3,4;2'',3'':5,6] benzo [1,2-c] [1,2,5] thiadiazole (DTBT). DTBT has a large molecular ring compared with DTTP and affords D18 with higher hole mobility. Y6, (2,20-((2Z,20Z)-((12,13-bis(2-ethylhexyl)-3,9-diundecyl-12,13-dihydro- [1,2,5] thiadiazolo [3,4-e]thieno [2,30':4',50] thieno [20,30:4,5] pyrrolo [3,2-g] thieno [20,30:4,5] thieno [3,2-b] indole-2,10-diyl) bis (methanylylidene)) bis (5,6-difluoro-3-oxo-2,3-dihydro-1H-indene-2,1-diylidene))dimalononitrile), is based on the dithienothiophen [3.2-b]-pyrrolobenzothiadiazole (TPBT) central unit. The blend of Y6 and a wide bandgap donor-acceptor copolymer donor can harvest most of the visible and NIR light and give high PCEs [87].

## 2.4 Stability Tests

### 2.4.1 ISOS stability test protocols

As OPVs lacked the stability of mature technologies such as silicon PV, and research was driven by academic groups, new consensus standards were developed to support research and development. The International Summit on Organic PV Stability (ISOS) test protocols were developed to provide a standardised method for testing solar cell performance and stability via an internationally agreed methodology, derived from a number of international round robin tests which collated the results from several groups. Through adoption of these, reduced variability of reported testing methods in the literature could be achieved with improved benchmarking. The ISOS test protocols give guidance on how to perform shelf-life, outdoor, laboratory weathering and thermal cycling testing. The criteria for each test can be found in Table 2.1.

Three levels						
Basic (Level 1)			“Hand held” measurements using the simplest equipment and few conditions Fixed conditions and protocols suited for most labs Standardized tests applied in certified labs. Extended range of parameters to monitor, etc.			
Intermediate (Level 2)						
Advanced (Level 3)						
Test type	Dark			Outdoor		
Test ID	ISOS-D-1 Shelf	ISOS-D-2 High temp. storage	ISOS-D-3 Damp heat	ISOS-O-1 Outdoor	ISOS-O-2 Outdoor	ISOS-O-3 Outdoor
Light source	None	None	None	Sunlight	Sunlight	Sunlight
Temp. <sup>a</sup>	Ambient	65/85 °C	65/85 °C	Ambient	Ambient	Ambient
Relative humidity (R.H.) <sup>a</sup>	Ambient	Ambient (low)	85%	Ambient	Ambient	Ambient
Environment <sup>a</sup>	Ambient	Oven	Env. chamber	Outdoor	Outdoor	Outdoor
Characterization light source	Solar simulator or sunlight	Solar simulator	Solar simulator	Solar simulator	Sunlight	Sunlight and solar simulator
Load <sup>b</sup>	Open circuit	Open circuit	Open circuit	MPP or open circuit	MPP or open circuit	MPP
Test type	Laboratory weathering testing			Thermal cycling		
Test ID	ISOS-L-1 Laboratory weathering	ISOS-L-2 Laboratory weathering	ISOS-L-3 Laboratory weathering	ISOS-T-1 Thermal cycling	ISOS-T-2 Thermal cycling	ISOS- T-3 Thermal cycling
Light source	Simulator	Simulator	Simulator	None	None	None
Temp. <sup>a</sup>	Ambient	65/85 °C	65/85 °C	Between room temp. and 65/85 °C	Between room temp. and 65/85 °C	−40 to +85 °C
Relative humidity (R.H.) <sup>a</sup>	Ambient	Ambient	Near 50%	Ambient	Ambient	Near 55%
Environment/setup	Light only	Light & Temp.	Light, Temp. and R.H.	Hot plate/oven	Oven/env. chamb.	Env. chamb.
Characterization light source	Solar simulator	Solar simulator	Solar simulator	Solar simulator or sunlight	Solar simulator	Solar simulator
Load <sup>b</sup>	MPP or open circuit	MPP or open circuit	MPP	Open circuit	Open circuit	Open circuit
Test type	Solar-thermal-humidity Cycling					
Test ID	ISOS-LT-1 solar-thermal cycling		ISOS-LT-2 solar-thermal-humidity cycling		ISOS-LT-3 solar-thermal-humidity-freeze cycling	
Light source	Simulator		Simulator		Simulator	
Temp.	Linear or step ramping between room temp. and 65 °C		Linear ramping between 5 and 65 °C		Linear ramping between −25 and 65 °C	
Relative humidity (R.H.)	Monitored, uncontrolled		Monitored, controlled at 50% beyond 40 °C		Monitored, controlled at 50% beyond 40 °C	
Environment/setup	Weathering chamber		Env. chamb. with sun simulation		Env. chamb. with sun simulation and freezing	
Characterization light source	Solar simulator		Solar simulator		Solar simulator	
Load <sup>b</sup>	MPP or open circuit		MPP or open circuit		MPP or open circuit	

<sup>a</sup> The ambient conditions are defined as 23 °C/50%RH in general, and 27 °C/65%RH accepted in tropical countries according to ISO 291(2008): Plastics—Standard atmospheres for conditioning and testing.

<sup>b</sup> Open circuit refers to a simply disconnected device or device connected to a sourcemeter set to 0 current.

*Table 2.1: Overview of different types of ISOS test protocols [88].*

ISOS test protocols were primarily developed in order to test organic solar cells but can be applied to a wide range of other device types [88]. In 2020 these test protocols were updated for testing of perovskite solar cells (PSCs) such that the properties, specific to PSCs could be tested, for example, ionic redistribution in response to an applied electric field, reversible degradation, impact of light cycling and to distinguish between ambient – induced degradation from other stress factors [89].

## 2.4.2 Stability issues in OPVs

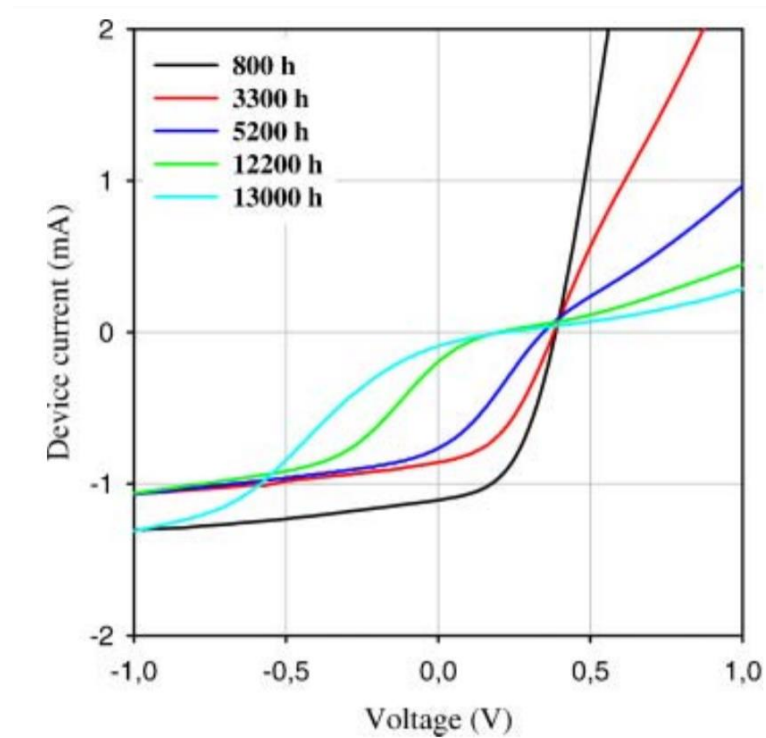
### *Extrinsic and Intrinsic Degradation*

The degradation of OPV solar cells is related to several stress factors and can be separated into both extrinsic and intrinsic factors. *Extrinsic factors* relate to the degradation which occurs when OPVs are aged in ambient conditions [90]. *Intrinsic factors* relate to the inherent degradation of the constituent materials of an OPV, unrelated to the extrinsic stress factors applied. Intrinsic degradation can be induced by both light and heat, either in dark conditions or under illumination. Dark storage degradation typically involves the movement of materials in the films whilst light induced degradation is caused by photochemical reactions which occur in the absorbing layer [90]. There have been several review articles on OPV stability and it is clear that degradation is not due to a single failure mechanism. This has been demonstrated by consideration of the combined effect of humidity and temperature on OPV degradation, leading to an interaction effect. When an OPV is stressed by both temperature and humidity, a greater degradation is observed than when each factor is increased individually [91]. Furthermore, by not considering the interaction effects, misleading conclusions can be reached due to the significant impact of the interactions on the degradation [92]. The effects of applying multiple stress factors on OPV modules simultaneously using a design of experiment approach was performed to demonstrate predictive aging of OPVs based on multistress testing using a log-linear life model [93]. Table 2.2 gives common intrinsic and extrinsic stress factors considered during OPV stability studies.

The severity of conventional architecture OPV degradation was demonstrated by Heeger *et al.* who fabricated conventional architecture OPVs but found that the devices lost all their PCE within 25 hours of ambient air storage; in comparison, an inverted device maintained more than 70% of its initial efficiency after 15 days of air storage [94]. One of the earliest promising results was that of Garcia – Belmonte *et al.* who found that OPVs stored in a glove box for approximately 1 year degraded from 3% efficiency to 1.5% [95]. The operation of OPVs was investigated under continuous 1 Sun illumination at 72°C for 13,000 hours and only 2% of the initial efficiency was retained [96]. The I-V characteristics at different points of degradation are shown in Figure 2.8.

Intrinsic	Extrinsic
Active layer decomposition	Oxygen infiltration
Hole Transport layer diffusion	Humidity/water infiltration
Electron Transport layer diffusion	Irradiation
Electrode diffusion	Heating
Metastable film morphology	Mechanical Stress

*Table 2.2: List of different Intrinsic and Extrinsic stress factors which influence OPV degradation. The list is not exhaustive but highlights some of the key degradation mechanisms.*



*Figure 2.8: Light I-V curves ( $1000 \text{ Wm}^{-2}$ , AM1.5,  $72 \pm 2^\circ\text{C}$ ) for the device at selected time intervals [96].*

The challenges and strategies involved in the stability enhancement of OPV solar cells can be understood in greater detail by considering some of the many review articles

written on the subject. One such review was presented by Cheng *et al.* [97]. Therein, several different intrinsic and extrinsic stress factors were identified as being influential in the stability of OPVs. The intrinsic factors included the metastable morphology, the diffusion of the electrodes and buffer layers into the active material; the extrinsic factors included oxygen and water infiltration, irradiation, heating and mechanical stress. Other work for increasing the stability includes material design, device engineering of the active layers, employing an inverted architecture, buffer layer optimisation, using stable electrodes and using encapsulation. However, from such a review, ranking of the different intrinsic and extrinsic factors is difficult in terms of severity; machine learning, however, presents a possible methodology for considering such literature sources, and using analytical techniques to quantify and rank the significance of each factor. The light, humidity and thermal stability of OPVs will now be discussed to highlight the effects of these extrinsic factors.

### *Light Stability*

OPV devices have often displayed poor photostability especially when using a P3HT:PCBM active layer. For example, devices using P3HT:PCBM active layers were tested using either PEDOT:PSS or MoO<sub>3</sub> as the HTL [98]. The two types of devices were tested under continuous illumination; the PEDOT:PSS device retained only 58% of its initial efficiency after 3 hours whilst the MoO<sub>3</sub> device retained over 88% of its initial efficiency over 3 hours. Several approaches and new active materials have been tested in recent years. The photostability of ternary blend OPVs was studied in terms of the role of PC<sub>70</sub>BM. Using a ternary blend was shown to improve the efficiency as well as the photostability compared with binary blend OPVs. PC<sub>70</sub>BM was added to PBDB-T:ITIC and PTB7-Th:ITIC blends in various ratios to form the ternary blend devices. It was found that the ternary blend devices outperformed the binary blend devices in terms of efficiency and photostability with only a 10% average loss in efficiency under continuous illumination. This was attributed to changes in the molecular structure of the active layer blends with the ternary blend displaying greater resilience to photoinduced molecular changes [99].

The light stability of OPVs using non-fused electron acceptors was studied using a PBDB-T donor material and PTIC and ID4F acceptor materials. Under continuous

illumination from a metal halide lamp without UV filtration, the device using PTIC retained approximately 70% of its original efficiency over 50 hours whilst the device using ID4F dropped to approximately 25% of its original efficiency over the same period. The photostability of these devices was further investigated by comparing the neat PTIC and ID4F films under continuous 1-Sun illumination. The PTIC film maintained its green colour with a steady absorption at 747nm over 32 hours. The ID4F film became transparent after 16 hours of illumination, indicating a breakdown of the conjugated system [100].

Another interesting study was conducted to provide a comparison of the indoor and outdoor stability of OPVs, where the OPVs are based on a blend of donor – acceptor polyfluorene copolymer and a fullerene derivative (PFDT2BT-8:PC71BM). Outdoor testing was performed for a period of 12,000 hours with  $T_{S80}$  determined to be greater than 10,000 hours. Indoor testing was performed for 650 hours under a constant solar simulator illumination of 1 Sun and the degradation under the two conditions was found to be approximately dependent on the optical energy dose, which is defined as being the optical radiant energy absorbed per unit area. The performance of the devices was monitored under inert conditions by testing the devices in a nitrogen filled chamber [101]. Whilst the difference in indoor and outdoor stability is likely to vary for different material systems, this is one of the few studies where both tests have been conducted.

Non fullerene acceptors have emerged over the past 5 years; the stability of OPV devices based on an EH-IDTBR acceptor layer, under continuous illumination, was studied and it was found that the device retained 50% of its initial efficiency after 100 hours; this was attributed to the increased resistance of the organic semiconducting layer [102]. The increased resistance was caused by a decrease in the number of available charge carriers implying that structural changes had ensued in the EH-IDTBR molecule and this strongly governed the degradation. The changes in the location of the resistances affecting the performance of the P3HT:EH-IDTBR cell were identified using EIS and equivalent circuit analysis [102]. The degradation of the devices are shown in Figure 2.9, which illustrates the I-V characteristics and the time dependence of the PCE,  $J_{sc}$  and FF.

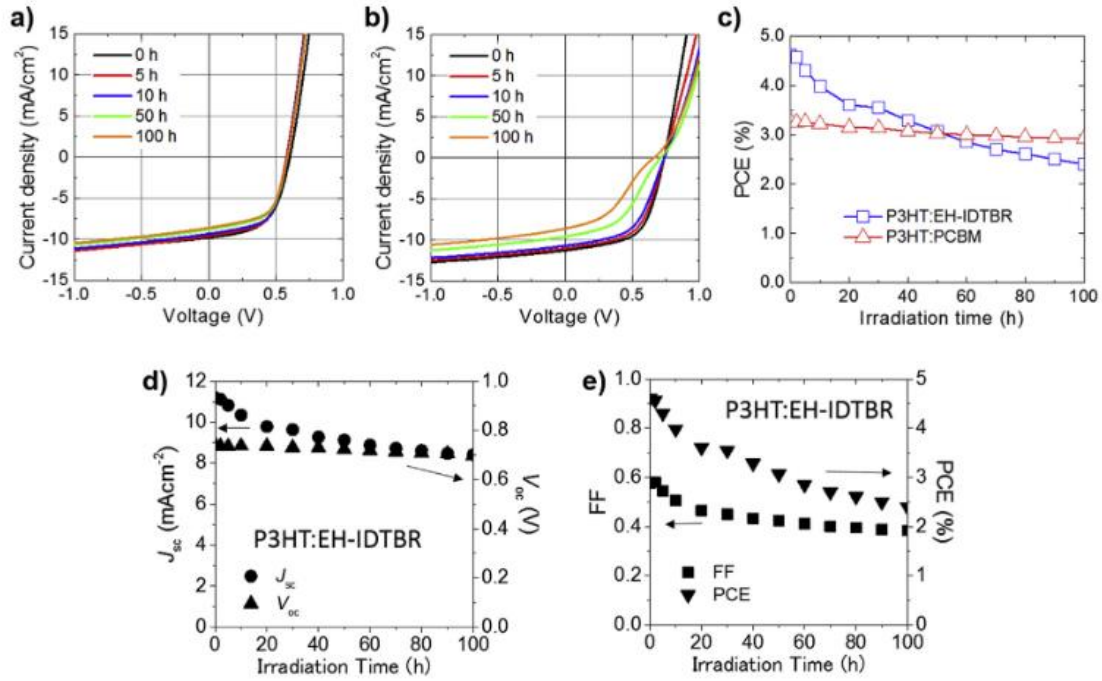


Figure 2.9: J-V characteristics for a) P3HT:PCBM and b) P3HT:EH-IDTBR cells, c) cell durability. d)  $J_{sc}$  and  $V_{oc}$ , and e) changes of FF and PCE for P3HT:EH-IDTBR cells under continuous light irradiation for 100 h [102].

### Humidity Stability

Humidity is one of the most prominent factors governing the stability of OPVs with moisture infiltration resulting in chemical breakdown of the active material [103] and corrosion of the electrode [104]. Electron microscopy has been employed as a means of studying the degradation of OPV solar cells when exposed to damp – heat conditions [105]; several HTL and cathode materials were utilised in the fabrication of the OPVs such that the water ingress could be monitored as well as the effect on the device. It was found that aluminium cathodes and those incorporating calcium both reacted with water, leading to voids and delamination. In the study, PEDOT:PSS was found to increase the rate of degradation by facilitating the ingress of water into the OPV device [105].

Encapsulation has been identified as a method to improve OPV stability against humidity and oxygen related degradation [90]. The long-term rate of oxidation of the electrodes and active material is ultimately limited by the diffusion rate of oxygen and



water, which outweighs any improvements gained by having more ordered and more dense film morphologies. Furthermore, for device lifetimes of several thousands of hours, glass – on – glass encapsulation is required to adequately reduce oxygen and water permeation [106]. A high glass transition temperature,  $T_g$ , in excess of 100°C, relative to the operating conditions is additionally required in order to achieve high dark stability since the temperature at which solar cells degrade is closely related to  $T_g$ . This has been demonstrated by comparing the thermal stability of polymer:fullerene bulk heterojunctions (BHJs) using P-Phenylene Vinylene (PPV) with high and low  $T_g$  values; solar cells made from low  $T_g$  PPV degrade at room temperature, whilst high  $T_g$  PPV solar cells do not due to their more stable morphologies [90], [107].

Several advantages, and disadvantages, for improving OPV stability can be provided by encapsulation [64]. Ethylene vinyl acetate (EVA) is discussed since this is one of the most widely used encapsulation materials in compact silicon solar cells. However, it is not necessarily well suited to OPVs. It has been noted that the adhesion strength of EVA is affected by processing conditions, damp-heat and exposure time. Despite SiO<sub>x</sub> and SiN<sub>x</sub> encapsulation leading to improvements in device performance, the deposition of these materials leads to plasma damage due to the sensitivity of the organic materials to radiation, charging and heating. It is also noted that permeation through the encapsulant depends on transport through pinholes. Overall, there is much ongoing work investigating new materials which can both satisfy the encapsulation requirement whilst also achieving high reliability and lifetimes. Many of the issues are common with organic light emitting diode (OLED) technology. In addition, the processing temperature of the encapsulant must be within a range which the organic and polymer materials can tolerate. One solution identifies the use of multilayer films consisting of alternating inorganic and organic layers such that the defects in the inorganic layer are interrupted and do not channel through the film structure. However, the barrier performance consisted of a steady state permeation rate as well as a transient rate which was found to exist over a specific period (lag time) that may be longer than the lifetime of the encapsulated OPV. The permeation rate in the transient region is usually lower than the steady state permeation. This means that characterising the barrier performance from the from only the initial transient period leads to an underestimate for the total permeation rate for long – term applications. Therefore, characterising the barrier performance should distinguish between the steady state and transient permeation

regions and account for the lag time to avoid overestimation of the barrier performance [64]. A promising method to achieve low cost and high throughput inorganic layers could be by using solution processing which is not as expensive as using vacuum deposition.

### *Thermal Stability*

Finding a means of stabilizing the active layer composition and morphology is key to realising thermally stable active layers. One cause of this instability is due to the intermixing of the donor and acceptor material; for example, reducing the PTB7 polymer and fullerene intermixing has been shown to produce thermally stable BHJ devices [108]. This is achieved by a sequential deposition of the nanostructured PTB7 and fullerene layer, as opposed to a blended deposition approach. X – ray diffraction has been used to reveal that the sequential deposition of the nanostructured PTB7 and fullerene layers leads to less intermixing. A PCE of 7.43% was achieved using this method and maintained this initial PCE after 10 days of thermal annealing at 140 °C. Blended PTB7 and fullerene deposition, which also show similar PCE values, maintained only 78% of its initial PCE under the same conditions.

Pentacene has been shown to enhance the thermal stability of inverted polymer solar cells when incorporated into the P3HT:PCBM active layer by producing more stable donor – acceptor interfaces [109]. The pentacene inclusion in the blends was found to suppress the crystallization of P3HT and PCBM, confirmed by microscopy and absorption spectroscopy. When stored at 120 °C for 24 hours, the initial PCE of the pentacene containing device reduced to 70%, whilst without pentacene, the device performance reduced to 13% of its initial value.

Employing new active layer blends have led to thermally stable, highly efficient and ultraflexible organic photovoltaics; Devices with PCEs of up to 10% have been developed which could endure temperatures of 100 °C with 80% retainment over the course of 500 hours. This was achieved by employing poly[4,8-bis(5-(2-ethylhexyl)thiophen-2-yl)benzo[1,2b;4,5-b']dithiophene-2,6-diyl-alt-(4-octyl-3-fluorothieno[3,4-b]thiophene)-2-carboxylate-2,6-diyl] (PBDTTT-OFT) blended with PC<sub>71</sub>BM. This blend exhibits a high degree of face – on orientation, thus enabling an improved percolating

network and maintaining a sturdy microstructure, even at 120 °C [110]. This is illustrated in Figure 2.10 (a).

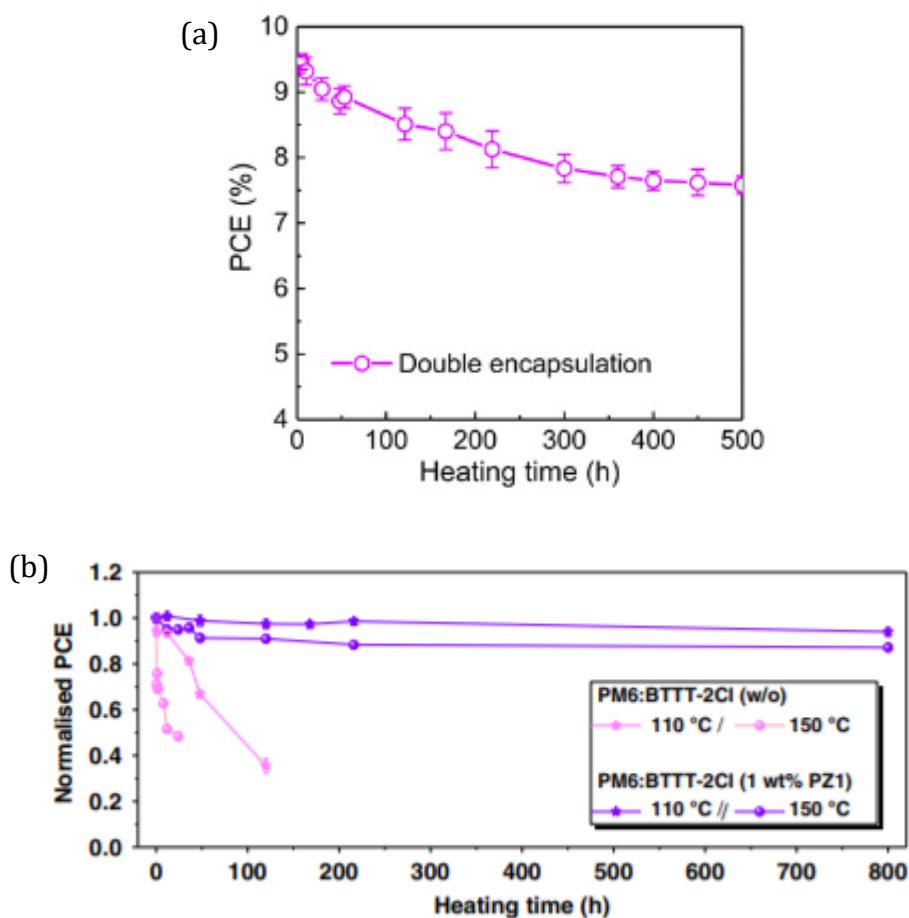


Figure 2.10: (a) Temperature stability of OPV at 100 °C employing a PBDTTT-OFT:PC<sub>71</sub>BM active layer [111]. (b) Normalised PCE of blends with different concentrations of PZ1 as a function of annealing time at 150 °C [110].

Thermal stability enhancement is often investigated in terms of active layer engineering and employing novel materials. OPV device efficiency and thermal stability was enhanced using a  $\pi$ -conjugated n-type conjugated polymer (PZ1) additive [110], composed of acceptor-donor-acceptor. This was constructed using an n-type organic semiconductor, IDIC-C16, as the key building block and thiophene as the linking units [112]. Commercially available PM6 was used as the donor material and BTBT-2Cl as the acceptor. In addition, this thermally stable bulk heterojunction photovoltaic system

displayed an improved PCE of 15.10% from 13.83% without the addition of PZ1. It was found that the improved film quality suppresses the thermally driven phase separation at high temperatures and was thermally stable at 150 °C for more than 800 hours in a nitrogen (N<sub>2</sub>) filled glovebox, as shown in Figure 2.10 (b). In addition, the PZ1 doping yielded more robust morphology against thermal cycling stress conditions [110].

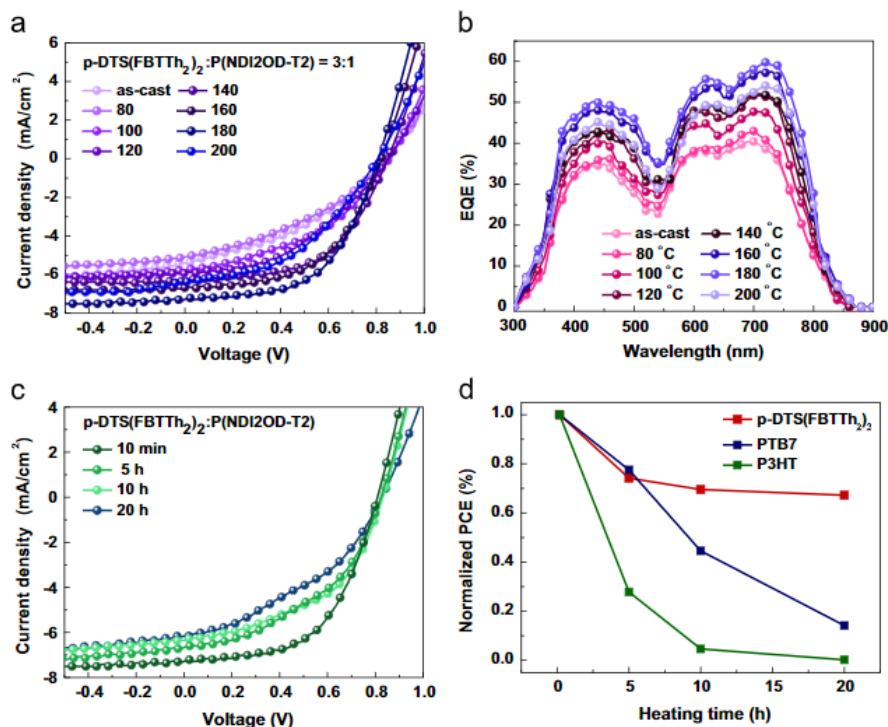


Figure 2.11: (a) J–V characteristics of the BHJ solar cell with a p-DTS(FBTTh<sub>2</sub>)<sub>2</sub>:P(NDI2OD-T2) (3:1) active layer at different annealing temperatures. (b) EQE spectra of the corresponding devices at different temperatures. (c) J–V curves of the BHJ solar cell based on the p-DTS(FBTTh<sub>2</sub>)<sub>2</sub>:P(NDI2OD-T2) blend after isothermal heating at 180 °C for 10 min, 5, 10, and 20 h. (d) Dependence of the efficiency on the heating time for devices made of p-DTS(FBTTh<sub>2</sub>)<sub>2</sub>, PTB7, and P3HT (heating temperature was 180 °C) [113].

Non – fullerene organic solar cells using a p-DTS(FBTTh<sub>2</sub>)<sub>2</sub>:P(NDI2OD-T2) bulk heterojunction have been found to be highly thermally stable [113]. It was determined that the p-DTS(FBTTh<sub>2</sub>)<sub>2</sub> compound enables stabilised blend morphology at high temperatures due to its lower diffusion kinetics. After thermally heating the OPV devices

for 20 hours at 180 °C, 70% of the initial PCE is retained. This compares favourably against devices using PTB7:P(NDI2OD-T2) and P3HT:P(NDI2OD-T2) BHJ active layers which retain 15% and 0% respectively under the same testing conditions. The results highlight the need for compounds which provide morphological stabilisation and provide promise for developing BHJ systems without the use of fullerenes. The J-V characteristics and EQE for different annealing temperatures are shown in Figure 2.11 (a) and (b) for a p-DTS(FBTTh<sub>2</sub>)<sub>2</sub>:P(NDI2OD-T2) bulk heterojunction, and the variation in J-V characteristics as a function of time for a p-DTS(FBTTh<sub>2</sub>)<sub>2</sub>:P(NDI2OD-T2) bulk heterojunction and the variation in normalised PCE over time for the three tested active layers are shown in Figure 2.11 (c) and (d) respectively.

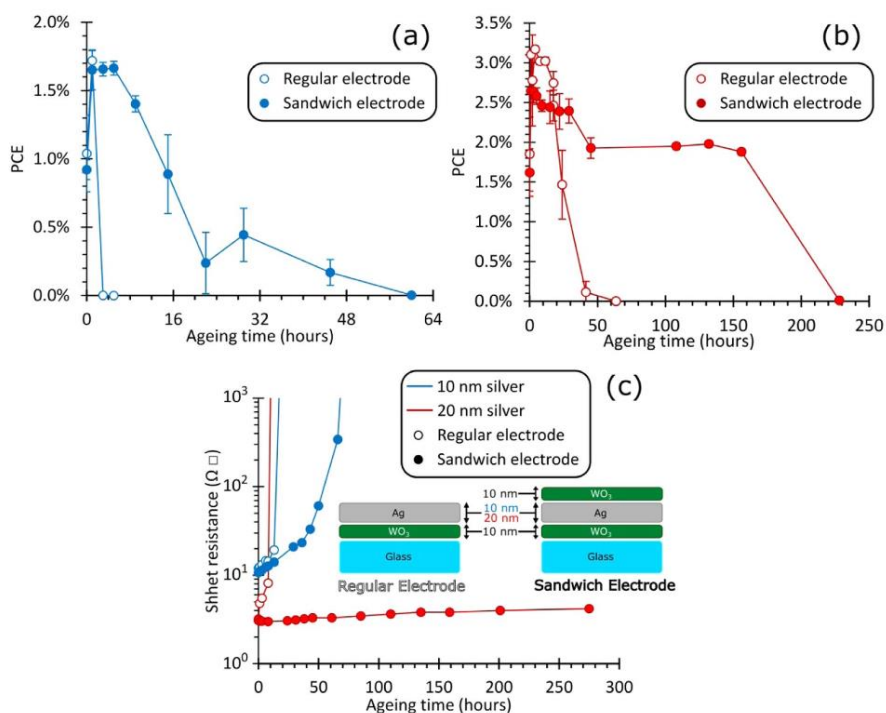


Figure 2.12: PCE data obtained from IV measurements of OPV devices with 10 nm (a) or 20 nm (b) silver electrodes in either the regular or sandwich configuration (c) obtained from measurements of 10 nm or 20 nm silver electrodes in either the regular or sandwich configuration as the devices are thermally aged at 85 °C in the dark [114].

OPV devices employing a thin silver electrode have been tested and it was found that they exhibited catastrophic failure even after modest levels of thermal ageing due to electrode delamination [114]. This failure was attributed to the formation of gaps in the silver electrodes, confirmed using atomic force microscopy and Rutherford backscattering [114]. It is concluded that these intrinsic vulnerabilities could be avoided using a sufficiently thick electrode. In addition, this issue can be mitigated using a thin layer of metal oxide on top of the electrode, significantly reducing the de-wetting process. This was achieved using tungsten oxide ( $\text{WO}_3$ ) to form a sandwiched electrode which leads to a stabilised electrode and was confirmed by observing a slower increase in sheet resistance when compared with the standard single layer electrode [114]. These results are shown in Figure 2.12.

### 2.4.3 Outdoor monitoring of OPVs

Outdoor monitoring has been a sparingly adopted approach for testing the stability of OPVs. By testing OPVs in outdoor conditions, multiple stress factors can be applied, and outdoor testing remains one of the best approaches to review OPV stability. One of earliest reports on outdoor testing of OPVs was conducted by Katz *et al.* in 2007 [115]. The long-term PV performance was investigated under outdoor conditions for three devices differing in terms of their active layer composition with encapsulation. The three active layers were a MEH-PPV:PCBM heterojunction, a P3HT:PCBM heterojunction and a P3CT- $\text{C}_{60}$  bilayer heterojunction.

In this work, the MEH-PPV:PCBM cell displayed the fastest degradation followed by P3HT:PCBM with significantly slower degradation and P3CT- $\text{C}_{60}$  being the most stable. The  $I_{sc}$  and  $V_{oc}$  was observed to restore for P3HT:PCBM and P3CT- $\text{C}_{60}$  devices after dark periods overnight. Overnight the devices were stored in a  $\text{N}_2$  filled glovebox in dark conditions. The same observations were made when the devices were kept in air in dark conditions overnight. Furthermore, when the P3CT- $\text{C}_{60}$  device was shadowed for 30 minutes during the daytime, complete recovery in  $I_{sc}$  and  $V_{oc}$  was observed. Katz *et al.* proposed new measurement protocols whereby the temperature dependence of PV parameters is recorded, the spectral response of the cells at various stages of degradation are measured and the variation in sunlight spectrum over the course of each day is measured.

Over the course of a month,  $I_{SC}$  values only recovered partly after each night whilst  $V_{oc}$  recovered completely after each night with almost no reduction in a long-term time scale [115]. Comparing the degradation of P3HT:PCBM and P3CT- $C_{60}$  devices, it was found that the evolution of FF differed significantly between the two devices; for P3HT:PCBM devices the FF reduced from 0.26 to 0.21 in a few hours, followed by a slow linear decay, whilst the FF for P3CT- $C_{60}$  devices remained approximately constant throughout the entire test process. In Figure 2.13, the  $I_{SC}$  and PCE variation is shown for MEH-PPV:PCBM cells in (a), the PCE and FF for P3HT:PCBM cells in (b) and PCE and FF for P3CT- $C_{60}$  cells in (c).

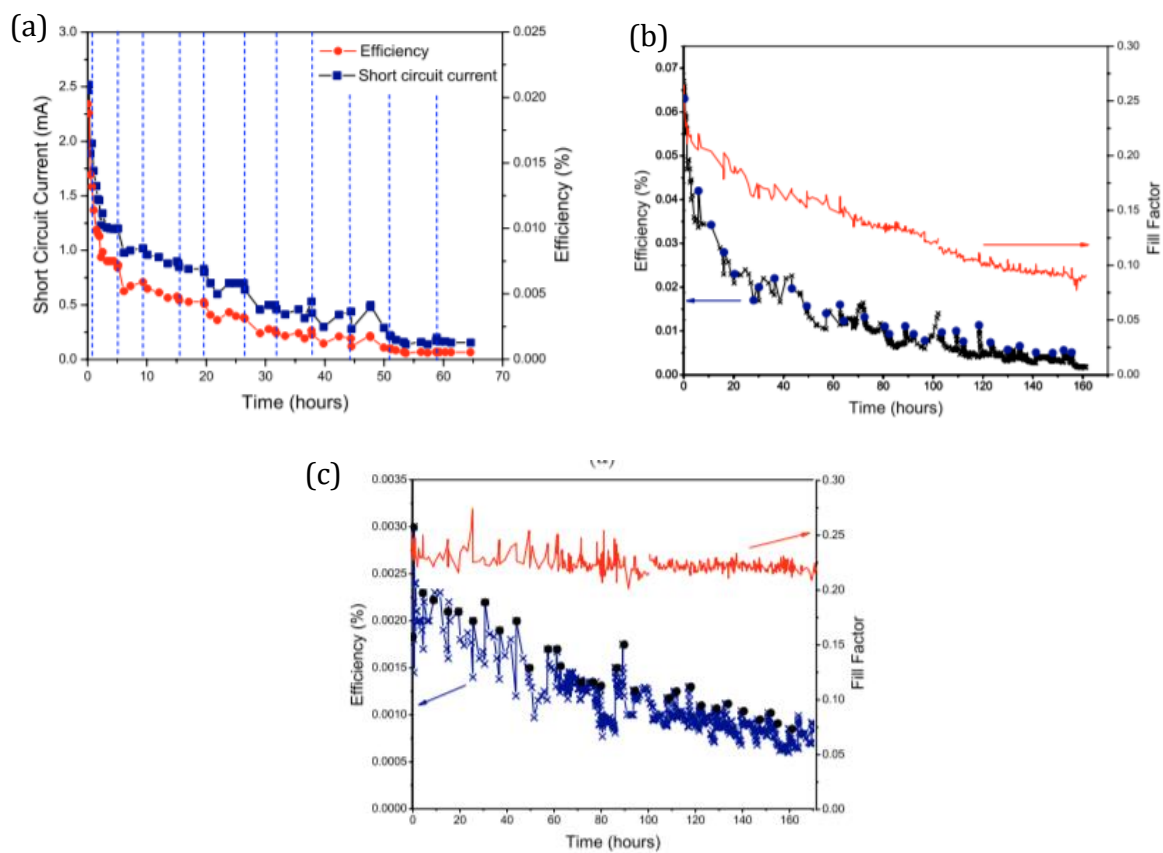


Figure 2.13: (a)  $I_{SC}$  and PCE variation for MEH-PPV:PCBM cells, (b) PCE and FF variation for P3HT:PCBM cells and (c) PCE and FF for variation for P3CT- $C_{60}$  cells, all tested in outdoor conditions during the day and stored in a dark,  $N_2$  glovebox overnight [115].

A number of outdoor stability studies were led by Krebs *et al.*. One of the most significant was the report on a solar park based on polymer solar cells, taking into

consideration the performance, the practicality, the installation speed and the energy payback time [116]. The analysis showed that a high voltage installation, where solar cells are all printed in series, enabled an installation rate that far exceeds any other PV technology which existed at the time. Figure 2.14 shows (a) the power extracted from 6 stretches of 100m solar cell foil at full sun, (b) the I-V curve from 126000 solar cells and (c) the efficiency over the course of 3000 hours.

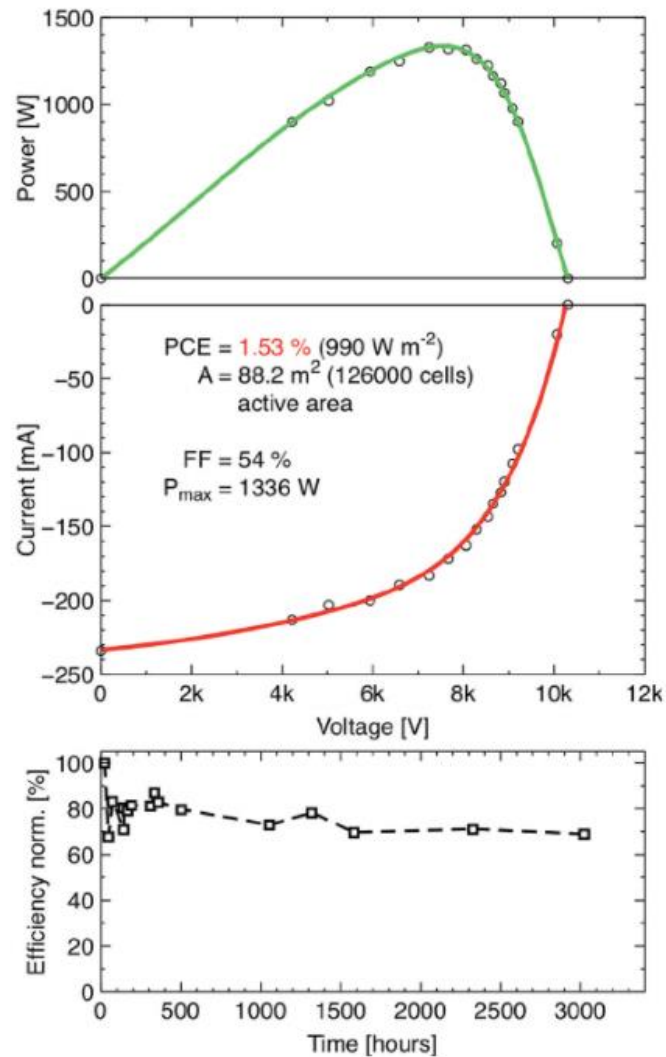


Figure 2.14: (a) The power extracted from 6 stretches of 100m solar cell foil at full Sun, (b) the I-V curve from 126000 solar cells shows a high fill factor of 54% and (c) after an initial drop in efficiency it stabilised at 1.5% mainly due to a drop in fill factor [116].



Life – cycle assessment (LCA) was used in this study to evaluate the solar park performance. LCA considers a number of aspects and allows for a comparison between the efficacy of the technology in relation to how quickly it can conquer back the energy used in the manufacturing [116] . This quantity is known as the energy payback time (EPBT). The EPBT for the installation of polymer solar cell foil on a wooden 250 m<sup>2</sup> platform was 277 days in Denmark and 180 days in southern Spain. Figure 2.15 shows the EPBT in days for the different components within the solar park, based on the results in this study.

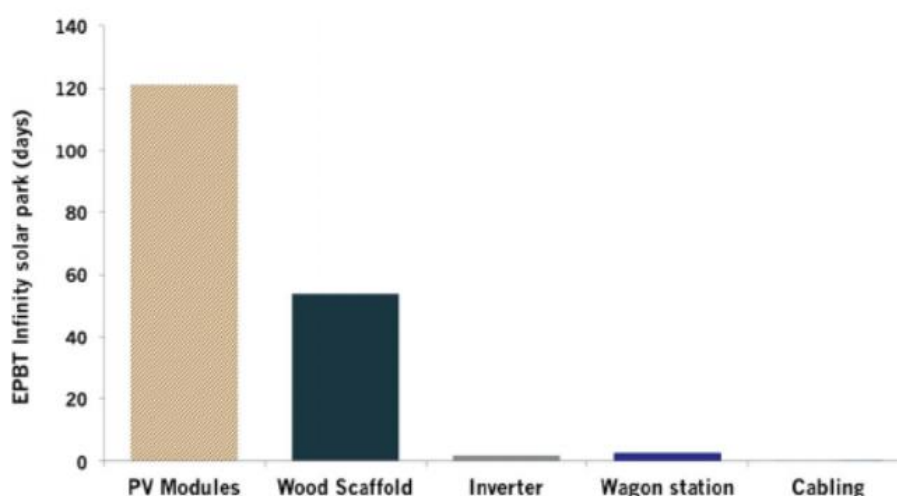


Figure 2.15: Energy payback time (EPBT) in days for the different components of the solar park [116].

The installation and deinstallation rates were estimated at 100 m min<sup>-1</sup>, which exceeds the total manufacturing speed of the polymer solar foil which was 1 m min<sup>-1</sup> for fully encapsulated and tested foil. Using the methods presented, simultaneous installation and deinstallation were possible, providing efficient schemes for decommissioning and recycling. New research directions are discussed by Krebs *et al.* whereby new and advanced materials must be developed with the potential for large scale application in solar parks, fast roll-to-roll processing using only abundant materials and finally the use of flexible substrates using low-cost barriers and adhesives. Finally,

Krebs *et al.* note how this research field is directed by LCA and that the short EPBT obtained in their research was achieved by choosing a wooden structure, roll-based installation and high voltage connections.

Studies on small molecules are rare. However, boron subphthalocyanines (BsubPc) and associated structural variants have been investigated for OPVs as light absorbing and electron donating materials paired with buckminsterfullerene ( $C_{60}$ ). These variants are chloro boron subphthalocyanine (Cl-BsubPC), phenyl-BsubPc (Ph-BsubPc), phenoxy-BsubPc (PhO-BsubPc) and  $\alpha$ -sexithiophene. The devices are subsequently tested according to ISOS-O-3 protocols. Constant current measurements and current – voltage sweeps, reinforced by irradiance and temperature tracking, are measured. These measurements reveal that a different chemical structure for the boron subphthalocyanine leads to different degradation rates [117]. These results suggest that the initial burn – in efficiency loss is due to  $C_{60}$ . However, the long – term degradation is attributed to the chemical breakdown of the subphthalocyanine donors through hydrolysis. Two main conclusions are that the molecular structure of boron subphthalocyanine significantly dictates the device lifetime, allowing a structure – property relationship to be established for stability. Furthermore, the results highlight the need to substitute  $C_{60}$  with an alternative for pairing with boron subphthalocyanine [117]. This work was continued on small molecule OPVs employing boron subphthalocyanine derivatives as the electron acceptor and  $\alpha$ -6T as the electron donor. The three derivatives tested were Cl-BsubPc, PhO-BsubPc and Ph-BsubPc. An epoxy glue and glass cover slip were applied to each device as encapsulation. The highest performance and stability were observed for devices using Cl-BsubPc. It is thought from this investigation that the stability of the other two derivatives is reduced by the phenoxy or phenyl functionalization in the molecular axial positions [118].

In Wales, Bristow *et al.* have performed several studies on the outdoor performance and characteristics of OPV. Firstly, they investigated the temperature and irradiance dependence of OPV module performance in [119]. The performances of OPVs are benchmarked against c-Si modules where OPVs are found to display lower performances under low light conditions. It is demonstrated that the low light performance of OPV is associated with the lower performance of  $V_{OC}$  and FF which are limited by the occurrence of inflexion points in the I-V characteristics; further

information regarding this behaviour and S-shaped I-V curves can be found in [120]. The occurrence of such S-shaped I-V curves often indicates a bottle neck in the charge transport capabilities [119]. However, these inflexion points can be removed and mitigated by photo – annealing at high irradiances, although this has a negative impact on the OPV performance during overcast conditions [119]. Furthermore, Bristow *et al.* show that OPV modules possess a positive temperature coefficient of  $+0.007\%K^{-1}$  as compared to  $-0.341\%K^{-1}$  for c-Si modules [119]. Figure 2.16 illustrates the variation of OPV PCE as a function of temperature at irradiances of  $600\text{ Wm}^{-2}$  and  $1000\text{ Wm}^{-2}$  [119]. However, the degradation over the course of 12 weeks was significantly higher than for c-Si modules and the study highlighted the need to develop improved barrier layers, UV filters and edge sealants. However, this study is highly dependent on the type of module studied and different values could lead to different values for temperature coefficient.

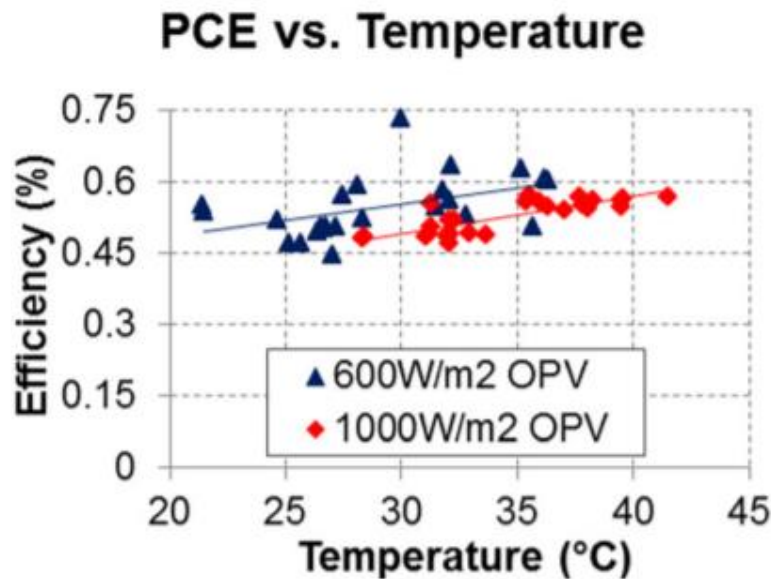


Figure 2.16: Temperature dependence of PCE measured at fixed irradiance of  $(600 \pm 10)\text{ Wm}^{-2}$  and  $(1000 \pm 10)\text{ Wm}^{-2}$ . Linear fitted curves are applied to measure the temperature coefficient of the module [119].

A comprehensive study comparing several different OPV technologies corresponding to 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> generations was conducted by Bristow *et al.* [121]. The

performance parameters of each of these technologies was studied as a function of irradiance and it was found that both dye-sensitised solar cells (DSSC) and OPV display notably distinct characteristics in comparison to crystalline technologies, as shown in Figure 2.17 (a) [121]. These differences are attributed to the differences in the charge transfer mechanisms between each of the technologies [121].

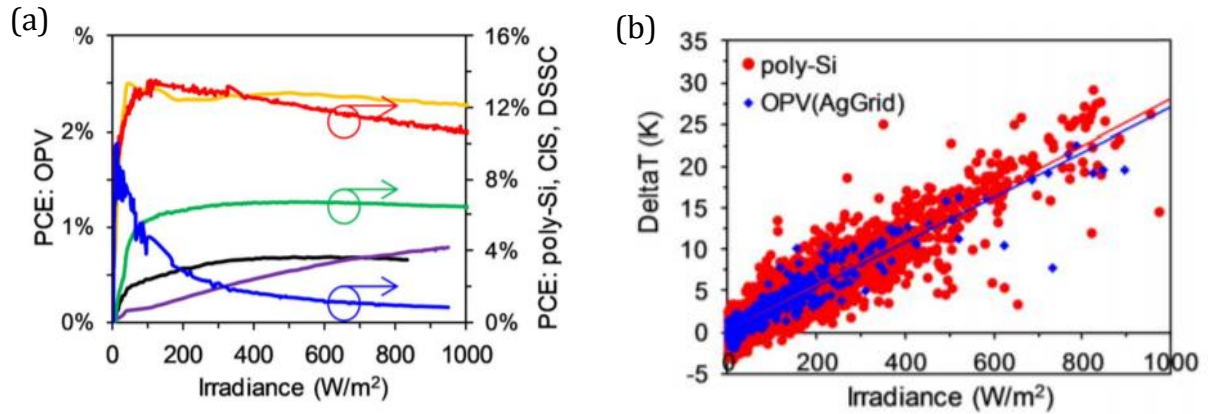


Figure 2.17: (a) Comparison of performance parameters for poly-Si, CIS, DSSC and three different OPV modules and (b) effect of module temperature rise above ambient as a function of irradiance. The gradient gives the Ross coefficient [121].

The change in module temperature as a function of irradiance is investigated for both poly-Si and the OPVs and the Ross coefficient for each are found to be  $0.028 \text{ K m}^2 \text{ W}^{-1}$  and  $0.027 \text{ K m}^2 \text{ W}^{-1}$  respectively, as illustrated in Figure 2.17 (b). The Ross coefficient was calculated using,

$$T_{\text{Delta}} = T_{\text{Module}} - T_{\text{Ambient}} = kG, \quad 2.6$$

where  $T_{\text{Delta}}$  is the temperature change of the module,  $T_{\text{Module}}$  is the module temperature,  $T_{\text{Ambient}}$  is the ambient air temperature,  $G$  is the irradiance and  $k$  is the gradient of the temperature change versus irradiance plot, representing the Ross coefficient [121]. In addition, the effect of wind speed on the Ross coefficient illustrated the impact of the different materials used in the fabrication of the modules [122]. The Ross coefficient is influenced by wind speed, as given by [123],

$$k = a + be^{-cv}, \quad 2.7$$

where  $v$  is the wind speed and  $a, b, c$  are fitting coefficients for the Ross coefficient as a function of wind speed. The module temperature is preferred over the ambient temperature since this gives a better indication of the heating effect of the OPV module.

Burn – in is affected by the irradiation which causes a photo – induced reaction in the active layer and the formation of sub-band gap states [124]. These sub-band gap states lead to a reduction in the  $J_{sc}$  and  $V_{oc}$  by increasing the recombination rate, reducing the exciton lifetime and reducing the diffusion length. In addition, the charge carriers can fill these sub-band states near to the quasi-Fermi level, leading to a reduction in the  $V_{oc}$ . This arises from the quasi-Fermi levels moving away from the donor HOMO and acceptor LUMO levels and into the energy gap between these two levels. Burn – in observed for OPV technologies has been attributed to several possible factors [124]. The origins of burn – in for OPV are further discussed in [125], and the possible mechanisms are reported as macro – phase separation of the polymer:fullerene layer and the formation of a charge blocking layer at the electrode interface.

## 2.5 Application of ML in OPV and PSC development

There has been an increasing prevalence in the use of machine learning to enhance the understanding of material selection in a wide range of application areas (battery, solar cell, chemical synthesis research) [126]. However, by utilising machine learning methods, predictions of emerging PV technologies can be made as well as predict the future lifetimes and performances of emerging technologies; ML can be used to find new and better combinations of materials and structures to enhance both stability and performance. Indeed, the hidden underlying causes of OPV degradation can be extracted that goes beyond the standard approach of acquiring specific information by directly measuring the stability due to one or more changes. By deploying data analytical approaches in this manner, it is possible to determine which materials and stress factors have the greatest impact on the device stability. In addition, by acquiring an

understanding of the materials and environmental attributes, which lead to more stable devices, the physical phenomena, which lead to degradation, can also be better understood. This could represent a potential new paradigm in understanding OPV reliability.

### 2.5.1 Use of Data Analytics for Material Studies

Material discovery is one of the main fields of research using ML approaches in PV. One of the first such studies report on the use of ML to pre-screen new OPV applicable materials using ML models and subsequently only focussing on molecules which “passed” the ML assessment [126]. Figure 2.18 illustrates the structure of the convolutional neural network. Several ML algorithms were employed including back propagation, neural networks, deep learning, support vector machine (SVM) and random forest (RF). Their analysis studied a total of 1719 potential OPV donor materials and their method successfully allows the OPV PCE class to be predicted. In addition, they determine the RF algorithm to be capable of handling long inputs even in the presence of noise, due to RF’s ability to choose multiple features rather than the complete content of the input [126].

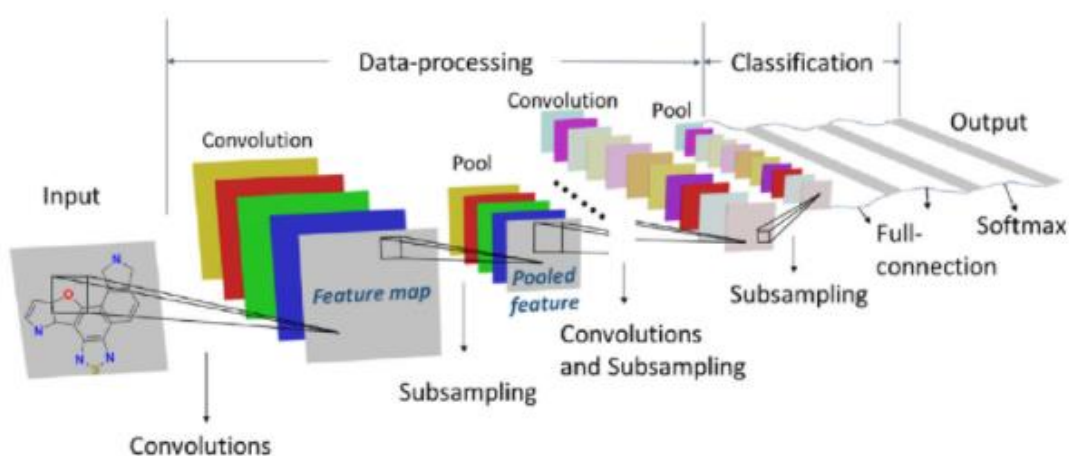


Figure 2.18: Structure of our convolutional neural network [126].

The multilayer perceptron (MLP) ML algorithm was used to analyse a dataset from the Harvard Clean Energy Project. MLP algorithms are used since they are known to be good at supervised regression problems [127]. This method is shown to substantially reduce the fraction of the screening library that is calculated. This is achieved using a high-throughput virtual screen. The training attributes used focussed on the molecular graph which contain knowledge of the type and connectivity of the atoms. Four molecular parameters are employed consisting of the hashed atom-pair fingerprint, the MACCS key-based fingerprint, the Morgan circular fingerprint and the hashed topological torsion fingerprint. Subsequently, the HOMO, LUMO and PCE are predicted based on the chemical properties of the materials and a Pearson  $r$  of 0.99 is achieved in each case for the validation set. The study does not return a general neural network to predict different molecules but does exploit the local nature of the chemical space used with high-throughput virtual screens and, therefore, led to significant gains in predictive power on relevant molecules. However, it is noted that this can come at the expense of transferrable neural networks [127].

ML can be applied to PV applications in order to improve the stability and performance. Employing ML in this way has proven successful in the field of PSCs for stability engineering of Halide Perovskite [128]. DFT is used in conjunction with ML to calculate the decomposition energies which are thought to correlate with the thermodynamic stabilities of PSCs. ML techniques allowed the stabilities of 14,190 perovskite materials to be investigated. Figure 2.19 shows the ML predicted decomposition energies for mixed perovskite (a)  $\text{ABi}_3$  and (b)  $\text{APbX}_3$ .

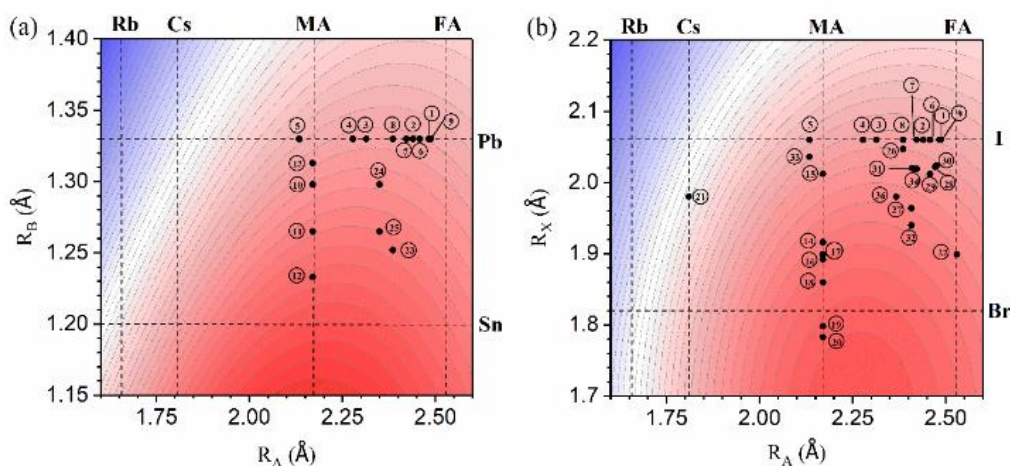


Figure 2.19: The ML-predicted decomposition energies for mixed perovskites (a)  $ABX_3$  and (b)  $APbX_3$ . The x/y axis are ionic radii of A/B and A/X in (a) and (b), respectively [128].

Data analytical techniques have been employed to analyse trends and patterns present in a database reporting on perovskite performance reports between 2013 and 2018 [129]. Therein, the key materials necessary to improve the performance of different architecture PSC are found. For example, poly – triarylamine (PTAA) as an HTL, a mixed cation perovskite and bathocuproine (BCP) as an ETL are significant for inverted architecture PSCs. For conventional architecture devices, using chlorobenzene as an antisolvent, using two or three spinning processes and using LiTFSI + TBP + FK209 as HTL additive and  $SnO_2$  as an ETL led to improved performance.

To design and produce efficient materials for OPVs, the parameters which control their material properties could be analysed. It has been shown that the PCE of OPVs can be predicted by developing a model which uses the largest number of these parameters, called descriptors [130]. In the study, 280 small molecule OPV systems are used to produce a dataset. Thirteen descriptors are employed as the attributes to predict the PCE of the OPV systems. The descriptors used were:

1. Number of unsaturated atoms in the main conjugation path of donor molecules
2. Vertical ionisation potential of donor molecules
3. Polarizability of donor molecules
4. Energy of the electronic transition to a singlet excited state with the largest oscillator strength
5. Reorganization energy for holes in donor molecules



6. Hole-electron binding energy in donor molecules
7. The energetic difference of HOMO of donor and LUMO of acceptor
8. Energy of the electronic transition to the lowest-lying triplet state
9. The energetic difference of LUMO of donor and LUMO of acceptor
10. Change in dipole moment in going from the ground state to the first excited state
11. The energetic differences of HOMO and HOMO-1 of the donor molecules
12. The energetic differences of LUMO and LUMO+1 of the donor molecules
13. The energetic difference of LUMO and LUMO+1 of the acceptors.

These 13 descriptors were selected since they are properties of the donor and acceptor material and can be easily computed from electronic structure codes, subsequently allowing for rapid screening of many new compounds [130]. The acquired model was able to predict the PCE based on the above descriptors with a correlation of  $r = 0.79$  for the test set of 30 molecules, using a gradient boosting regression tree (GB) ML algorithm. Several other algorithms were tested including linear regression,  $k$ -NN, ANN and RF, however, GB yielded the highest correlation. The descriptor importance of GB is analysed. The descriptor importance is estimated by keeping track of the reduction of mean square error for each descriptor when data passes through the trees and averaging it over all trees in the ensemble [130]. The regression fitting using GB for the testing set and using leave-one-out cross-validation are shown in Figure 2.20 (a) and (b) respectively. The descriptor importances for (c) GB and (d) RF are additionally shown in Figure 2.20.

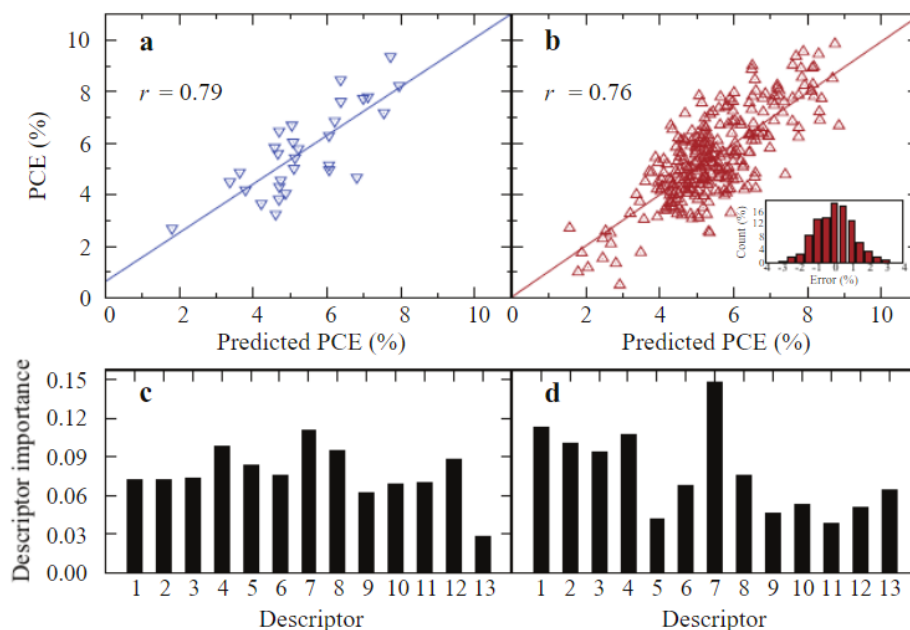


Figure 2.20: Theoretically predicted versus experimental PCE for (a) the testing set (30 molecules) and (b) all data points using the leave-one-out cross-validation technique for the GB model. Inset shows probability density of prediction errors. The descriptor importance for the (c) GB and (d) RF model are depicted [130].

## 2.5.2 Application of ML in Forecasting

ML can additionally be applied for forecasting and prediction tasks. To date, this has not been performed for OPVs or PSCs but has been applied to c-Si modules. This provides a method to determine the potential power outputs of PV technologies, as well as predicting the degradation of modules and devices based on significantly shorter durations of testing.

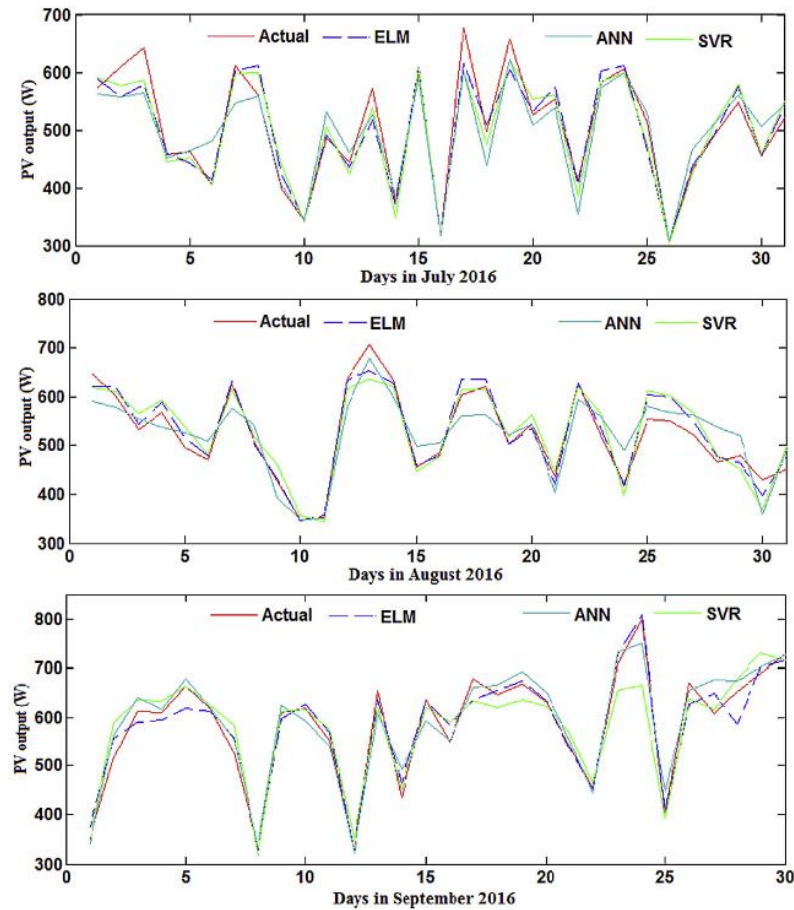
Fault detection in PV power stations needs to be rapid and effective; if a fault develops and remains undetected, then this can severely impair the plant's power generation ability. Fault detection strategies have been used where an ML technique is merged with statistical hypothesis testing [131]. A Gaussian process regression technique was employed as a modelling framework and a generalised likelihood ratio test

(GLRT) chart was used to detect PV system faults. Furthermore, a method for failure detection has been developed using ML and a probabilistic neural network classifier [132]. In this work, a database was developed which comprises both healthy and faulty periods of operation and the model is trained to classify operation as faulty or healthy. Subsequently the second classifier attempts to diagnose the type of fault.

The solar energy potential of rooftop photovoltaics has been quantified using ML techniques where the support vector machine (SVM) algorithm and geographic information systems (GIS) was used to provide an estimation of the total potential of such energy sources for urban areas representing the smallest administrative division in Switzerland [133]. The methodology allowed 1901 out of the total 2477 communes in Switzerland to be estimated. Using a 6-fold cross validation method, the normalised RMSE is employed as a metric to validate the SVM model predictions. Following this approach, it is found that 81% of the ground floor area of each building represents available roof top area for potential PV applications; this corresponds to 328 km<sup>2</sup> of roof space within  $\pm 90^\circ$  of due south. This implies a potential PV electricity generation of 17.6 TWh if all roofs have PV installed.

It has been shown that the PV power output from a solar plant can be forecasted based on the meteorological variables from a numerical weather forecast [134]. A model was built using weather forecast data as well as reference AC power measurements of PV plants. The method employed utilises a quantile regression forests machine learning algorithm to forecast the AC power. This provides a means of predicting the hourly AC power production one 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) [135]. 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. Figure 2.21 illustrates the results for three months in 2016 using ELM and an artificial neural network (ANN) and support vector regression (SVR) ML algorithms for comparison. The forecasting results for one plant is shown out of a total of three tested as an example.

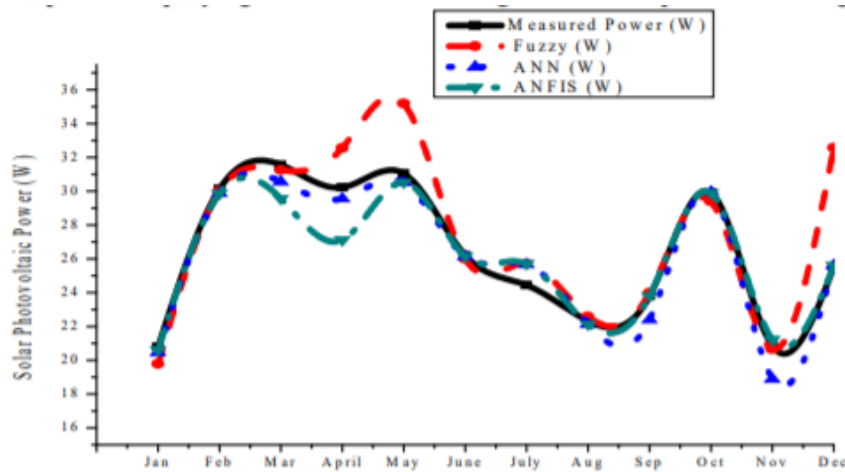


*Figure 2.21: Comparison of 1 day ahead forecast generated based on the ELM and other models using the data from plant 1 with data size  $N = 274$  and  $M = 92$  where  $N$  and  $M$  are the sizes of the training and testing sets respectively [135].*

Using an ELM ensures that a high learning rate is attained with a high performance and is found to outperform both ANN and SVR. When the results from the ELM prediction are compared with the results from the SVR model and the ANN model, it was observed that the ELM displays a higher accuracy and is less computationally demanding method for forecasting daily and hourly output power [135].

Forecasting the short – term PV power generation has additionally been tested using intelligent modelling techniques: fuzzy logic, ANN and adaptive neural fuzzy interface systems (ANFIS) [136]. This was performed based on the sky – conditions. Subsequently the intelligent modelling techniques were used to forecast the power output from a 210 W Heterojunction with Intrinsic Thin Layer (HIT) PV modules for several different sky – conditions, such as clear, hazy, partially foggy/cloudy and fully foggy/cloudy for composite climate zone. The ANFIS model is found to outperform the

other methodologies used. The ANFIS model has the benefit of integrating both fuzzy logic methods and ANN, thus increasing the system accuracy and provides a faster system response; this makes it more suitable for short term, forecasting and monitoring systems [136]. Figure 2.22 illustrates the measured and forecasted PV power employing fuzzy logic, ANN and ANFIS methodologies [136].



*Figure 2.22: Graphical analysis of measured and forecasted PV power employing fuzzy logic, ANN and ANFIS methodologies [136].*

There have also been efforts to improve the real – time control performance of PV systems using a weather – based hybrid method for 1 day ahead hourly forecasting of PV power output [137]. Three stages were implemented: classification, training and forecasting. For classification, self – organising map (SOM) and learning vector quantisation (LVQ) networks are used to classify historical PV output data. The SOM network is employed to select different weather types based on different PV power generation patterns; similar days of weather conditions are defined as having the same weather type. The LVQ network is then used to assign the features selected by SOM to individual classes. Combined, the SOM and LVQ networks work as a pattern classification system. Training is performed using an SVR algorithm to train on the datasets using temperature, probability of precipitation and solar irradiance of similar hours as input parameters. This is performed on several different weather types, producing six SVR sub-models for sunny, sunny and cloudy, cloudy and sunny, cloudy and rainy, and rainy days based on the historical weather and PV generation data. Finally, a forecasting stage is

conducted in order to more accurately predict the weather conditions for every 3 hours of the next day and to predict the PV power output. This is performed using the fuzzy inference method which selects the appropriate SVR sub-model in the forecasting stage [137]. It is noted that the PV power generation is highly fluctuating, due to variations in climatic conditions. The method updates the forecast in real – time, thus overcoming the problem of large PV output fluctuations.

From the literature it is apparent that many of the ML forecasting approaches tested thus far focus on the short-term or real-time prediction. Therefore, developing means of forecasting the long-term performance could prove useful where the performance can be predicted months, or even years in the future based on smaller training sets.

## Chapter 3. Experimental and Computational Methods

### 3.1 Introduction

The following chapter outlines the experimental and computational methods employed in this thesis; this includes the fabrication of inverted PSCs; the testing and characterisation techniques for stability assessment; the preparation and testing methods for OPVs in outdoor conditions; outdoor data acquisition using an Access database and finally the methods of data analysis and machine learning.

PSC device fabrication and characterisation were conducted in the School of Electronic Engineering at Bangor University in a Class 1000 cleanroom and nitrogen filled glovebox with oxygen levels maintained below 1 ppm and water (H<sub>2</sub>O) levels below 1%. Carbon based PSCs were also fabricated at Swansea SPECIFIC and delivered to Bangor university for testing. OPV modules for outdoor testing were supplied by CSEM Brazil and Infinity PV; testing of CSEM modules was performed at the School of Electronic Engineering at Bangor University as well as at CSEM Brazil, Belo Horizonte, by Gabriela Amorim. The initial collation of stability and performance data of OPVs was conducted by Dr Suren Gevorgyan at the Danish Technical University (DTU). All data analytical and machine learning approaches were performed at the School of Electronic Engineering at Bangor University.

### 3.2 CSEM module preparation

These methods outline the methods for the data presented in Chapter 5. OPV modules (21.6 cm<sup>2</sup>) were supplied by CSEM Brasil for outdoor testing, under ISOS – 0 – 2 protocols corresponding to the following conditions:

- Outdoor conditions
- Natural sunlight illumination for testing and characterisation
- Ambient temperature and humidity
- Kept at either maximum power point or open – circuit voltage)

The modules were made by Gabriela Amorim and team. 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) was used as the hole transport layer. Active layer formulations were acquired from Merck with the acceptor being based on a fullerene – derivative. 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 polyethylene terephthalate (PET) based barrier film with a water vapour transmission rate of the order of  $10^{-3} \text{ g cm}^{-2} \text{ day}^{-1}$  from Mitsubishi. A Delo epoxy-based UV curable adhesive with barrier properties of  $6 \text{ g cm}^{-2} \text{ day}^{-1}$  was applied using a roll-to-roll (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  $40 \mu\text{m}$ . Once received, modules were tested in Bangor and cabling added by ultrasonic soldering.

CSEM modules, consisting of 6 cells, with an inverted architecture were installed in outdoor conditions at the School of Electronic Engineering, Bangor University and at CSEM Brasil, Belo Horizonte. The cell architecture and module configuration are shown in Figure 3.1. CSEM modules were installed on the roof of the School of Electronic Engineering at an orientation angle of  $35^\circ$  facing South and at  $20^\circ$  facing North in Belo Horizonte; the angles are taken from the vertical and remain constant throughout the test. These angles are the optimal values to maximise the average incident solar illumination throughout the year and are determined based on the latitude of the location. Modules were connected to resistive loads such that at standard test conditions (STC) they operate at the MPP. An image of the installed modules is shown in Figure 3.2. Each OPV module is fixed on raised studs, thus allowing free flowing air around the modules.



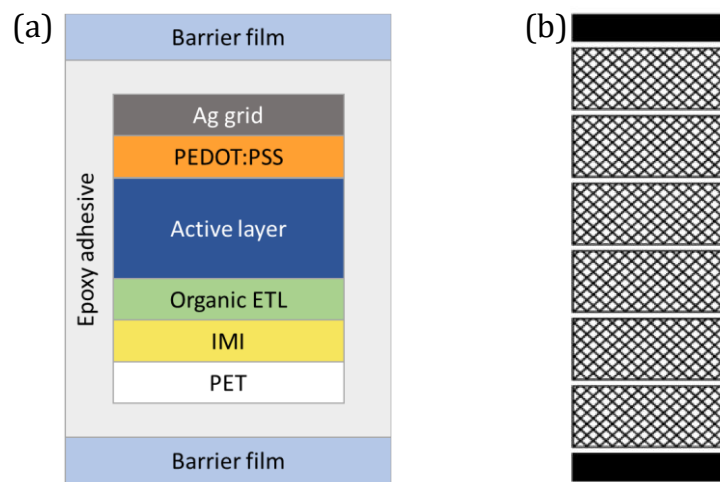


Figure 3.1: (a) Device architecture of CSEM devices and (b) Module layout consisting of 6 cells.

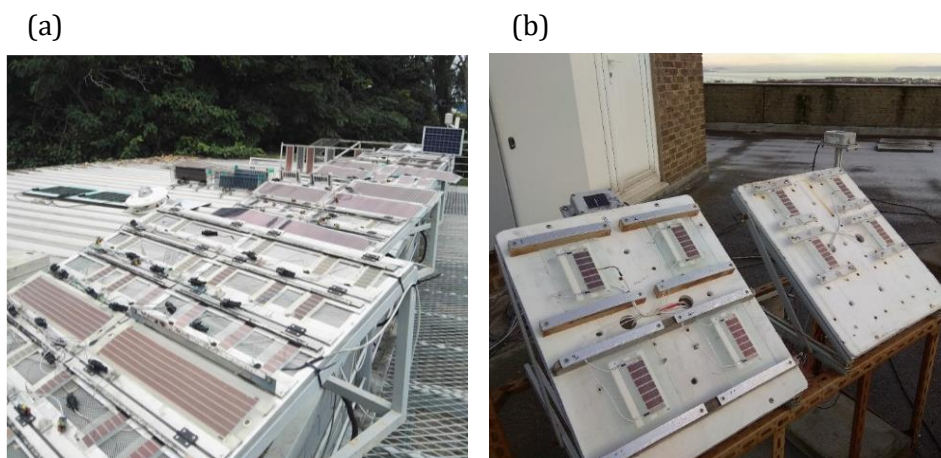


Figure 3.2: CSEM modules installed at CSEM, Brazil and on roof of the School of Electronic Engineering, Bangor University.

### 3.3 Characterisation Techniques

#### 3.3.1 I – V Characterisation

Current – voltage (I-V) measurements were performed in order to determine the solar cell performance parameters of all cells under test. These include the short circuit current density, the open circuit voltage, the maximum power point current and voltage, and the fill factor, as discussed in section 2.1.3. The measurements were performed using a Newport Oriel™ Sol1A class A solar simulator, having a xenon bulb. This solar simulator provides an irradiance of  $100 \text{ mWcm}^{-2}$  with an AM1.5G spectrum. The illumination area of the solar simulator is approximately  $10 \text{ cm}^2$ , providing ample area for testing multiple small area devices simultaneously. The I-V measurements were recorded using a Botest Systems GmbH source measure unit (SMU), supplied with a multiplexer allowing up to 8 devices to be measured sequentially. For stability assessment measurements, devices were placed under the solar simulator and regular I-V measurements were taken for ISOS-L-2 protocols and were covered for ISOS-D-2 and ISOS-D-3 protocols. Between successive I-V measurements, devices were kept at  $V_{oc}$ .

#### 3.3.2 Absorption Spectroscopy

UV-vis absorption spectroscopy measurements were performed using an Ocean – Optics USB – 4000 spectrometer with multiplexed optical fibre feed – through connections having a  $400\mu\text{m}$  core diameter. A sulphur plasma lamp with an irradiance of  $100 \text{ mWcm}^{-2}$ , calibrated using a silicon reference cell from RERA, Netherlands. All measurements were performed in the transmission mode. To measure the SEM, thin films of HTL and FMC were deposited on glass substrates without the ETL or electrodes deposited.

#### 3.3.3 Scanning Electron Microscopy

Scanning Electron Microscopy (SEM) measurements were performed using a Hitachi TM3000-table top SEM. To measure the SEM, thin films of HTL and FMC were deposited

on glass substrates without the ETL or electrodes deposited in order to determine the effect of the underlayer (HTL) on the perovskite film morphology.

### 3.3.4 Electrochemical Impedance Spectroscopy

Electrochemical Impedance Spectroscopy (EIS) measurements were performed using an AMETEK™ 1260A Impedance Analyser with an AMETEK™ 1296A Dielectric Interface System. The 1260A Impedance Analyser applies a sinusoidally varied variable voltage (AC) to the material with a controllable applied bias voltage (DC). The frequency of the applied AC voltage is varied from high frequency to a low frequency range; the 1260A Impedance Analyser is capable of scanning between 32 MHz and 10  $\mu$ Hz. The impedance is then calculated from the ratio of the voltage to the current, which varies as a function of the frequency according to the properties of the material under test. These measurements can shed light on the physical structure of the material as well as chemical processes occurring within the material. EIS allows data to be acquired in a non – destructive manner and, significantly, allow effects due to electrodes, diffusion and charge – transfer to be distinguished across different frequency domains.

### 3.3.5 Environmental Test Chambers For accelerated testing of PSCs

Accelerated lifetime testing was performed using a CME Envirosystems™ PAC series climatic chamber. The climatic chamber allows for both temperature and humidity control with a temperature range of -40°C to 180°C and a humidity range of 0% to 100%. This allowed devices to be tested according to ISOS – D protocols.

The climatic chamber was employed for testing PSCs under damp – heat conditions. The chamber was set to a temperature of 65°C and 35% relative humidity (RH) and J-V characteristics of the PSCs was determined as a function of time; the EIS spectra were also measured at the same time in order to identify the changes occurring within the PSCs due to damp – heat exposure. The lower RH was chosen, in contrast to the standard ISOS – D – 3 conditions of 65°C and 85% RH due to the rapid onset of degradation observed when devices were tested at these conditions. The devices would

degrade within a few hours of being exposed; in order to gain meaningful insight into the changes occurring within the PSCs, a longer test duration was required.

## 3.4 Data Analytical and Machine Learning Approaches

### 3.4.1 Principal Component Analysis

Principal component analysis (PCA) was used in Chapter 5 to elucidate trends in the datasets and to determine which environmental factors contribute to the performance parameters and degradation rates of OPV modules. PCA is a dimensionality reduction technique for analysing large datasets developed by Karl Pearson in 1901 [138]. 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 [139].

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 [140]. The dimensionality of the dataset can be reduced by removing redundant dimensions, keeping only the most important ones.

The variance in a quantity  $x$  is defined as,

$$\text{var}(x) = \frac{\sum (x_i - \bar{x})^2}{N}, \quad 3.1$$

where  $N$  is the number of data points. Similarly, the covariance for quantities  $x$  and  $y$  is defined as,

$$cov(x) = \frac{\sum (x_i - \bar{x})(y - \bar{y})}{N}, \quad 3.2$$

where  $\bar{x}$  and  $\bar{y}$  are the mean values of each quantity.

Variance is a measure of the variability and measures the spread of the data; it averages the squared deviation from the mean score. Covariance measures the extent to which corresponding elements from sets of ordered data move in the same direction. In order to find patterns in the data sets, we require the data to be spread in all dimensions. We also require the dimensions to be independent. If the data has a high covariance, when represented in some  $n$  number of dimensions, then we replace those dimensions with linear combinations of those  $n$  dimension. Each of the new dimension determined from PCA are orthogonal to each other – linearly independent – and ranked according to the variance of the data along them [141]. The new dimensions found are referred to as the principal components (PCs) and represent the new variables describing the dataset. In order to perform PCA, a normalisation process is required for the dataset [142]. The normalisation is achieved *via* mean centring, where the mean value is subtracted from the value of each instance. This is required such that the sample mean is zero and the data is centred on the origin of the origin of the coordinate system. In addition, the variance has been scaled to unity. These pre-processing procedures ensure that each variable contributes equally, and no single variable dominates the analysis.

The process of PCA is summarised below,

1. Calculate the covariance matrix  $X$  of data points.
2. Calculate eigenvectors and corresponding eigen values.
3. Sort the eigenvectors according to their eigenvalues in decreasing order.
4. Choose first  $k$  eigenvectors and that will be the new  $k$  dimensions.
5. Transform the original  $n$  dimensional data points into  $k$  dimensions.

A schematic illustration of PCA applied in three dimensions, for visualisation, is shown in Figure 3.3.

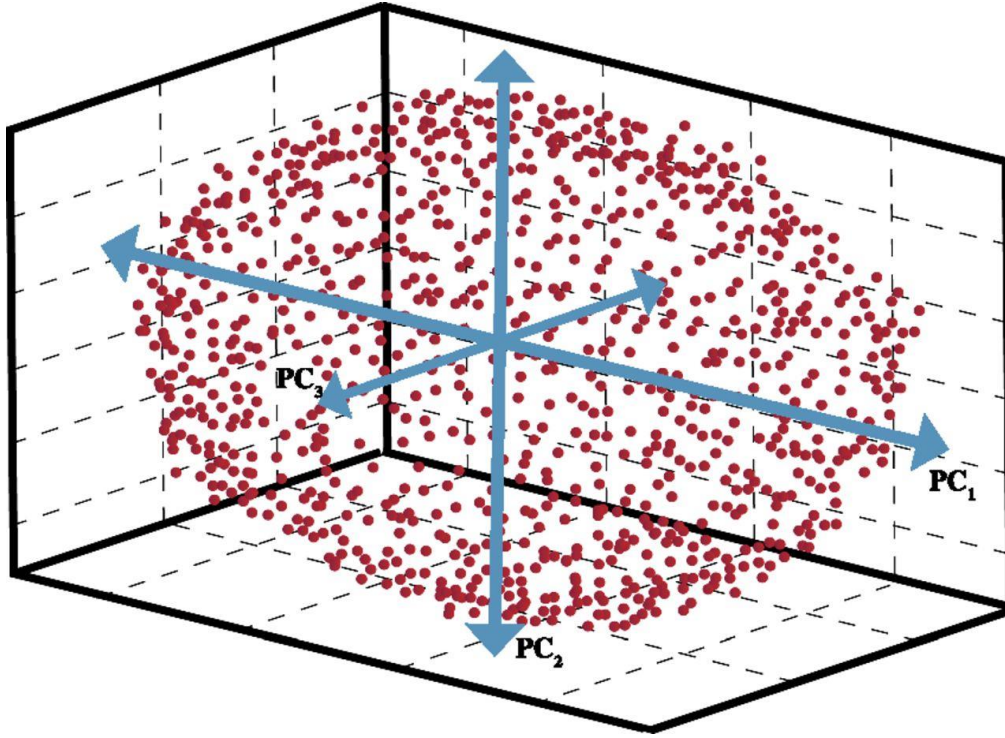


Figure 3.3: Schematic illustration of PCA applied in three dimensions [141].

Given a matrix  $\mathbf{X}$  of  $n \times m$  dimensions, such that there are  $n$  data points of  $m$  dimension, then the covariance matrix is calculated from,

$$\mathbf{C}_x = \frac{1}{n-1} (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^T, \quad 3.3$$

where  $n$  is the number of scores in each set of data,  $\bar{\mathbf{X}}$  is the mean of  $n$  scores in the first set of data and  $\mathbf{X}^T$  is the transpose of  $\mathbf{X}$ .

The covariance matrix contains the variance of the dimensions as the diagonal elements and the covariance of the dimensions as the off-diagonal elements. We need to remove correlated dimensions; covariance among dimensions should be zero. This implies that there should be a large number of main diagonal elements whilst off diagonal elements should be zero. Before, performing PCA, the data should be normalised. If the features are of different scales, a correlation matrix can be used instead. The objective of PCA is summarised below.

1. Find linearly independent dimensions (or basis of views) which can represent the data points without loss.
2. Those newly found dimensions should allow us to predict/reconstruct the original dimensions. The reconstruction/projection error should be minimised.

Consequently, the original variables can be projected onto each principal component, and a set of score values are associated with each principal component. In addition, the “loadings” can be determined which, geometrically, represent the direction of each principal component in the  $N$  dimensional feature space. These loadings represent how each of the original variables contribute towards each principal component.

### 3.4.2 Multivariate Linear Regression

Multivariate Linear Regression (MLR) was used in Chapter 4 to produce predictive models for  $E_0$ ,  $T_{80}$  and  $T_{S80}$  and determine the most significant factors governing these responses based on OPV architectures and test conditions. MLR was additionally used in Chapter 5 to determine the degradation rates of OPV modules tested in Outdoor conditions.

An MLR model uses a least square polynomial model to express the effects in terms of the stress factors or environmental conditions. Least square mean values can be calculated by assuming the model [142],

$$y_i = f_0 + \sum_{k=1}^z f_k x_{nk} + \varepsilon_n, \quad 3.4$$

where  $y_i$  represents the response,  $f_0$  is a constant fitted parameter,  $f_1$  to  $f_k$  are the fitting parameters for the individual coefficients,  $x_{nk}$  represents the  $n^{\text{th}}$  level of the  $k^{\text{th}}$  predictor variable and  $\varepsilon_n$  represents the standard variance error. Multiple ‘y’ observations can be expressed with the matrices,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdot & x_{1n} \\ 1 & x_{21} & x_{22} & \cdot & x_{2n} \\ \vdots & \vdots & \cdot & \cdot & \cdot \\ 1 & x_{n1} & x_{n2} & \cdot & x_{nn} \end{bmatrix}. \quad 3.5$$

The matrix  $\mathbf{X}$  is denoted as the design matrix containing information about the level of the predictor variable. In order to identify a regression line, the standard error must be accounted for. The standard error of the estimate is a measure of the accuracy of predictions and given by,

$$\sigma_{est} = \sqrt{\frac{\sum (Y_{real} - Y_{est})^2}{N}}, \quad 3.6$$

where  $\sigma_{est}$  is the standard error of the estimate,  $Y_{real}$  is an actual value,  $Y_{est}$  is an estimated value and  $N$  is the number of samples. The numerator is the sum of the squared differences between the reference values and the predicted values. The method of least squares is employed in MLR, whereby the coefficients in the matrix, shown in equation 3.5, can be estimated by finding values of the coefficients that minimise the sum of the squared residuals.

In order to implement a MLR model, several criteria must be fulfilled:

1. For all instances of the independent variable, the error in the dependent variable, plotted as a function of the independent variable, should be normally distributed with constant variance (normality).
2. As the dependent variable scales linearly, the independent variable should scale linearly as well (linearity).
3. All observations should be independent of one another; the measurement of one sample does not influence the measurement of another sample (independence).
4. The error term (the random disturbance in the relationship between the independent and dependent variables) is constant across all values of the independent variable (homoscedasticity) [143].



The efficacy of these criteria can be tested through residual analysis. If the criteria are fulfilled and the model is valid, then the residuals should be normally distributed with a mean of zero and a constant standard deviation.

Applying MLR analysis to a data set allows for the determination of the most significant factors affecting the performance and stability of OPV devices and find the best combination of factors in order to improve the performance and stability. The t-value can be utilised as a metric assessing the significance of each factor. The t-value is defined as being [142],

$$t = \frac{\bar{x} - \mu}{SE}, \quad 3.7$$

where  $\bar{x}$  is the mean of the sample,  $\mu$  is the mean of the population, and  $SE$  is the standard error. The higher the t-value, the more significant the factor is in the prediction of the related response.

### 3.4.3 Sequential Minimal Optimisation

Sequential minimal optimisation regression (SMOreg) was used in Chapter 4 to produce predictive models for  $E_0$  and  $T_{80}$  and determine the most significant factors governing these responses based on OPV architectures and test conditions. The SMOreg algorithm is a useful machine learning (ML) algorithm for analysing both categorical and numerical data sets by allowing the acquired model to be analysed in terms of the weight of each attribute and, therefore, the significance of each attribute can be understood; this is ideal for the problem of predicting performance and stability of OPVs where we want to find which attributes have the biggest and least impact upon OPV stability. The weights of each attribute are obtained through the method of variational calculus using Lagrange multipliers to determine an optimum hyperplane, which separates the dataset into classes [144]. Initially the method of sequential minimal optimisation for classification will be discussed to introduce the method and, subsequently, the modification of this algorithm for regression will be outlined.

### Sequential Minimal Optimisation for Classification

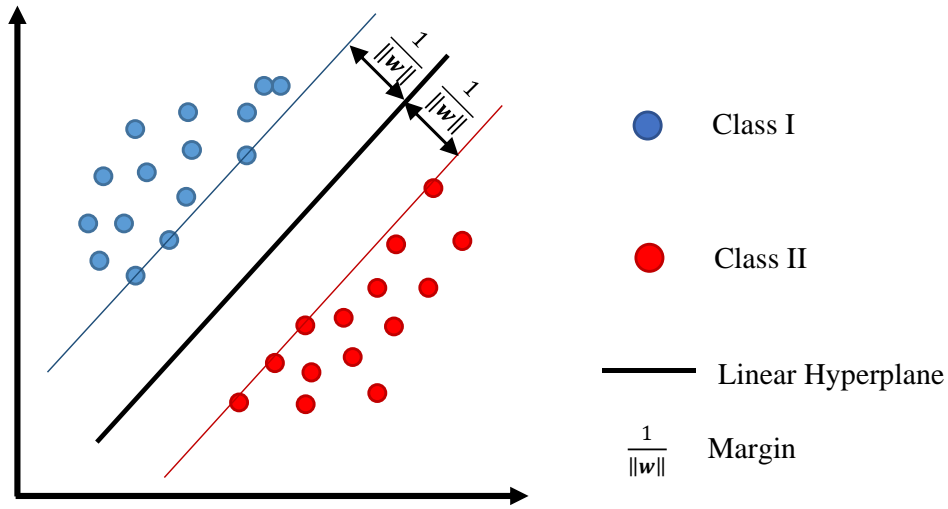


Figure 3.4: Schematic of classification during SMO.

To separate data, a hyperplane is used which can be linear (for linearly separable data) or can be more complex, thus requiring a kernel function to define the form of the hyperplane [144]. Figure 3.4 illustrates the classified data with a linear hyperplane separating the two classes, shown in blue and red. Maximising the margin is equivalent to maximising the term  $1/\|\mathbf{w}\|$  which is equivalent to minimising  $\|\mathbf{w}\|$ . The optimisation of the hyperplane fitting is achieved through variational calculus and Lagrangian methods [144]. The Lagrangian is defined as,

$$\mathcal{L}(\mathbf{w}, \mathbf{b}, \alpha) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^N \alpha_i y_i (\mathbf{w} \cdot \mathbf{x}_i - b) + \sum_{i=1}^N \alpha_i. \quad 3.8$$

For equation 2.1,  $\mathbf{w}$  is the weight vector,  $\alpha_i$  is the Lagrange multiplier,  $\mathbf{x}_i$  is the instance variable,  $y_i$  is the response and  $b$  is an intercept term. The objective problem is to find a hyperplane such that the above Lagrangian is maximised with respect to  $\alpha$  with the constraints that  $\alpha \geq 0$  with,

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \mathbf{0} , \quad 3.9$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 . \quad 3.10$$

This, subsequently, allows  $w$  to be determined from equations 3.9 and 3.10 as,

$$\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i , \quad 3.11$$

and,

$$\sum_{i=1}^N \alpha_i y_i = 0 . \quad 3.12$$

Therefore, via substitution, we can obtain,

$$\mathcal{L}_D = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j . \quad 3.13$$

Throughout the Lagrangian iteration process, the requirement is that the condition  $\sum_{i=1}^l y_i \alpha_i = 0$  is enforced and implies that the smallest number of multipliers that can be optimised at each step is 2; whenever one multiplier is updated, at least one other multiplier needs to be adjusted in order to keep the condition true. At each step, SMOReg chooses two elements  $\alpha_i$  and  $\alpha_j$  to jointly optimise, finds the optimal values for those two parameters given that all the others are fixed, and updates the  $\alpha$  vector accordingly whilst maximising the objective function, represented by equation 3.13 [144].

The distance between the hyperplane and the edge of a group of classed data is known as the margin where the blue and red lines represent two additional hyperplanes mapping the edges of the two classes of data, representing two different attributes. Data which lie on one of the two boundary hyperplanes are referred to as the

support vectors, which govern the form of the separating hyperplane. By definition, no data can exist between the separating hyperplanes. The goal of the SMOreg algorithm is to find a hyperplane which maximises the margin [144].

The effectiveness of the SMO algorithm in characterising the data can be quantified by the correlation coefficient. The correlation coefficient (also known as Pearson Product-Moment correlation coefficient),  $r$ , determined in statistical regression analysis yields information regarding the linear dependence of one variable on the value of another [145]. The correlation coefficient can take values between -1 and 1 with 1 corresponding to a perfect positive correlation, whilst -1 corresponds to a perfect negative correlation. A value of 0 for the correlation coefficient signifies that there is no dependence of one variable on another. The strength of the correlation between two variables is given directly by the magnitude of  $r$ . Generally, a value of  $r$  between 0.5 and 1 is considered to be a strong positive correlation. The value of  $r$  can be calculated using equation 3.14.

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}. \quad 3.14$$

In equation 3.14,  $x_i$  is the reference  $x$  value and  $\bar{x}$  is the mean value of the reference values. Similarly, for  $y$ , the predicted response. In the equation above,  $\sum_i (x_i - \bar{x})(y_i - \bar{y})$  corresponds to the product of the covariance of  $x$  and the covariance of  $y$ . The term  $\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}$  corresponds to the product of the standard deviations of  $x$  and  $y$  respectively. It should be noted that the correlation coefficient can be applied as a statistical measure for any machine learning algorithm with the same procedure.

#### *Sequential Minimal Optimisation for Regression*

The SMO algorithm can also be applied for regression tasks as shown by Vapnik by introducing an  $\varepsilon$ -insensitive loss function ( $\varepsilon$ -SV or SMOreg) [146]. The objective in SMOreg is to find a function,  $f$ , which has a maximum deviation,  $\varepsilon$ , from the actual target variables for all training data. This implies that only errors which are smaller than  $\varepsilon$  are considered important and any larger errors are rejected. Similarly to SMO, the objective in SMOreg is to minimise  $\|\mathbf{w}\|$ . However, it may not always be possible to find a function

within these bounds. Therefore, slack variables  $\xi_i, \xi_i^*$  to handle the otherwise infeasible constraints. This formulation was given by Vapnik as,

$$\text{minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*), \quad 3.15$$

$$\text{subject to } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}.$$

The constant  $C$  dictates the trade-off between the flatness of the function  $f$  and the limit to which deviations larger than  $\varepsilon$  are tolerated. This can be represented by the  $\varepsilon$  - insensitive loss function  $|\xi|_\varepsilon$ ,

$$|\xi|_\varepsilon = \begin{cases} 0 & \text{if } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{otherwise} \end{cases} \quad 3.16$$

The method of applying the SMOreg algorithm using a soft margin loss for a linear SVM can be represented as shown in Figure 3.5 [147].

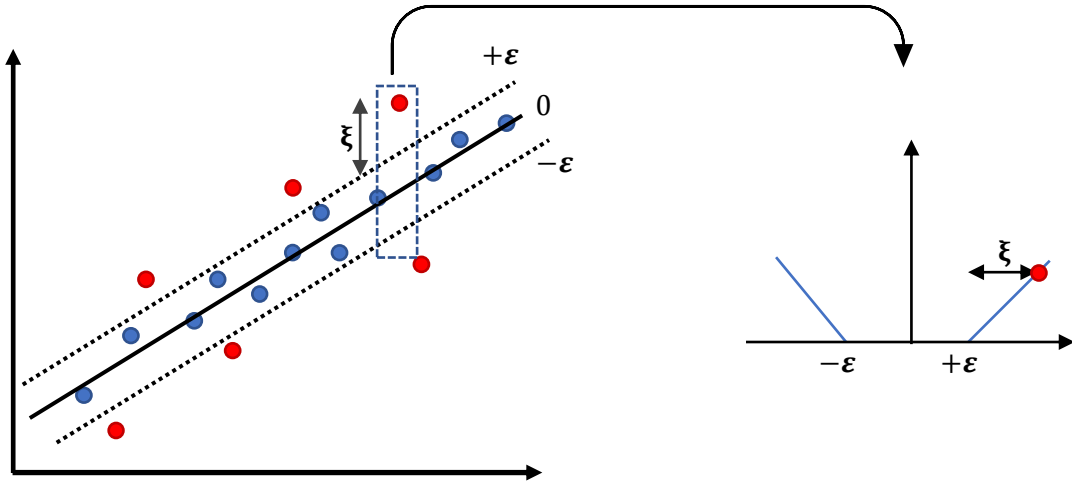


Figure 3.5: Schematic of regression during SMOreg with soft margin loss setting.

### 3.4.4 Multilayer Perceptron

The multilayer perceptron (MLP) algorithm was used in Chapter 5 to predict the diurnal variation in OPV performance parameters tested outdoor, based on the diurnal weather conditions. The MLP is a feedforward artificial neural network proposed by Rosenblatt in 1958 [148]. The network relies on 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 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 [149],

$$f(x) = \frac{1}{1 + e^{-x}}. \quad 3.16$$

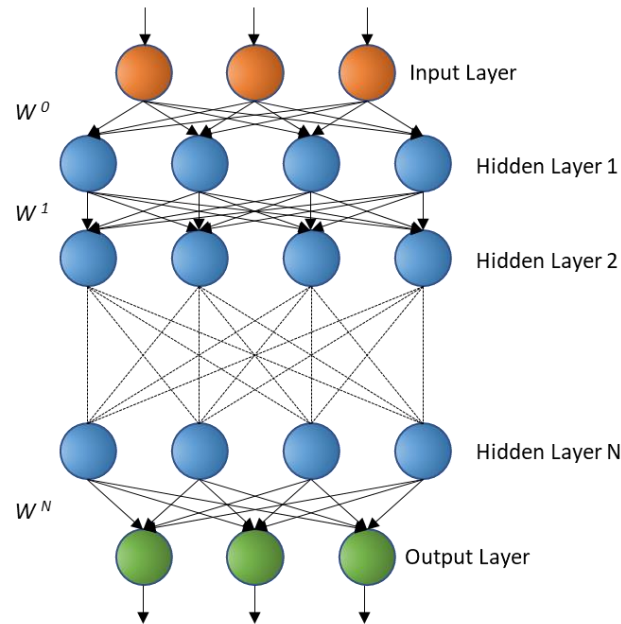
Therefore, the output from the  $i^{\text{th}}$  hidden layer, can be represented by

$$h_i^j = f\left(\sum_{k=1}^{n_{i-1}} w_{k,j}^{i-1} h_{i-1}^k\right); i = 2, \dots, N \text{ and } j = 1, \dots, n_i, \quad 3.17$$

where  $w_{k,j}^i$  is the weight of the interconnection between node  $k$  in hidden layer  $i$  and node  $j$  in hidden layer  $i + 1$ .  $n_i$  is the number of neurons in the  $i^{\text{th}}$  hidden layer [149]. The final output from the entire neural network can be represented by,

$$y_i = f\left(\sum_{k=1}^{n_N} w_{k,j}^N h_N^k\right); Y = (y_1, \dots, y_{N+1}) = F(W, X), \quad 3.18$$

where  $w_{k,j}^N$  is the weight between node  $k$ , in the  $N^{\text{th}}$  hidden layer, and node  $j$ , in the output layer.  $n_N$  is the number of nodes in the  $N^{\text{th}}$  hidden layer.  $Y$  is the vector of the output layer,  $F$  is the transfer function and  $W$  is the matrix of weights for all hidden layers and  $W^i$  represents the vector of weight connecting each layer of nodes. A schematic illustration of the MLP neural network is shown in Figure 3.6.



*Figure 3.6: Schematic illustration of multilayer perceptron artificial neural network showing input layer, hidden layer and output layer, all interconnected by weights.*

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. Learning for the MLP is the process of adapting the connection weights in order to minimise the difference between the network output and the desired output. This is achieved using backpropagation where the weights are recalculated for the last hidden layer and subsequently recalculated for each previous layer from the last layer to the input layer with each recalculation reducing the error term. In the case of forecasting, the MLP model is first trained *via* supervised learning on a certain set of training data. This is achieved by training the MLP network to predict the solar cell performance parameters based on the weather conditions. In the final analysis this is performed using two weeks of weather conditions and associated performance parameters as the training set. The trained MLP network is then applied to unseen data consisting of only weather conditions and the corresponding performance parameters are predicted by the network. The unseen weather conditions correspond to data in the future and represents the testing set. The testing data used spanned a period of approximately 6 months after the 2-week training set.

### 3.4.5 Random Forest

The Random Forest algorithm was used in Chapter 5 to predict the degradation of OPV performance parameters tested outdoor weather conditions. The Random Forest (RF) algorithm is employed as an algorithm for time series forecasting [150]. The RF algorithm can be considered as an ensemble of individual decision trees [151]. A decision tree operates by applying a cascade of criteria to each feature of the dataset and identifying which of these criteria are fulfilled. For categorical data, the decision tree splits the data into nodes based on the predictor variables. To achieve this, several possible split points are identified for every predictor variable. The algorithm then determines the purity of the splits in terms of the classes present in each node. If classification were applied to data consisting of OPV architecture, the decision tree would split the data to achieve pure classification of each material. For example, the substrate would be classified as being glass, PET or quartz. If a node contains multiple classes, the node is split further until optimum classification is attained. For regression, a decision tree splits the data at certain percentiles of the distribution of values. For each split, the decision tree calculates the mean square error and selects the result with the lowest mean square error value. The mean square error is calculated in relation to the known output value during supervised learning. The decision tree comprises of several components, namely decision nodes and prediction nodes. Dependent on the criteria which have been identified for each feature, the decision tree applies a class label, based on the fulfilled criteria. In the case of performance prediction and forecasting, a continuous variable decision tree is employed where the decision tree has a continuous target variable [152] and the performance parameter output is predicted based on continuous variable inputs. Finally, the RF collects all the outputs from each individual tree reaches a collective prediction based on the majority vote from each decision tree. In the case of continuous variables, the output value is determined through averaging of the outputs from each tree [153].

The goal of the RF algorithm allows for averaging of multiple deep decision trees such that overfitting of the data is avoided; deep decision trees tend to become highly atuned to irregular patterns in datasets and, consequently, have very low bias, but high variance [154]. Bias in ML algorithms relates to the tendency of an algorithm to incorrectly favour certain attributes of the dataset, due to limited flexibiity, sometimes



due to a constraint on the number of possible parameters in the model. Consequently, the algorithm may not identify the significant relationships and patterns within the dataset. Parametric algorithms commonly experience high bias since the set of available parameters is constrained in breadth. Examples of ML algorithms which experience high bias are linear regression, linear discriminant analysis and logistic regression. The averaging process in the RF algorithm is referred to as feature bootstrap aggregation bagging [155][156]. Initially the dataset is split into a training set and a testing set, where the RF algorithm is first trained and a predictive model developed. Feature bagging is a modified version of the bagging process used for decision trees, where for each split in the decision tree learning process, a random subspace of the features is selected [157]. This reduces the likelihood of a limited number of strong predictors becoming selected by the majority of the decision trees, leading to correlation of these features. The process of feature bagging is summarised below.

Considering a training set consisting of  $X = x_1, x_2, \dots, x_n$  attributes and  $Y = y_1, y_2, \dots, y_n$  responses, bagging selects  $L$  random samples of  $X$  and applied a decision tree to each sample. For  $l = 1, \dots, L$ ,  $n$  training samples are randomly selected with replacement, meaning that a feature may be repeated more than once in a particular sample. A regression decision tree ( $f_l$ ) is subsequently trained on the random samples. Once the model has been trained, it can be applied to the unseen testing data with samples  $x'$  [155],

$$\hat{f} = \frac{1}{L} \sum_{l=1}^L f_l(x'), \quad 3.19$$

where  $\hat{f}$  is the averaged prediction from all the sampled decision trees. This methodology of bagging RF algorithms allows for predictions to be made with low variance whilst minimising the bias.

### 3.4.6 Hypothesis Testing and Statistical Significance

Hypothesis testing was used in Chapter 5 to determine the relative significance of the weather conditions governing the OPV degradation. In many situations when trying to analyse large multivariate datasets, it is useful to be able to identify which factors, used

in the model, govern the prediction most significantly. This can be performed using statistical techniques and hypothesis testing methods. This is achieved by first defining a null hypothesis ( $H_0$ ) and computing the associated  $p$ -value for each factor in the model. The null hypothesis is defined as being no relationship between the response parameter and each of the factors employed; the results are due to chance and that there is no causality between the factors. In contrast, the alternative hypothesis ( $H_1$ ) asserts that the factor, or independent variable, does govern the response and is, therefore, significant. The level of statistical significance of each factor can be assessed using the  $p$  – value for a two – tailed distribution.

The significance of each factor can be determined from inspection of the acquired  $p$  – value, where the smaller the value, the greater the significance and provides stronger evidence that the null hypothesis should be rejected. Typically, a  $p$  – value of 0.05 or less is considered to be significant, where there is less than a 5% probability that the null hypothesis is correct [142]. In contrast a  $p$  – value greater than 0.05 suggests that the null hypothesis is correct and the alternative hypothesis is rejected [142].

### 3.4.7 Software for Data Analysis and Machine Learning

Multivariate linear regression was implemented using the Unscrambler 7.0 software (for performance and stability assessment, Chapter 4) as well as using Python 3.7 (for OPV lifetime forecasting, Chapter 6). The SMOreg algorithm was implemented using the Waikato Environment for Knowledge Analysis, WEKA 3.8 (for performance and stability assessment, Chapter 4). MLR, MLP, RF, principal component analysis and statistical significance analysis were all implemented using Python 3.7 (for diurnal performance prediction and degradation forecasting, Chapter 6).

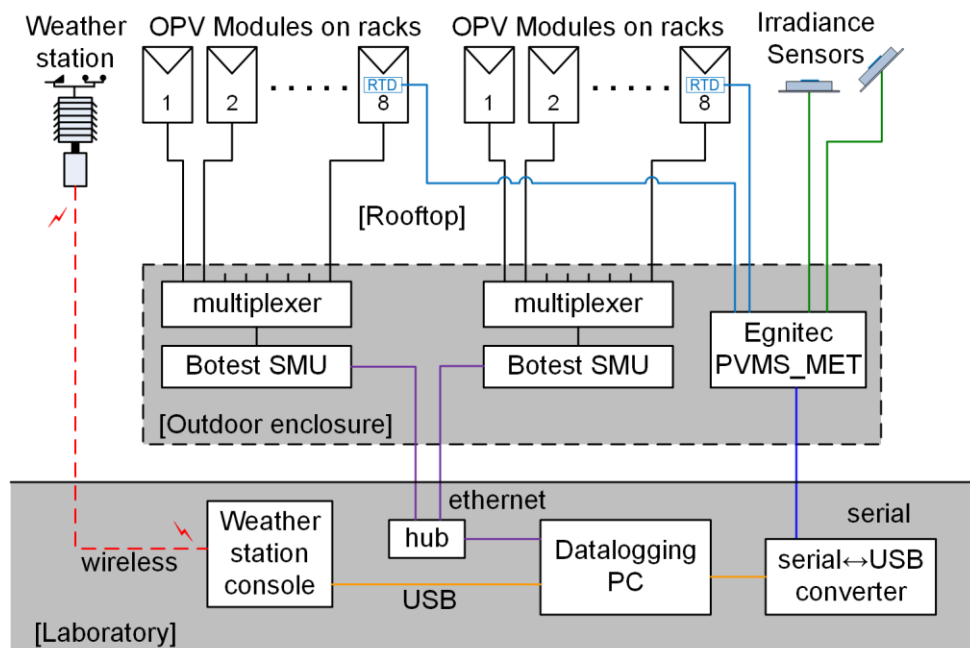
## 3.5 Outdoor Measurement

### 3.5.1 Outdoor Measurement System

For outdoor monitoring data acquisition, an outdoor monitoring system, setup at The School of Electronic Engineering, Bangor University, was used for monitoring module performance, irradiance levels and local weather conditions and is designed such that

tests can be conducted according to ISOS-O-2 protocols. OPV modules are mounted on adjustable racks which allow the inclination angle to be chosen. The OPV modules are connected *via* 8 – channel multiplexers to SMUs, supplied by Botest Systems GmbH, positioned within a weatherproof cabinet. This allows a total of 8 modules to be monitored synchronously and allows for I-V characterisation and performance parameter extraction for the modules under outdoor testing conditions. The irradiance levels are monitored using commercially bought irradiance sensors supplied from IMT Solar. Two silicon irradiance sensors were installed, one mounted horizontally and one mounted in – plane with the OPV modules being tested. The irradiance sensors are monitored using a PVMS-250(MET) measurement system supplied by Egnitec Ltd.

The weather conditions are monitored using a commercially bought Davis Vantage Pro 2 weather station. This system allows the temperature, relative humidity, wind speed, wind direction, solar radiation, UV levels, rainfall, and air pressure to be monitored. A solar powered integrated sensor suite is employed to wirelessly connect with the weather station console indoors which communicates with the primary data logging PC (which also collects the OPV I-V characteristics). The console downloads the data onto the PC at regular intervals. A schematic illustration of the setup is shown in Figure 3.7.



*Figure 3.7: Schematic of the outdoor monitoring system installed on the roof the School of Electronic Engineering, Bangor, North Wales.*

The Botest SMUs and Egnitec measurement units employ independent datalogging software and is run on a single PC, thus maintaining synchronous timestamps between the measurements. A Microsoft Access database was used to import and synchronise the different measurement files. The database uses various import routines to process and synchronise all the datafiles and allows each of the performance parameters (PCE,  $J_{sc}$ ,  $V_{oc}$  and FF) to be extracted, as shown in Figure 3.8. The synchronisation step is necessary since the timestamp corresponding to the irradiance level must be synchronous with the I-V datafile timestamp, in order to calculate the PCE. The synchronisation results in each of the modules measurements being associated with the nearest irradiance and weather station record.

Figure 3.8 consists of two screenshots of a Microsoft Access database interface. Screenshot (a) shows the 'Import Folder' section with a text box for the folder path and a 'Processing Status' area. Below this, there are checkboxes for 'AutoLoad' and 'Egnitec MPP\_IV', 'Egnitec MET\_IV', 'Egnitec MET\_Irr', 'Egnitec OPV\_Temp', 'BOTest', and 'Weather Station'. There are also fields for 'File Type', 'File Count', 'Record Count', and 'Import File'. Screenshot (b) shows the 'Query Automation' section with radio buttons for 'Display AND Copy', 'Display Query (NO Copy)', and 'Copy To Clipboard (No Display)'. Below this, there are 'Date-based Queries' and 'Crosstabs' sections. The 'Date-based Queries' section includes fields for 'DATE', 'Start Time', 'End Time', 'OPV Data', 'OPV Module', 'Weather (Month)', 'Weather (Day)', 'IRR Data', 'PV Data', and 'PV Module'. The 'Crosstabs' section displays a grid of buttons for various data combinations, such as 'OPV\_Efficiency\_Diurnal', 'OPV\_Efficiency', 'PV\_Efficiency\_Diurnal', 'OPV\_Irr\_vs\_Isc', 'OPV\_Irr\_vs\_Voc', 'OPV\_Irr\_vs\_FF', 'OPV\_Voc\_Diurnal', 'OPV\_Irr\_Diurnal', 'PV\_Efficiency', 'OPV\_Irr\_vs\_Voc/Cell', 'OPV\_Voc/Cell\_Diurnal', 'OPV\_Power\_Diurnal', 'PV\_Power', 'OPV\_Irr\_vs\_FF', 'OPV\_FF\_Diurnal', 'OPV\_Irr\_vs\_Eff', 'OPV vs Eff (at Irr)', 'OPV vs Isc (at Irr)', 'OPV vs Voc (at Irr)', 'OPV vs FF (at Irr)', 'OPV vs MPP (at Irr)', 'OPV Daily Yield', 'OPV Daily Yield (timed)', 'OPV Daily Yield per Area', 'OPV Daily Yield/Area (timed)', 'Daily Insolation', 'Sampling Ratios', 'Extract Irradiances', and 'OPV and WS Data'.

Figure 3.8: Microsoft Access databases used to (a) import and synchronise the different measurement files and (b) select different crosstabs.

## Chapter 4. Enhancing the Indoor Performance and Stability of Emerging Photovoltaics through Machine Learning

### 4.1 Introduction

The nature of OPV research is such that a vast array of results has been gathered, leading to data being generated with a plethora of materials and device structures. The body of work has led to substantial improvements in efficiency and stability, even if these still remain below those of silicon and other mature thin film technologies. As demonstrated by the literature review previous studies have made significant progress in improving OPV performance and stability. However, it is often difficult to quantify the key factors governing OPV performance and rank them in terms of their significance. In addition, directly comparing results across the breadth of literature can be problematic and does not provide an analysis of OPV performance in totality. By adopting a holistic approach, and applying machine learning techniques to analyse the literature, greater insight can be achieved.

As discussed, one of the key metrics for a solar cell technology's commercial potential is stability, which has been explored by utilising machine learning (ML) techniques on a large dataset of OPV data. This allows interesting trends to be extracted from the underlying patterns within the dataset that goes beyond the standard approach of acquiring specific information by directly measuring the stability due to one or more changes. By deploying data analytical approaches in this manner, it is possible to determine which materials and stress factors have the greatest impact on the device stability.

Developing ML models for predicting efficiency and stability of OPV devices based on large datasets comes with a number of challenges. One of those are inconsistencies that arise from the fact that data comes from multiple authors working in different labs which introduces biases and random noise in the data. Others are inconsistencies in the set of materials that result in optimum stability, variability in material properties, and non-uniformly employed testing protocols, with the result that the information

concerning the relative impact of different layers upon the stability may be limited. A final challenge is that factors that have a significant impact on device performance are not adequately described or reported, or not yet properly known, and therefore it will not be possible to directly incorporate in the ML-models. One of the important questions that arises, and which the performance of the ML-models can answer, is to what extent we can understand and report the factors that really are important for device performance, and how much is still “dark knowledge” or just unknown.

In this chapter, OPV stability data, based on ISOS testing protocols, has been utilised which has been acquired from work undertaken between 2011-2019. Using ML methods, described in Chapter 3, trends within this OPV dataset can be elucidated and models produced that enable OPV stability and performance to be estimated based on the device architecture and environmental testing conditions. This methodology could be used to compare stability studies more fully and to analyse the relative significance of various environmental stresses and materials and, ultimately, identify key failure mechanisms in devices. An analysis of this type would allow the challenges experienced in assessing the literature to be overcome by employing analytical models to identify the trends and patterns in the development of OPV technologies and rank the most influential factors governing the performance and stability.

## 4.2 Methodology

### 4.2.1 Overview of Data Acquisition and Ethics

The dataset used for this work was partly obtained from the Danish Technical University (DTU) who ran the “lifetime predictor” on the Plastic photovoltaics website from 2011-2017 [156]. Additional papers between 2017 and 2019 were manually scraped, with an estimated 65%/35% split between the two data sources. To source papers, ‘Web of Science’ was used to identify papers with OPV stability data. Device information and device data was subsequently extracted from the papers and added to the dataset. Information concerning device architecture and testing conditions was also determined. Device data was found in terms of performance and stability, which was extracted from tables or in figures; in which case a plot digitizer was employed in order to extract the relevant information from the figures. This allows for parameters such as the initial

efficiency ( $E_0$ ), time taken for the device to reach 80% of the initial value ( $T_{80}$ ) and time taken for the device to reach 80% of the stabilised value ( $T_{S80}$ ) to be extracted. Whilst automated scraping using Python or R is possible, all the analysed data have been manually extracted from the papers. In total, data from 1850 devices were used for this work. Papers were selected where device stability data is reported.

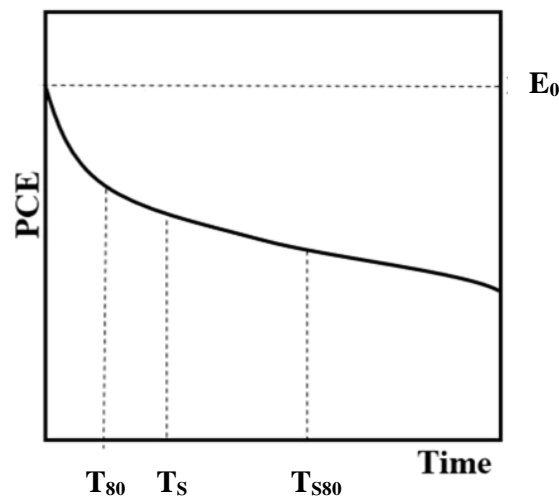
#### 4.2.2 Data Acquisition and Description of Data Format and Categories

For the analysis methodology used in this chapter, consistency in the format is essential and data has therefore been modified using the ‘OSEMN’ (Obtain, Scrub, Explore, Model, Interpret) process [158]. The “obtain and scrub” phase entails the acquisition and formatting of the data in a format suitable for the application of ML algorithms, such as a CSV file with consistency of format enforced throughout the dataset. In the case of the material/structural properties, a schematic of the generic structure of the OPV devices is as shown in Figure 2.4. Note that the active layer can comprise of both a single layer and a blend. This is represented by using two active layers in the dataset, represented by Active1 and Active2.

Initially, the appropriate attributes are acquired from each paper. The attributes for each device include the structure and materials, encapsulation and substrate type, test protocols, environmental conditions, light source. In excess of the device related attributes, there are also attributes related to measurement conditions such as temperature, light level, bias conditions and relative humidity. In total, the dataset contains 17 attributes, each of which contains a number of categories. These are detailed in Appendix A. Given the number of attributes and the associated categories, the total dimensionality, i.e. the number of possible combinations of labelled device attributes, are 28725. That is far more than the 1850 combinations realised in the dataset, and even if many of those will be experimentally inaccessible, this illustrates the potential for extrapolation and suggestions for new promising combinations worth exploring experimentally that can come out of an analysis of this dataset.

In addition to the categorical features describing the devices, three lifetime and performance metrics have been extracted, namely  $E_0$ ,  $T_{80}$  and  $T_{S80}$ , with  $T_{80}$  being focused on primarily since, for OPVs,  $T_{80}$  is commonly used to assess the lifetime of these devices

(see Figure 4.1). Figure 4.1 provides a schematic illustration of the typical variation in OPV performance with time during degradation and highlights the lifetime metrics considered in this chapter. It should be noted that this degradation curve could vary when different structures and materials are used.  $T_{50}$  might be a more meaningful metric to use, but many papers do not report this or show graphs that cut off before this metric is reached, meaning that extrapolation is required, which can be subjective or misleading. In order to ensure consistency of reported data, all stability measurements correspond to tests performed under ISOS protocols.



*Figure 4.1: Schematic illustration of typical OPV degradation curve, illustrating key lifetime metrics employed in dataset.*

Once the datasets have been obtained, the ‘scrub’ phase can be implemented whereby the data is cleansed and filtered, thus producing data which possess a consistent format. Therefore, careful inspection of the data is required in order to ensure that no unexpected values are present which could significantly affect the final result. For example, if the same materials are reported with slightly different names, or trade names, then these two quantities will be classified as two different features. At the scrub stage, the researcher needs to use their judgement and decide what level of detail they wish to investigate using ML. For example, this approach could easily be used to allow groups to compare the stability of the same material from two different suppliers or identify the optimum thickness of a layer.



In addition to the data having an inconsistent format, some data may be missing or erroneous. There are several reasons as to why the dataset may contain missing values or contain errors made during the construction of the dataset. However, there could be a valid reason for having an empty cell in the dataset. For this work, a process called “imputation” was adopted which can be used to infer a missing value from the nature of the other attributes of the dataset [159]. There are various techniques to deal with missing data and values in machine learning such as using mean values [160], k nearest neighbours [161] and multivariate chained equation [162]. In this study, the data was cleansed and manual deductive imputation [163] was performed. The analysis also used the machine learning algorithm (SMOreg), whose implementation in WEKA of Alex J. Smola and Bernhard Scholkopf’s sequential minimal optimization algorithm globally replaces all missing values. For example, considering the encapsulation or transport layer, a blank or missing entry could signify that this element was not present in the device. Therefore, inferring a categorical value of ‘none’ would be a suitable imputation for this quantity. In the case that the quantity cannot be inferred, due to lack of information concerning the device, the quantity “unknown” was imputed.

Subsequently, after the dataset has been formatted correctly, exploration and modelling can be implemented, whereby patterns and trends in the data can be investigated. The best ML algorithm can be chosen and applied for data analysis (discussed in section 2.3). An overview of the process adopted is shown in Figure 4.2.

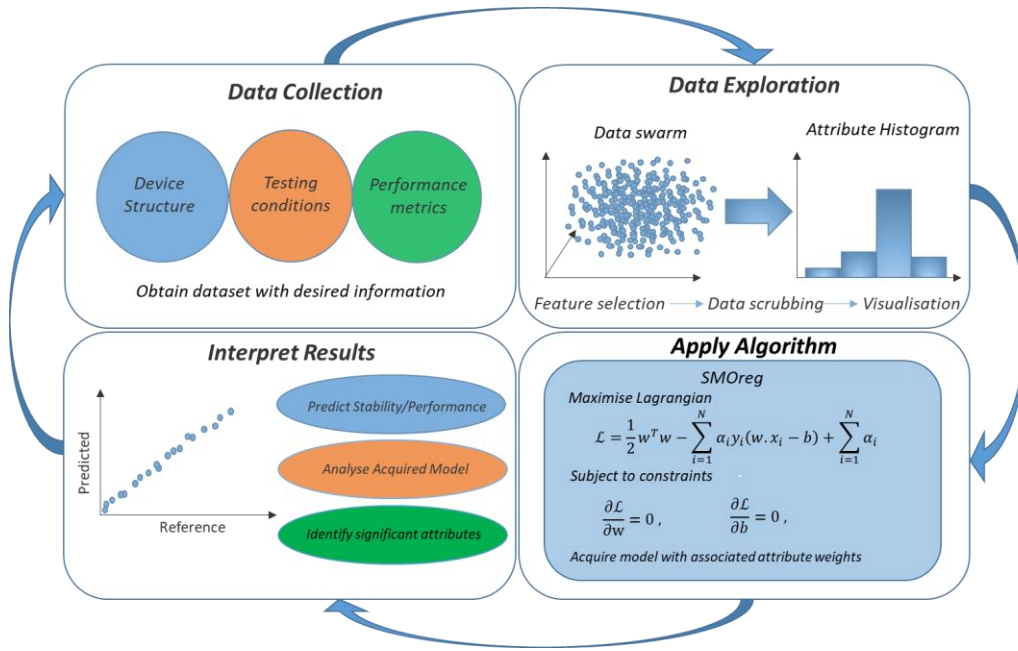


Figure 4.2: Schematic illustration of data science life cycle.

## 4.3 Results and Discussion

### 4.3.1 Initial Data Exploration

Figure 4.3 shows the categorical violin plot for the full distribution of  $T_{80}$ ,  $T_{S80}$  and  $E_0$ . These plots intuitively show the median, interquartile range and  $1.5 \times$  interquartile range. In addition, the plot shows the kernel density estimation to illustrate the distribution of each parameter. Analysis of this data allows the statistics of each of these variables to be determined as shown in Table 4.1. This demonstrates that the  $T_{80}$  and  $T_{S80}$  distributions are very similar with the near uniform distribution up to 10 days of stability, whilst  $E_0$  displays a broad distribution for efficiencies of 3.5%. In addition, the interquartile range of the  $E_0$  values is significantly less than for both  $T_{80}$  and  $T_{S80}$ . The plot shown in Figure 4.3 demonstrates that most stability data is very poor with the greatest proportion of  $T_{80}$  and  $T_{S80}$  lifetime being less than 10 days.

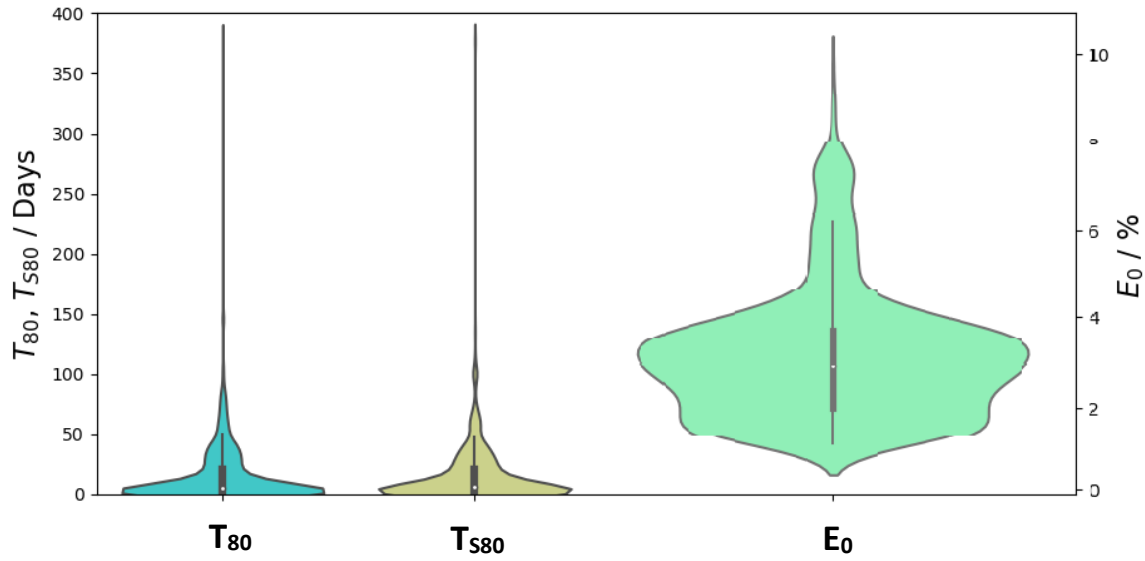


Figure 4.3: Categorical distribution of  $T_{80}$ ,  $T_{s80}$  and  $E_0$  for the full dataset.

	$T_{80}$	$T_{s80}$	$E_0$
Mean / Days, %	14.9	15.0	3.1
Median / Days, %	4.6	5.7	2.9
Max / Days, %	379.2	379.2	9.75
Range / Days, %	378.8	379.0	8.65

Table 4.1: Statistics for  $T_{80}$ ,  $T_{s80}$  and  $E_0$  distributions.

#### 4.3.2 Application of the Multivariate Linear Regression for understanding stability data

In the first instance, a Multivariate Linear Regression (MLR) method for data analysis was used to find trends in the OPV stability dataset. This is performed to predict  $E_0$ ,  $T_{80}$  and  $T_{s80}$  and subsequently determine the attributes which most significantly govern each of the performance parameters. The attribute classes for the dataset are shown in Table 4.2.

Attributes	Categories (Low to High values)
Configuration	Normal (L)   Inverted   Tandem (H)
Encapsulation	N/A   None   Flexible   Rigid
ISOS Compatibility	L,D-1   L,D-2   D-3
Substrate	Glass   PET
First Electrode	ITO   Ag   Other Metal
First Transport Layer (TL1)	None   PEDOT:PSS   Low $\Phi$ Metal Oxide   High $\Phi$ Metal Oxide
Active Layer	Polymer   Small Molecule
Second Transport Layer (TL2)	Inorganic   Small Molecule   None
Second Electrode	Al   Ag   Au
Publication Year	2019 - 2009
Light Spectrum	N/A   Metal halide lamp   Sun   LED   Halogen lamp

*Table 4.2: Attributes and Categories used in the Dataset.*

Numerical labels are applied to each of the attribute classes such that the dataset can be processed using the Unscrambler X software. For example, considering “configuration”, “Normal”, “Inverted” and “Tandem” are labelled as “0”, “1”, and “2” respectively. This will inherently impose some bias in the MLR analysis since the numerically high and low attributes are imposed. However, the validity of the choice for “high” or “low” attributes can be examined from the fitting parameters; a good  $R^2$  value can provide confidence in the choice as well as confidence in the model fitting.

The MLR technique has been applied to the accelerated light-degradation studies (ISOS-L) and accelerated temperature/humidity studies (ISOS-D) separately. Figure 4.4 shows the regression results obtained from papers restricted to light-degradation studies only, showing how the attributes affect the  $E_0$ ,  $T_{80}$ , and  $T_{s80}$  values. In Figure 4.4 (a), the results of the predicted versus reference for the  $E_0$  calculation are shown calculated from the ISOS-L data. The correlation between the predicted and actual values can be represented by the Pearson correlation ( $R^2$  value). However, Figure 4.4 (a) shows an  $R^2$  value of 0.427 for the prediction of  $E_0$  using MLR.

Using MLR analysis, the regression t-values can be calculated from the individual regression coefficients of each attribute, divided by the standard error. A positive t-value means that as the attribute varies from “low” to “high”,  $T_{80}$ ,  $T_{S80}$  or  $E_0$  also increase, and as the attribute varies from “high” to “low”,  $T_{80}$ ,  $T_{S80}$  or  $E_0$  decrease. Therefore, a positive t-value indicates that the higher level of the attribute increases stability, and the lower level reduces stability (with respect to the low and high values in Table 4.2). A negative t-value correspondingly means that as the attribute varies from “low” to “high”,  $T_{80}$ ,  $T_{S80}$  or  $E_0$  decrease, and as the attribute varies from “high” to “low”,  $T_{80}$ ,  $T_{S80}$  or  $E_0$  increases. Therefore, a negative t-value indicates that the higher level of the attribute reduces stability, and the lower level increases stability. Figure 4.4 (a) shows the t-values for  $E_0$  obtained from the ISOS-L. The regression t-value distribution indicates that the encapsulation and light spectrum are the attributes, which have the most significant negative impact on the  $E_0$  prediction, while the most positive for improving the  $E_0$  prediction is the second transport layer (TL2).

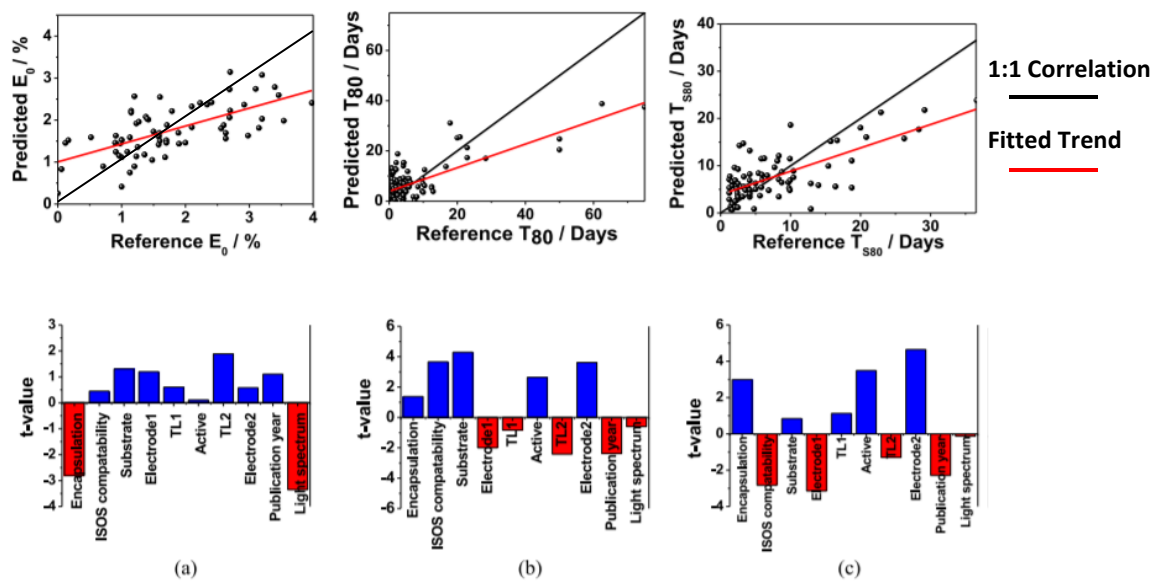


Figure 4.4: MLR prediction versus reference values and model t-values for (a)  $E_0$ , (b)  $T_{80}$  and (c)  $T_{S80}$ , tested under ISOS-L standards.

Figure 4.4 (b) shows the regression results for  $T_{80}$  values obtained only with ISOS-L data with the corresponding regression t-values shown in Figure 4.4 (b). Figure 4.4 (b)

displays a good  $R^2$  value of 0.59 for the prediction of  $T_{80}$  using MLR, showing a moderately better fit than the  $E_0$  data. In rank order, this indicates that “Substrate,” “ISOS compatibility,” “Electrode 2,” “Active layer,” and “Encapsulation” types positively affect stability, and that “Transport Layer 2,” “Electrode 1,” and “Publication year” negatively affect stability (in rank order). The negative impact identified for TL2 indicates that as the material varies from inorganic to small molecule to none, then the  $T_{80}$  lifetime decreases.

A contradictory pattern is observed for  $T_{80}$  and  $T_{S80}$  in the t-value graph, where ISOS compatibility shows a positive effect on  $T_{80}$  and a negative effect on  $T_{S80}$ . This may be attributed to the fact that 65% of the dataset possess P3HT:PCBM as the active layer. In P3HT:PCBM based OPVs, an improvement in performance when annealed is due to an increase in carrier mobility and reduction in trap density in the active region [164]. In the case of ISOS-L-2, the devices are aged at a higher controlled temperature of 65/85 °C. Therefore, during the “burn-in” stage of the devices, the trap density within the active layer reduces, leading to higher  $T_{80}$  times. However, temperature in the long term can cause loss of conjugation in polymers and can damage the contacts, which is reflected in the t-value graph for  $T_{S80}$  [165].

The data also indicate that overall, for “Active layer,” small molecules provide better stability than polymers, that inorganic transport layers (i.e., low-workfunction metal oxides) are more stable than other types of transport layers (TL1).

Figure 4.4 (c) shows the predicted versus reference regression t-value analysis for  $T_{S80}$  ISOS-L data, which yield the worst  $R^2$  value of 0.348. Overall, the trends in the  $T_{S80}$  data follow the  $T_{80}$  data, although the ranking does change. For,  $T_{S80}$  the three most important factors are the top electrode, the active layer, and the encapsulation. The increased significance of the active layer for the  $T_{S80}$  prediction could be associated with the fact that  $T_{S80}$  is the time to reach 80% efficiency after the burn-in process. This means that for  $T_{S80}$ , the trapped charge states in the active layer can now participate in charge extraction [165].

Figure 4.5 (a) and (b) displays the regression analysis for samples only subjected to temperature/humidity studies (based on ISOS-D data) for  $T_{80}$  and  $T_{S80}$ , respectively. Figure 4.5 (a) displays a strong correlation in the prediction with an  $R^2$  value of 0.589 for

$T_{80}$  prediction and an  $R^2$  value of 0.517 for the  $T_{S80}$  prediction. From the analysis, the stability of polymer based OPVs is greater under thermal/humidity testing than small molecules.

The next most impactful attributes resulting in a negative t-value are TL1 and the Electrode indicating that using indium tin oxide (ITO) and metal oxide provides the best stability. The regression analysis for  $T_{S80}$  using the dark data displays very similar trends as observed for  $T_{80}$ , which shows the same contributing factors. The effect of burn-in processes can also explain the differences between the ISOS-L and ISOS-D  $T_{80}$  and  $T_{S80}$  t-values. It is noteworthy that there is a change in parity when comparing the ISOS-L and ISOS-D  $T_{80}$  and  $T_{S80}$  t-values. For ISOS-L testing, the use of a small molecule active layer has a significant impact on enhancing the stability, while for ISOS-D testing, using a polymer active layer improves the stability. This indicates that small-molecule-based active layers have better light stability and are better suited to higher irradiance condition.

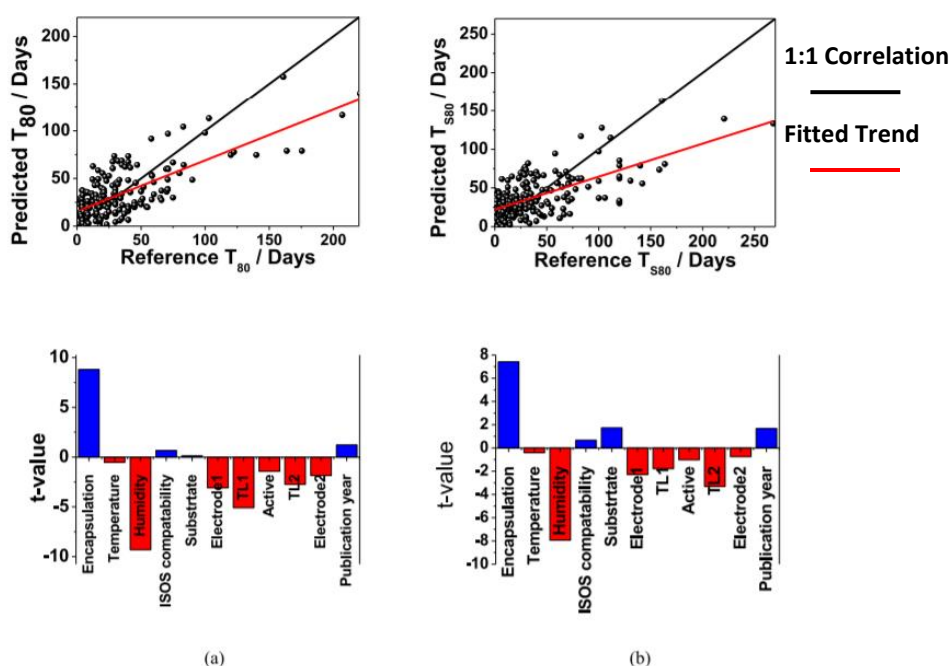


Figure 4.5: MLR prediction versus reference values and model t-values for (a)  $T_{80}$  and (b)  $T_{S80}$ , tested under ISOS-D standards.

### 4.3.3 Algorithm optimisation

In order to optimise the ML algorithm, both the learning method, as well as the dataset, require optimisation. This has been performed in order to determine the training protocols that yield the highest correlation coefficient. This was achieved by training the ML algorithm using percentage split and cross validation. Dataset optimisation was performed using the full set of attribute features, shown in Appendix A, without any further re – classification, thus allowing for greater knowledge extraction whilst maintaining a high model validity.

Percentage Split is a re-sampling method that leaves out a random percentage of data in the training model for testing. This allows for us to assess the accuracy of the trained data based upon the test data. Cross-validation is a resampling procedure for training and testing an ML algorithm. Initially the dataset is reorganised and subsequently split into new groups, called “folds”. For each fold in turn, it is taken as a test set and the remaining folds as the training set. This is then repeated for each fold in turn. The model performance is then summarised for all iterations. In this way each entry in the dataset is assigned to one group, where it remains for the entire iteration.

The performance parameters are determined as a function of the proportion of the dataset used for training and as a function of the number of folds used for cross validation. For the dataset optimisation, this was achieved by varying the minimum value of each response parameter in each case.

When the SMOreg algorithm is applied to the entire dataset, this results in a predictive model which gives a correlation coefficient (CC) of 0.46 and relative absolute error (RAE) of 54%. It is required that any model that is developed to predict the stability, in terms of  $T_{80}$ ,  $T_{S80}$  and the performance, in terms of  $E_0$ , includes only stable devices in the investigation. Therefore, in order to obtain the optimal model predicting  $T_{80}$ ,  $T_{S80}$  and  $E_0$ , the CC is initially maximised by sequentially removing lower  $T_{80}$  and  $T_{S80}$  instances. This process reduced misclassification of the data when high stability and performance devices are being analysed and will allow the properties of these devices to be more readily determined. Therefore, a minimum  $T_{80}$  of 0.4 days is selected since this yields the greatest initial CC of 0.54. For  $T_{S80}$  the maximum CC is achieved when  $T_{S80}$  data less than 0.2 is removed, giving a maximum CC of 0.48. The same procedure of optimising the



initial CC is performed for the  $E_0$  prediction. The variation in CC as a function of minimum  $E_0$  demonstrates that there is a significant increase when  $E_0$  values of greater than 1.1% are used, with very little increase with higher cut-off ranges. Therefore, optimisation of the  $E_0$  prediction for all the data is performed within this range. The variation in minimum CC for  $T_{80}$  and  $E_0$  are shown in Figure 4.6. The optimisation of CC for  $T_{80}$ ,  $T_{S80}$  and  $E_0$  was performed using cross-validation, thus accounting for lower CC values than quoted later in sections 4.3.4 and 4.3.5.

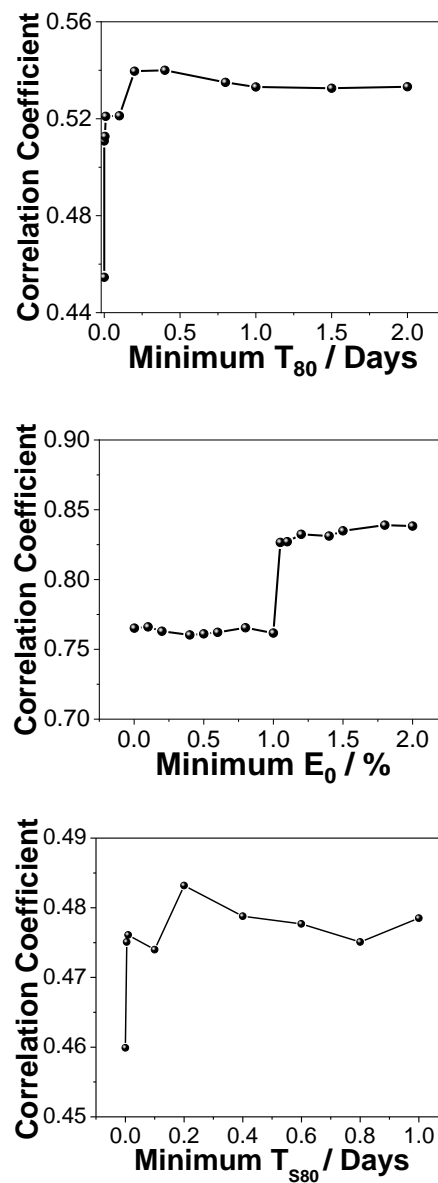


Figure 4.6: Optimisation of dataset by varying the minimum  $T_{80}$ ,  $E_0$ , and  $T_{S80}$  occurrence in the dataset. This optimisation was performed using cross-validation.

Low performing data has been removed such that noise within the dataset is minimised and, therefore, the algorithm does not overfit and make predictions based on this noise. In addition, by removing low performance devices, the algorithm will focus primarily on identifying the materials which lead to improved performance and stability.

The method by which the algorithm is applied to the dataset can significantly alter the predicted results. For example, training can be performed via cross validation, or percentage split. However, the choice of training procedure does not alter the acquired model used for predicting the target variable. Percentage split corresponds to splitting the data into different proportions and training on the remaining proportion. Therefore, the training methods are compared to illustrate their effectiveness in correctly classifying the data. Figure 4.7 illustrates the variation in correlation coefficient (CC), root mean square error (RMSE) and relative absolute error (RAE) for the prediction of  $T_{80}$  under varying percentage splits of the dataset. Figure 4.8 illustrates the variation in correlation coefficient (CC), root mean square error (RMSE) and relative absolute error (RAE) for the prediction of  $T_{80}$  for varying number of folds used during cross validation.

Figure 4.9 illustrates the variation in CC, RMSE and RAE for the prediction of  $E_0$  under varying percentage splits of the dataset and Figure 4.10 shows the corresponding variation for varying number of folds. In summary, percentage split was used to analyse the data since this provided the highest correlation coefficient with lowest error. The optimised algorithm and dataset were subsequently used to analyse the data. The SMOReg algorithm is considered in subsequent sections due to its benefits over the MLR approach in elucidating more detailed information regarding the beneficial and detrimental materials. In addition, the SMOReg algorithm is able to fit a more robust model, as compared with MLR, using the methods of variational calculus and Lagrange multipliers.

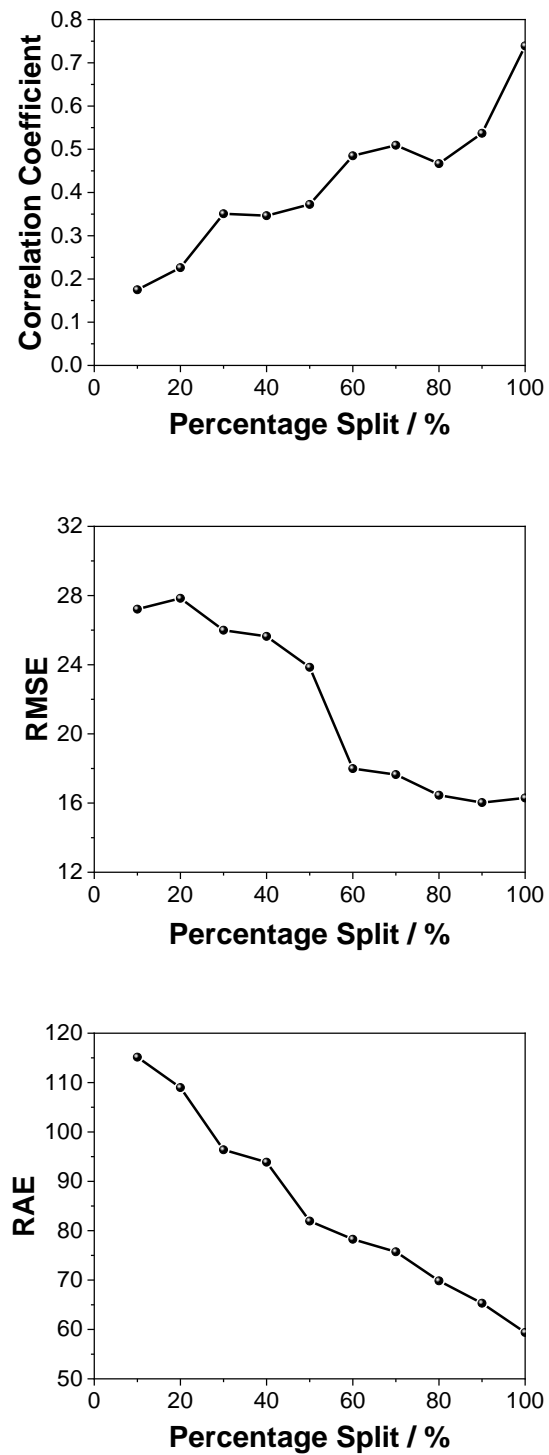
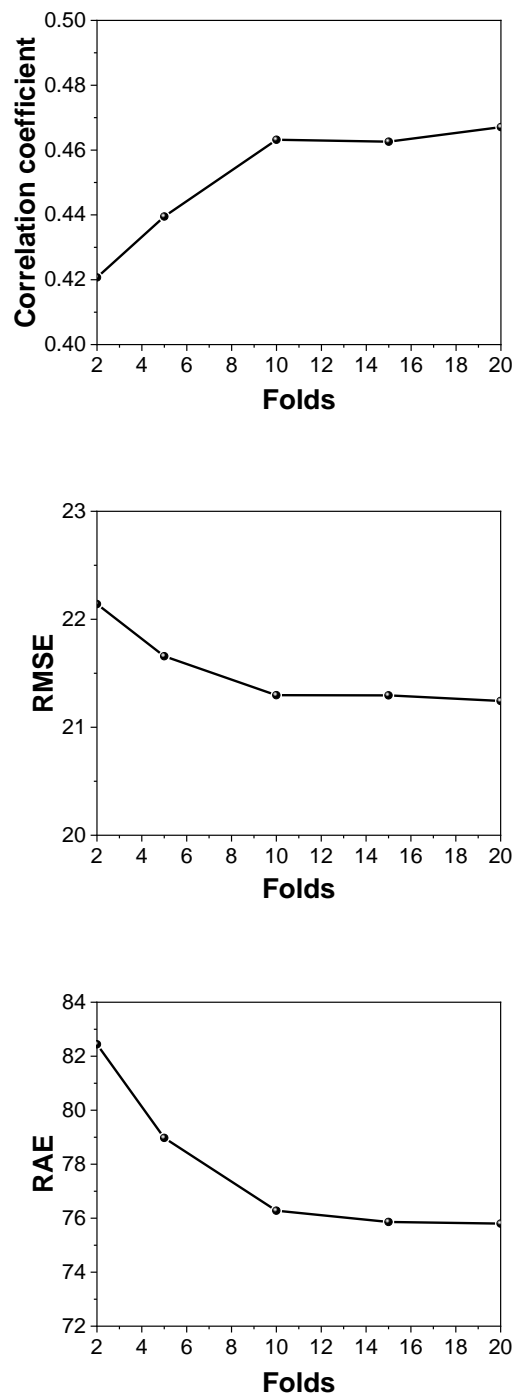


Figure 4.7: Variation in correlation coefficient, root mean square error and relative absolute error for the prediction of  $T_{80}$  under varying percentage splits of the dataset.



*Figure 4.8: Variation in correlation coefficient, root mean square error and relative absolute error for the prediction of  $T_{80}$  under varying number of folds of the dataset.*

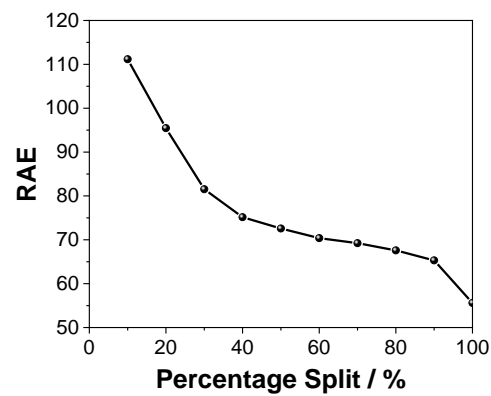
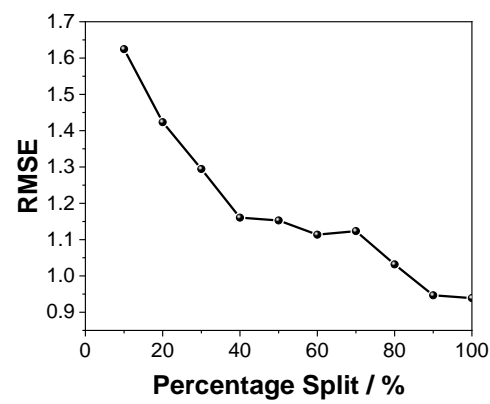
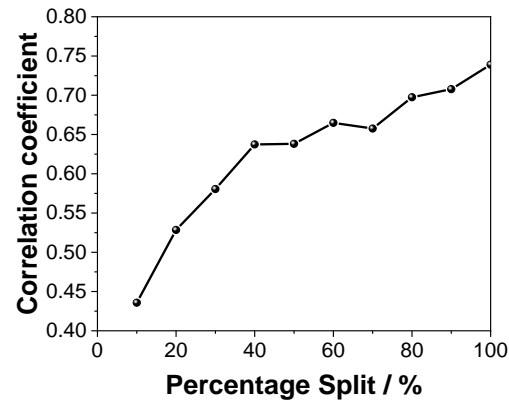


Figure 4.9: Variation in correlation coefficient, root mean square error and relative absolute error for the prediction of  $E_0$  under varying percentage splits of the dataset.

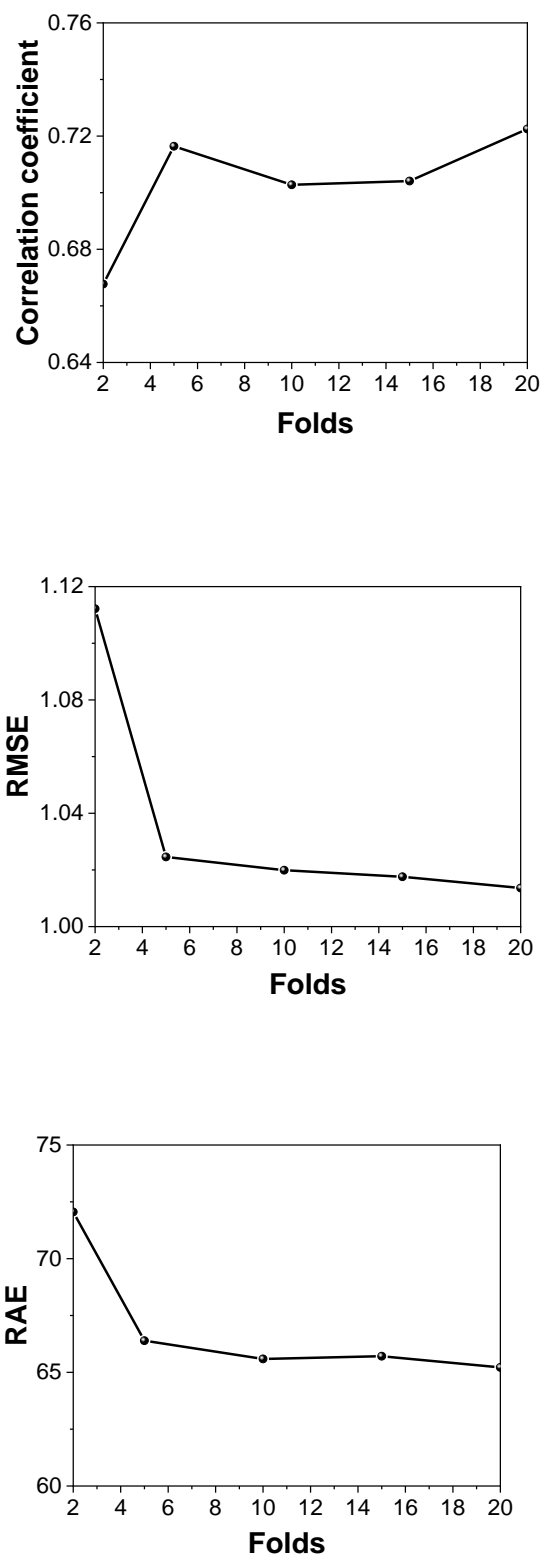


Figure 4.10: Variation in correlation coefficient, root mean square error and relative absolute error for the prediction of  $E_0$  under varying number of folds of the dataset.

#### 4.3.4 Application of the SMOREG algorithm for understanding stability data after dataset optimisation

Subsequently, the SMOREG algorithm has been applied to the optimised full dataset as well as ISOS – L only and ISOS – D only data. The details of the SMOREG algorithm are given in Chapter 3, section 3.4.3. During SMOREG, the regression error is being minimised, as opposed to maximising class distance (as in the case of SMO). The regression model predictions for the three subsets of data are shown in Figure 4.11 along with the residual plots for each model.

The algorithm performance metrics for each analysis are summarised in Table 4.3. The metrics indicate that the correlation coefficient improves when ‘subsets’ of data are analysed (i.e. when data is separated into only ISOS-L data or ISOS-D data), rather than the dataset as a whole. The correlation coefficient can be employed to give a measure of this quality. The overfitting of the algorithm can be verified by comparing the correlation coefficients for training and testing. Verification on unseen data is always important to ensure that the trained model can be generalised to other OPV performance and stability datasets. Overfitting often occurs when a trained model is highly specialised to the training set and cannot make accurate predictions when applied to new datasets which may display different trends and patterns. The correlation coefficient for training and testing can be determined by splitting the data into two subsets (67% for training and 33% for testing). This allows a model to be derived on the larger split of data which is subsequently reapplied to the smaller split. Therefore, the model is being tested on data which it has not encountered before and was not included when training and deriving the model. This will indicate its generalizability to new datasets.

The correlation coefficients acquired for the training and testing sets are given in Table 4.4 and display similar values for both training and testing, giving confidence that the model is not substantially overfitting. In addition, these values compare well with the other literature sources such as Sahu *et al* where values between 0.66 and 0.79 were achieved [130] and outperforms the application of MLR algorithms in section 4.3.2.

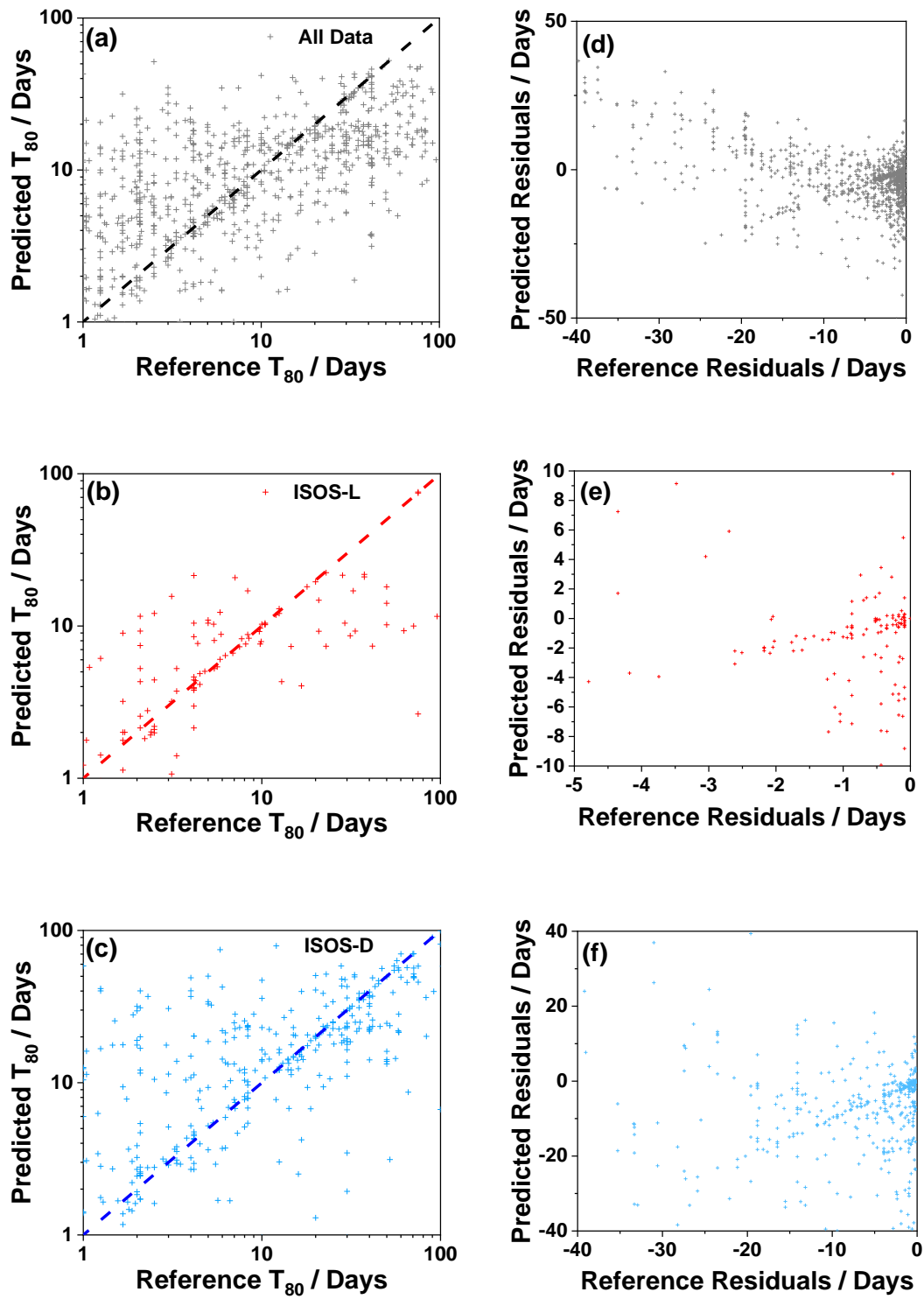


Figure 4.11: Predicted  $T_{80}$  lifetimes plotted as a function of reference  $T_{80}$  lifetimes for (a) the full dataset, (b) ISOS-L testing and (c) ISOS-D testing. The dashed line represents a perfect correlation of 1:1, where the data is plotted on a logarithmic scale. (d)-(f) Residual plots for all data, ISOS – L and ISOS – D respectively.



Metric	All data	Only ISOS-L data	Only ISOS-D data
Correlation coefficient	0.739	0.819	0.767
Mean Absolute Error (days)	9.31	7.64	12.1
Root Mean Squared Error (days)	16.2	20.9	23.5
Relative Absolute Error	59.4%	44.1%	52.9%
Root Relative Squared Error	69.3%	59.7%	65.1%
Number of Instances	1149	155	489

*Table 4.3:  $T_{80}$  fit parameters for SMOreg algorithm applied to full dataset, ISOS-L and ISOS-D, based upon the training set.*

	All data	Only ISOS-L data	Only ISOS-D data
Training	0.739	0.819	0.767
Testing	0.713	0.734	0.659

*Table 4.4: Correlation coefficients for training and testing sets for the SMOreg algorithm.*

Using the weights obtained from the SMOreg model, the 10 most beneficial and 10 most detrimental attributes for each subset of data is shown in Table 4.5. Yellow shading represents testing conditions. Green represents architectural components. Several important features can be identified from the SMOreg weights. For the entire dataset, the attributes that most influence the stability positively are materials used in the first transport layer ('TL1') and the choice of active layer. In addition, the use of an LED light source is found to be beneficial for extending  $T_{80}$ . This is as expected due to the absence of UV light in the LED light sources. Significantly, all of the negative influences correspond to the ISOS testing protocol (and weighted by their relative severity) [166] and the light intensity. What one would expect *a priori* is that harsher test conditions would have a negative impact on stability. This is what is seen and provides confidence in the approach of using the SMOreg algorithm to evaluate how the attributes affect the stability. The

effect of thermal cycling (ISOS-T-3) is found to be the most detrimental attribute, followed by the ISOS-L and then ISOS-D testing protocols. This highlights the importance of not only improving the material stability but also making the devices more robust against harsh environmental testing conditions.

Another factor used in the initial data analysis was the 'year of publication.' This showed a positive weighting for performance and stability; showing that more recent results in the dataset possessed better performance and stability. However, this factor is excluded from analysis henceforth as the primary aim of this research is to report how the material and device properties impact stability.

Entire Dataset				ISOS-L				ISOS-D			
Best 10 Attributes		Worst 10 Attributes		Best 10 Attributes		Worst 10 Attributes		Best 10 Attributes		Worst 10 Attributes	
Name	Weight	Name	Weight	Name	Weight	Name	Weight	Name	Weight	Name	Weight
TL1 = NDP2 doped PV-TPD	0.2974	ISOS-T-3	-0.1033	Configuration = Tandem	0.3826	Conditions = Inert	-0.2431	Active2 = PTCDI	0.1954	TL1 = PEDOT:PSS	-0.1926
Active2 = PTCDI	0.0823	ISOS-L-1	-0.1006	Active1 = ZnPc	0.1873	Configuration n = Normal	-0.1928	Configuration = Tandem	0.1375	Active2 = None	-0.1466
Active1 = PBDTTT-c	0.0571	ISOS-L-2	-0.0982	Light Type = Xenon	0.1138	Configuration n = Inverted	-0.1897	TL2 = BPhen	0.0987	Configuration = Normal	-0.0864
Electrode1 = FTO	0.0554	ISOS not compatible	-0.0902	Light Type = Halogen	0.1041	Active1 = Unknown	-0.0576	Electrode2 = Au	0.0913	Temperature	-0.0745
TL1 = ZnO (Spray coated)	0.0483	ISOS-D-2	-0.0891	Light Type = FL	0.0935	Active1 = P3HT	-0.0416	Electrode1 = Ag Grid	0.0737	TL2 = PEDOT:PSS	-0.0647
Active1 = ZnPc	0.0308	ISOS-L-3	-0.0884	Active2 = PCBM (Slot die)	0.0167	Active1 = Other	-0.0374	Substrate = PEN	0.0647	TL2 = VO5	-0.0594
Active2 = ICBA	0.0267	ISOS-D-3	-0.088	Electrode1 = Ag grid (flexible)	0.0163	TL1 = PEDOT:PSS	-0.0366	Encapsulation = Glass	0.0508	Active1 = PTB7	-0.0539
TL1 = NDP2(Noveled) doped DiNPB	0.0267	ISOS-D-1	-0.087	Substrate = Glass	0.0134	Active2 = C60	-0.0215	Encapsulation Adhesive = Epoxy	0.0416	Configuration = Inverted	-0.0511
Substrate = PEN	0.0214	ISOS-O-1	-0.0431	Type = Module	0.0095	Temperature	-0.0185	Encapsulation Adhesive = UV curable	0.0392	Active1 = PBTTTz-4	-0.0485
Light Type = LED	0.0205	Intensity	-0.0356	Electrode2 = Ag	0.0087	TL1 = None	-0.0181	Active1 = Pentacene	0.0295	Active1 = CuPc	-0.044

Table 4.5: Best 10 attributes and worst 10 attributes for  $T_{80}$  enhancement for entire dataset, ISOS-L and ISOS-D  
*D* Yellow shading represents testing conditions. Green represents architectural components.

When considering data only obtained using ISOS-L standards, the most influential attributes are also the structural components (i.e. substrate, transport layers, active layer, and electrodes), materials and configuration. The choice of light source was also found to play a significant role in improving the  $T_{80}$  lifetime.

When considering data only obtained from ISOS-D standards, the attributes that most positively and negatively affect the degradation are found to be materials, architecture and the encapsulation method, illustrating the importance of protecting the device materials from environmental conditions during thermal or damp-heat testing. This demonstrates how the degradation during dark tests is dominated by the intrinsic stability of the materials used. The use of Tandem configuration leads to better dark stability (discussed later) along with the use of Perylenetetracarboxylic dianhydride (PTCDI) acceptor material and Bathophenanthroline (BPhen) as the top transport layer. Several interesting attributes can also be identified as being detrimental; the use of PEDOT:PSS, 'normal' device configuration and unsurprisingly, temperature.

To provide a summary of the data, Table 4.6 shows the three materials possessing the highest SMOreg weights for each layer within the device, when predicting  $T_{80}$  and considering 1) the full dataset, 2) the data from only ISOS-L tests and 3) the data from only ISOS-D tests. These top three weights illustrate the three most influential materials governing the stability and performance of the OPV devices, as predicted by the SMOreg algorithm.

Greater insight into the usefulness of ML approaches in elucidating trends and identifying key components can be gained from comparison of the highest weighted factors and the distribution of the  $T_{80}$  lifetimes for each attribute, shown in Figure 4.12. Inspection of the best electrode 1 components demonstrates that the three highest weighted materials are FTO, combined Cr and Al, and Ag. However, Figure 4.12 would suggest that these materials would not yield the greatest stability. Therefore, the ML algorithm has identified potential materials for stability enhancement which would not normally have been identified. This ability of the ML algorithm stems from the mechanism by which it finds the optimum hyperplane, whereby at least two Lagrange multipliers are simultaneously minimised. Inspection of the distributions for TL1, Active 1, Active 2 and TL2 illustrates the difficulty in manually identifying key components. Except for some high performing devices, very little variation in the  $T_{80}$  lifetimes can be

identified. This is where the ML approaches are most useful; the most influential factors can still be extracted, as determined from the greatest SMOreg weights. The three materials identified for TL1 can be seen to correspond to experimental results which show significant stability. This is also true for the identified Active 1 and Active 2 materials, as well as the TL2 materials. This highlights the usefulness of this approach whereby significant features can rapidly and effectively be identified. By looking at the most stable devices found in that dataset, it can be seen that the categories with high weights are used.

Data set	Electrode 1	TL 1	Active 1	Active 2	TL 2	Electrode 2
Full	FTO	NDP2(Novaled) doped PV-TPD	ZnPc	ICBA	ZnO-np	Al
	Cr/Al	ZnO- spray-coated	PBDTTT-c	PCBM	ZnO	Ag
	Ag	NDP2(Novaled) doped DiNPB	PC-TBT-TQ	C60	PEDOT:PSS	Au
ISOS-L	Ag grid flexible	BF-DPB	ZnPc	PCBM slot-dye	BPhen	Al
	Other	PEDOT:PSS	Other	PCBM	BaF2	Ag
	ITO	Unknown	P3HT	PCBM - 71	PEDOT:PSS	Ag grid
ISOS-D	Ag grid	PEDOT:PSS	PECz-DTQx	PTCDI	BPhen	Au
	ITO	ZnO – spray-coated	MDMO-PPV	PCBM slot-dye	AlQ3	Ag
	Other	TiOx	ZnPc	ICBA	BCP	Other

*Table 4.6: Top three attributes for each layer in device architecture for the full dataset, ISOS-L and ISOS-D, when predicting  $T_{80}$ .*

However, it should be noted that the analysis performed using the SMOreg algorithm does not take into account the compatibility of the different materials identified. This is because the best attributes for each device layer are considered individually and the possibility of combining materials is not included as an attribute

in the analysis. For example, ZnO cannot be used as both the HTL and ETL in an OPV device, as suggested in row 2 of Table 4.6. In addition, ZnO and AlQ3 cannot be used in the same device since both are ETLs. Care should also be taken when considering the identified donor – acceptor blends identified since it is critical that the two materials have favourable band alignments. However, most of the identified combinations in Table 4.6 have been trialled previously, except for the blend ZnPc:ICBA.

Subsequently, the literature can be examined for examples where materials identified as being beneficial for stability are also identified in Table 4.5 and Table 4.6. For example, the use of a thin layer of chromium with its accompanying oxide has been shown to significantly improve air stability when used with Ag, Al or Au electrodes [167]. Using a ZnPc was identified for all data, ISOS-L and ISOS-D data to be beneficial for improving the  $T_{80}$  lifetime. ZnPc has also been reported in the literature has having a high photostability [168] and improved stability under high ambient temperature and humidity [169]. Zinc oxide nanoparticles (ZnO – np) have been reported using a solution processed annealing free method for producing high stability devices which showed no drop in initial efficiency after 50 days of storage in air [170].

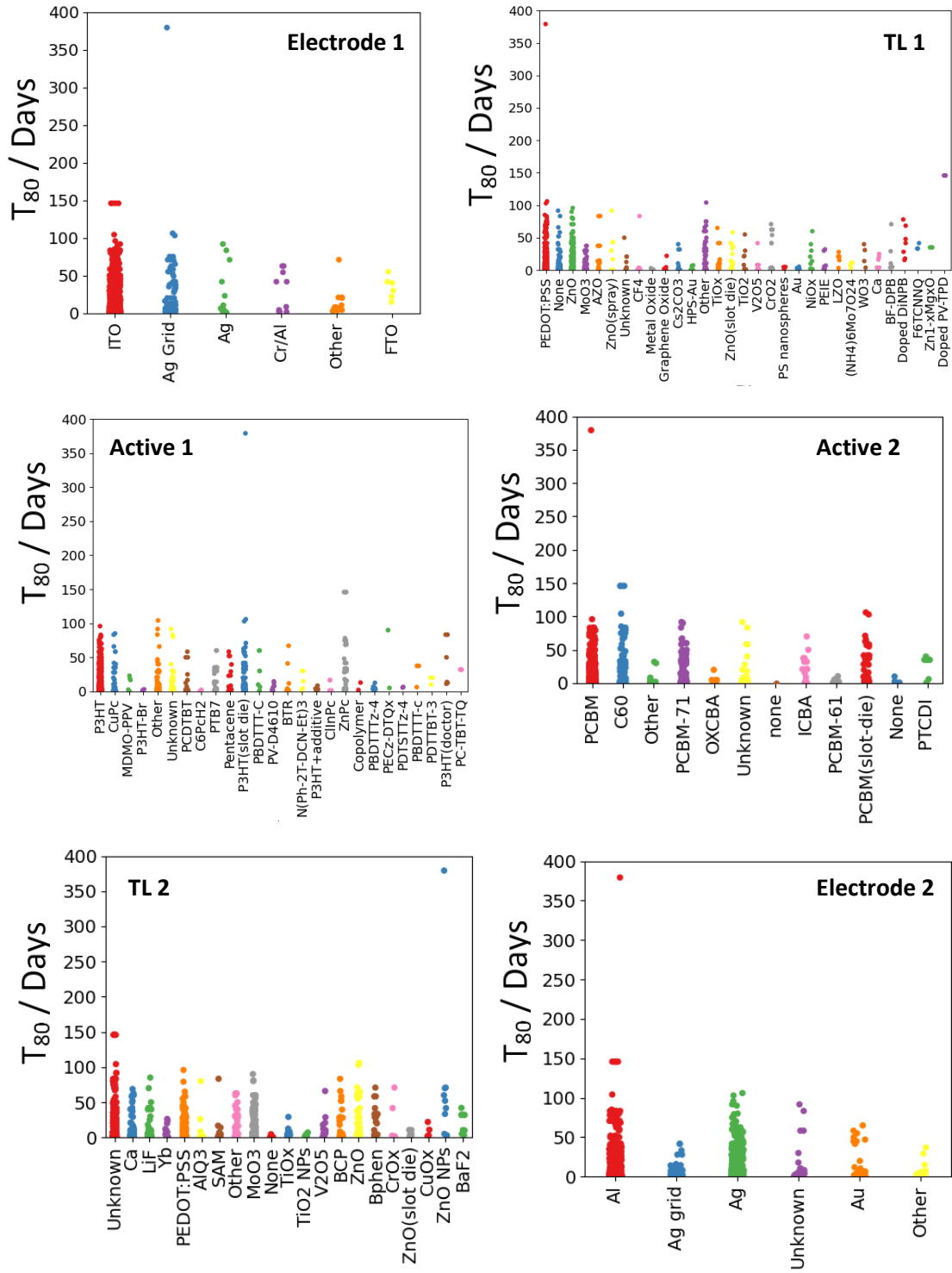


Figure 4.12: Distribution of  $T_{80}$  lifetimes for each structural class present in dataset.

#### 4.3.4.1 SMOREG Oddities and Biases Encountered

From analysis of the SMOREG output and weights, several examples of data biases were evident. The following discusses their possible origins and why they might arise. Using a Tandem configuration plays the most significant part in the  $T_{80}$  improvement along with using ZnPC as an active layer material. The former is potentially a limitation within our dataset; there are only three reports in our dataset of stability testing using tandem configurations and, in all cases, the reported  $T_{80}$  value is high. Tandem devices are also primarily targeted by groups that already master the art of single junction devices. Despite the thoroughness of our dataset, a small number of highly stable device will detrimentally bias the algorithm in favour of those architectures and testing conditions. This can occur when the data is asymmetrically distributed towards certain classes within attributes. For the highly stable tandem cells, the algorithm computes that the weightings of the normal or inverted configuration attributes have a negative impact on the  $T_{80}$  time. In addition, testing in ‘inert conditions’ is found to be a detrimental influence for stability, which is highly contradictory of scientific evidence that oxygen and water could degrade materials. However, greater inspection of the dataset shows this is associated with poorer performing devices, which were tested in early stages of development and where the general foci of the papers are not on enhancing major gains in stability. The SMOREG algorithm will yield a large number of attribute weights, related to many of the features present in the dataset. Whilst the weightings do provide guidance on their impact on stability, discretion must still be exercised when identifying the most beneficial and detrimental features, such that the interpretation of the weights is meaningful.

#### 4.3.5 Using the SMOREG algorithm for predicting the initial Stability, $E_0$

Whilst the focus of this work has been on stability data, the same methodology can be applied to understanding how different materials/architectures impact the efficiency prior to stability tests starting (defined as ‘ $E_0$ ’). The SMOREG algorithm has been applied to the dataset in order to predict  $E_0$ . The regression model prediction is shown in Figure 4.13 along with the residual plot for the model. The algorithm performance metrics can



be found in Table 4.7. The 10 most influential and 10 most detrimental attributes have also been found and are listed in Table 4.8. The distribution of  $E_0$  performances for each attribute can be seen in Figure 4.14.

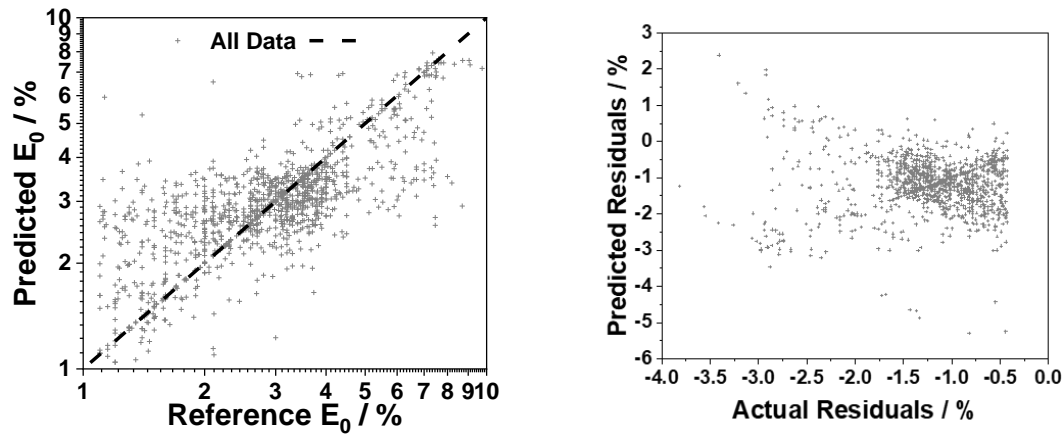


Figure 4.13: Predicted  $E_0$  performance plotted as a function of reference  $E_0$  performance for the full dataset, plotted on a logarithmic scale. The dashed line represents a perfect correlation of 1:1.

Metric	Full Dataset
Correlation Coefficient	0.739
Mean Absolute Error / %	0.605
Root Mean Squared Error / %	0.939
Relative Absolute Error	55.6%
Root Relative Squared Error	63.9%
Number of Instances	1347

Table 4.7:  $E_0$  fit parameters for  $SMO_{reg}$  algorithm applied to full dataset based on training.

Inspection of the weights for improving  $E_0$  illustrates that all attributes are architectural components such as the active layer, the transport layers and the electrodes. Similarly, most of the attributes for deteriorating  $E_0$  also correspond to architectural components. This provides a method of determining the optimum device architecture for maximised device performance.

Entire Dataset			
Best 10 Attributes		Worst 10 Attributes	
Name	Weight	Name	Weight
Active1 = PTB7	0.3572	TL1 = MeO-TPD	-0.1689
Configuration = Tandem	0.2315	Active1 = PBPTTT-T	-0.1551
Active1 = PBDTTT-c	0.2097	TL1 = NDP2(Noveled) doped PV-TPD	-0.1522
Active1 = PCDTBT	0.1769	Active1 = CuPc	-0.1372
Active2 = ICBA	0.1581	TL1 = AZO	-0.1262
Active2 = OXCBA	0.158	Configuration = Inverted	-0.1221
Electrode1 = Ag	0.1445	Intensity	-0.1185
TL1 = Graphene oxide	0.1042	Active1 = P3HT (Slot die)	-0.1133
TL2 = V <sub>2</sub> O <sub>5</sub>	0.0889	Electrode1 = Ag grid (Inkjet)	-0.1012
Electrode2 = Pt	0.0788	Active2 = None	-0.0989

*Table 4.8: Best 10 attributes and worst 10 attributes for  $E_0$  enhancement for entire dataset.*

It can be seen that the choice of active layer has the most significant impact on the value for  $E_0$  along with using a Tandem configuration. The most commonly used polymer, P3HT, is determined to have a negative impact whilst other polymers such as PTB7 and PBDTTT-c display a positive impact. By performing the regression to predict  $E_0$ , the model is being trained to understand which material combinations will result in higher efficiencies. However, care must be taken since the model does not, necessarily, know which combinations are impractical due to manufacture constraints. Therefore, interpretation of the acquired results still requires a certain level of discretion when drawing conclusions.

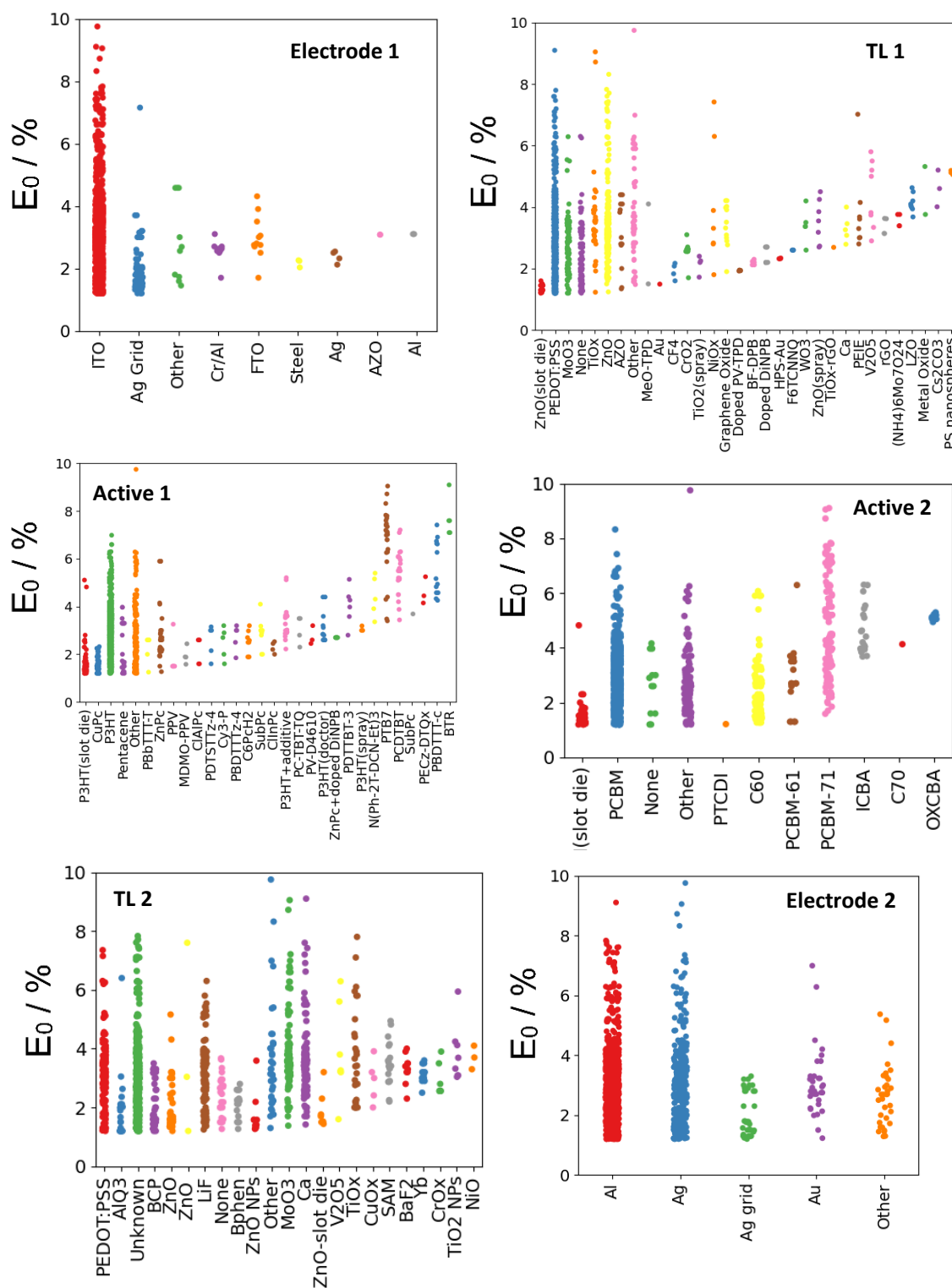


Figure 4.14: Distribution of  $E_0$  lifetimes for each structural class present in dataset.

## 4.4 Challenges and Perspectives

This work highlights some challenges that lie ahead. The success of the supervised approach can partly be attributed to the choice of the data acquired. The recently published ISOS consensus standards provide some discussion about this and state that for approaches set out in a paper to be of greater future success, standardised reporting is key [166]. The importance of training data selection was demonstrated by comparing the ‘entire dataset’ with the data only obtained from ‘ISOS-L testing’; selecting subsets of data improves the models trained on the dataset.

The attributes and categories used in the modelling represent a large portion of what is commonly used to describe OPV devices when reported in the literature. There are further factors that could be identified that could be of importance. For example, layer thickness, synthesis temperatures, chemical providers, etc. The problem with many of those parameters are twofold. One is that they are often not reported. Another is that with too many categories, the number of devices for which the data is available is too small to properly train such a model.

The models presented here have an accuracy of 40.6% and 44.4% with respect to  $T_{80}$  and  $E_0$  prediction, respectively. This is far from perfect and other phenomena which improve device performance could thus be attributed to other factors. Given the difference in performance of similar devices made at different labs a substantial, but yet unquantified, part of the discrepancy between the model and the reality could be attributed to either “dark” unreported knowledge or to hidden variables that are unknown. By looking to the future where the training data set can be expanded with more device data and with a more fine – grained meta – data mechanism, it is likely that more of those parameters can be elucidated.

The methods presented in this chapter additionally pave the way for analytical models to be developed for the performance parameters of OPV technologies. For example, material properties such as hole mobility, conductivity, electric dipole moment and diffusivity could be determined for each of the materials in the dataset and subsequently used in a quantitative approach where ML techniques are used. By modelling the performance and stability based on the physical and chemical properties, this could potentially allow the optimum set of material properties to be determined

without knowledge of the exact material previously tested. Furthermore, this could allow new materials to be developed based on the calculated set of optimised parameters. Physical and chemical models have previously been developed for silicon-based PV where the solar cell performance parameters are modelled based on the above parameters. The electrical and optical properties of silicon solar cells have been taken into account using a one-dimensional diffusion model to calculate the electron current density, spectral response and external quantum efficiency (EQE) [171]. Furthermore, the effect of illumination on charge carriers and electric currents has been investigated in [172] by using an analytical kinetic model for silicon solar cells. Using a ML approach to model similar material properties, the optimised properties could be predicted for OPV technologies.

## 4.5 Conclusion

Machine learning has been applied to a comprehensive OPV dataset as a rapid and effective screening technique for identifying the primary attributes in OPV degradation which have a positive or negative impact on the initial efficiency as well as the  $T_{80}$  lifetime. Supervised learning and regression have been implemented using the SMOreg ML algorithm and has yielded predictions for  $E_0$  and  $T_{80}$  based purely on the device architecture and testing conditions. This methodology allows the optimum material for each device layer to be identified as well as highlighting the role of the environmental conditions on stability and performance. The SMOreg algorithm has been applied to the dataset separated for tests conducted under both ISOS-L and ISOS-D testing compatibilities for  $E_0$  and  $T_{80}$ . By separating the data in such a way, the role of the different materials and stress attributes on both the operational stability as well as the intrinsic material stability can be separately identified. For ISOS-L testing, the choice of light spectrum as well as several active materials were found to be attributes which govern the modelling of performance significantly when using SMOreg. In contrast, for tests conducted under ISOS-D, the primary attributes affecting the stability were predominantly material dependent along with the use of encapsulation.

From inspection of the most promising materials and device architectures identified via this methodology, it is hoped that future studies will build on these findings

in order to experimentally verify and test these predictions. In addition, the materials identified herein will provide inspiration for other research groups to develop and enhance the stability of their devices.

Scientific progress relies on the efficient assimilation of results from a plethora of research outputs in order to choose the most promising way forward and to minimise re-invention. The capability of identifying the best material combinations and the role of the testing condition in the performance and stability of OPV devices is crucial for identifying the priorities of research. In addition, classes of materials that are more robust against environmental stresses can be identified whilst classes of materials with lower stabilities can be avoided.

## Chapter 5. Outdoor testing and forecasting of Diurnal Performance and Degradation Rates

### 5.1 Introduction

The performance and stability of organic photovoltaic (OPV) devices and modules depends on a wide range of environmental factors and conditions [173]. However, due to the relatively short period that OPV and other emerging technologies have been under development, their long-term stability at different stresses is not fully understood [93]. Several climatic conditions may influence the performance and subsequent power output [173]; outdoor monitoring provides a platform under which the combined effect of a multitude of different stresses can be tested [91], [174], [175]. In order to fully realise the potential for OPV technologies, it is highly desirable for consumers and industries alike to be able to predict the expected power output of an array of modules, given only the climatic conditions. The power output from an OPV module is, fundamentally, governed by the irradiance (power per unit area) incident on the module. However, in many circumstances, the data available may be limited and incomplete, with not all sites having the ability to measure the irradiance, or even the local climatic conditions.

Machine learning provides an ideal methodology to address this task since the quantity of data that can be gathered from both weather stations, deployed in various climates, as well as solar arrays, is vast. In this chapter, a variety of ML algorithms, such as those based upon multivariate linear regression (MLR) [142] and multilayer perceptron (MLP) [149], have been used to predict the performance of OPV modules based on climatic conditions and consequently the degradation rates can be determined based on both the cumulative and instantaneous climatic stress factors. This is the first example where cumulative stress effects have been used to predict the degradation of OPV modules in real-world scenarios using ML.

ML approaches are well placed to perform forecasting since the determination of OPV outdoor degradation is intrinsically a multifactorial problem. By considering many attributes in conjunction with each other, an improved understanding of the role of these

attributes can be achieved which may not be possible by testing the effect of each attribute individually in lab settings, or in outdoor conditions. The ML method provides the possibility of identifying trends and patterns which may not be evident from traditional analytical techniques as well as providing information about how the attributes operate with each other.

The analysis presented forthwith is organised into four main topics: qualitative analysis of performance and degradation using PCA, diurnal predictions of performance parameters and daily yield *via* the multilayer perceptron, degradation modelling *via* multivariate linear regression and finally a novel method of combining the results of the multilayer perceptron with the multivariate linear regression for improved daily yield forecasting.

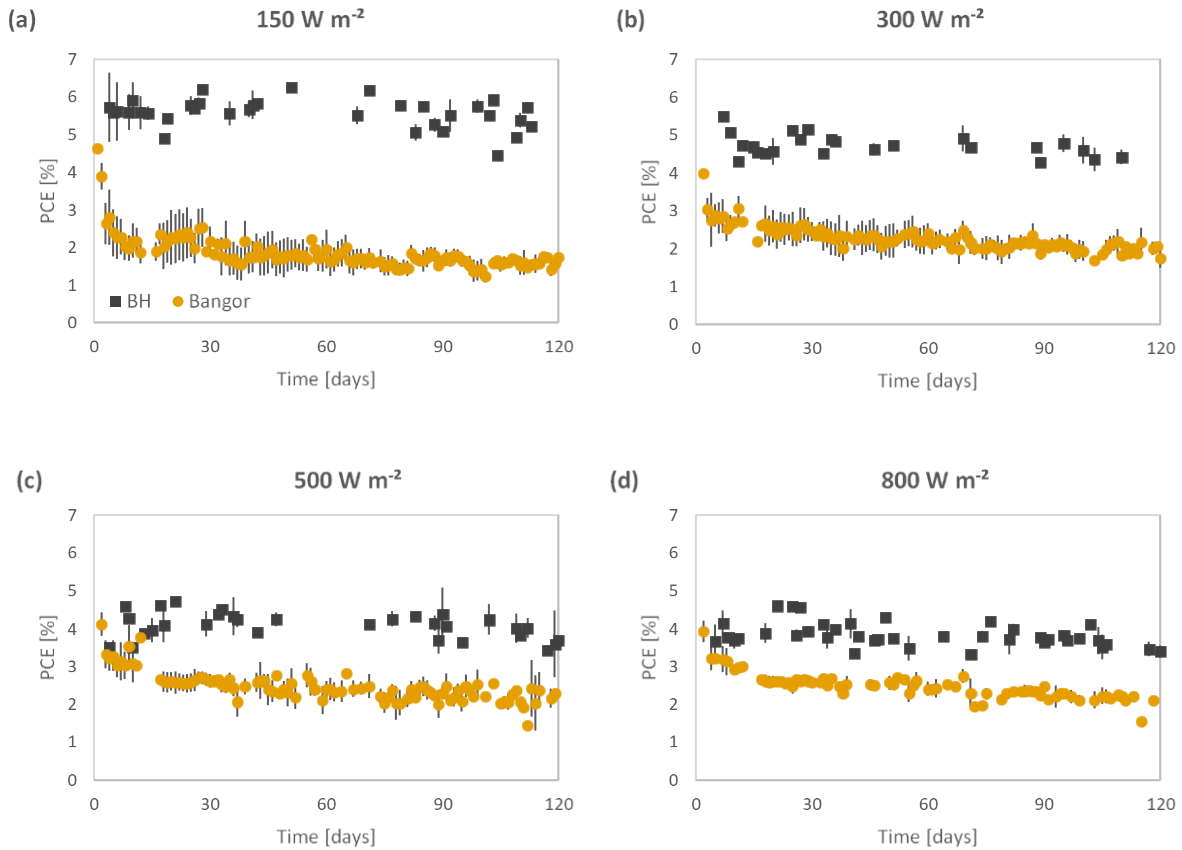
## 5.2 Results and Discussion

### 5.2.1 Outdoor testing of OPV modules at different latitudes

Initially, data had to be sourced for ML analysis. Outdoor testing was conducted on OPV modules at Bangor, North Wales and at Belo Horizonte Brazil in order to test the device stability at two different locations. Advice was given to the group in Belo Horizonte about how to undertake these measurements and a setup communicated based on the setup at Bangor. This will allow for an experimental determination of the effects of different climates on the degradation of OPVs, as well as allow for a collation of a large dataset of OPV performance in outdoor conditions which can subsequently be used for forecasting and applications using machine learning techniques.

The devices being tested in outdoor conditions are described in Chapter 3 section 3.2 along with the setup for outdoor testing. These modules were encapsulated with a multilayer PET based barrier film with a water vapour transmission rate of the order of  $10^{-3} \text{ gcm}^{-2}\text{day}^{-1}$ . Monitoring started in early Spring 2018 in Belo Horizonte and in Bangor and the PCE results are shown in Figure 5.1.





*Figure 5.1: Degradation for modules tested in Belo Horizonte and Bangor in May, measured at irradiances of (a) 150 Wm<sup>-2</sup>, (b) 300 Wm<sup>-2</sup>, (c) 500 Wm<sup>-2</sup> and (d) 800 Wm<sup>-2</sup>.*

It is observed that modules tested in Bangor exhibit significantly faster degradation as compared to modules tested in Belo Horizonte. In Bangor, the module degradation exhibited significant burn – in, resulting in a ~40% loss in PCE, followed by a linear degradation thereafter. In Belo Horizonte, the burn-in could not be easily seen since the first days of data were cloudy. However, considering that the samples had very similar initial parameters, it is likely that these modules did not experience a high initial degradation and were, thus, more stable in Belo Horizonte. By considering the electrical parameters, shown in Figure 5.2, it can be seen that  $J_{SC}$  values degraded at a similar rate at both sites, whilst the  $V_{OC}$  and FF dropped at a greater rate in Bangor.

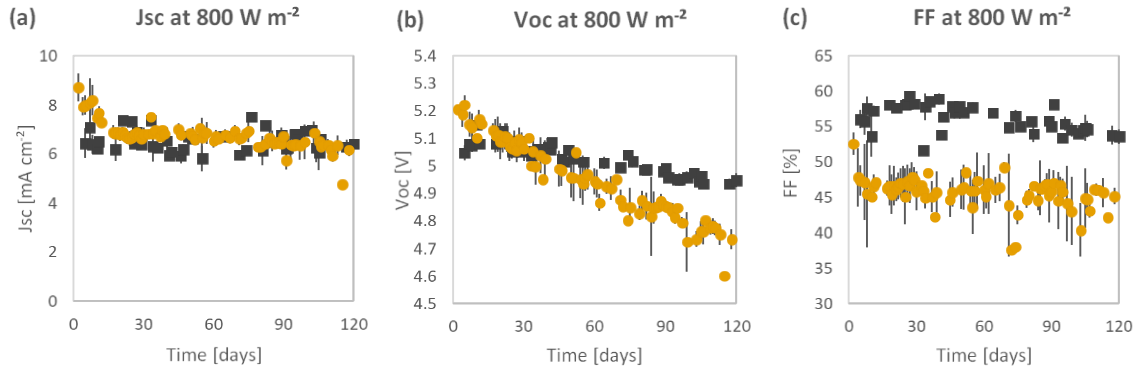


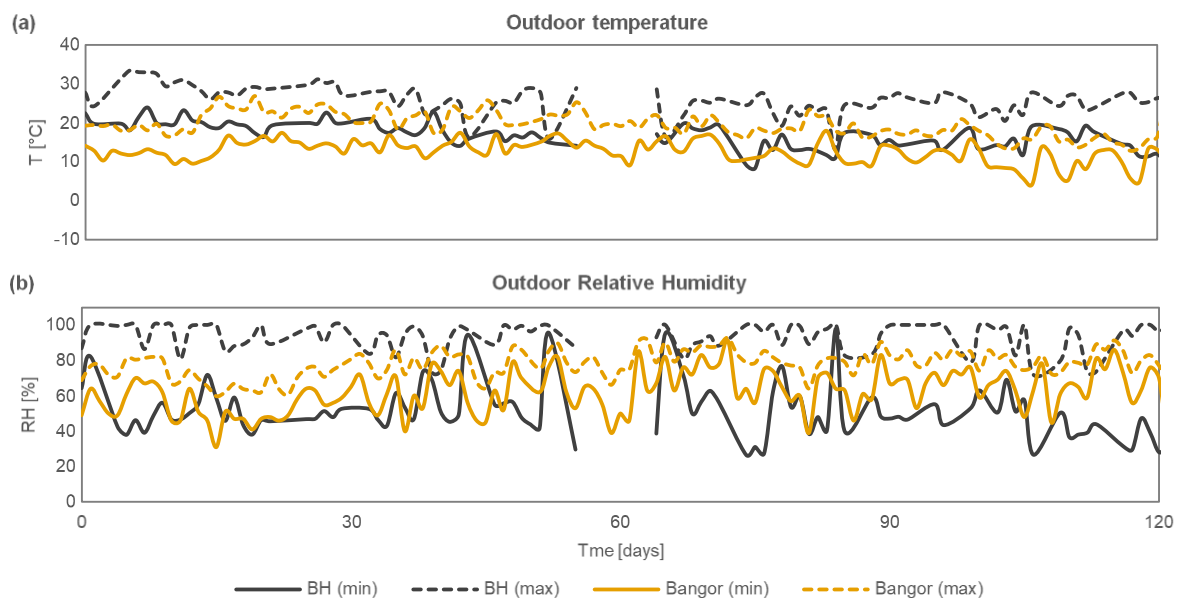
Figure 5.2: Degradation of the electrical parameters of the modules at 800 W m<sup>-2</sup>: (a)  $J_{sc}$ , (b)  $V_{oc}$  and (c) FF.

Figure 5.3 illustrates a comparison between the variation in environmental conditions in Bangor and in Belo Horizonte. The temperature and maximum levels of irradiance daily values were higher in Belo Horizonte, as well as the energy dose: 1800 MJ m<sup>-2</sup> against 1505 MJ m<sup>-2</sup> in Bangor, although the relative humidity was similar. However, the elevated ambient temperature and relative humidity in Belo Horizonte do not seem to consequentially increase the degradation rate of the OPV modules.

Relative humidity represents the proportion of moisture in air, in comparison to the total moisture which the air can potentially hold at any specific temperature. As air temperature increases, the total quantity of moisture that can be held by the air also increases. Therefore, for a constant quantity of moisture, if the air temperature increases, the relative humidity decreases since the air can now hold more moisture. Similarly, if the air temperature decreases, the relative humidity increases since the air can hold less moisture. High relative humidity will generally occur when air temperature drops sufficiently, such that it approaches the dew point. Therefore, under this circumstance it would be more likely that condensation forms, due to the ambient temperature reaching the dew point. This can subsequently lead to greater degradation due to damage to encapsulation, contacts and device layer materials.

Given that both sites used the same experimental procedure and modules, the reasons for the variation in stability are limited. Potentially transportation could induce some mechanical issues by causing damage to the modules during packaging and handling; the modules would be in transit for an extended period. If the devices have been

damaged during transit, then this would facilitate easier infiltration of moisture, especially if exposed to greater condensation and higher levels of relative humidity. This could arise due to damage to edge sealant or delamination of the PET based encapsulant. Damage to devices could potentially arise due to mechanical stress during handling and transportation weakening the bonded PET. Additionally, the contact solder bond could be weakened due to mechanical stress. The measurement system in Bangor is located around 400m east of the Menai straits, so salinity is likely to be higher (although prevailing winds come from the west). This could be a contributing factor to the greater degradation observed in Bangor, but the increased levels of condensation on modules in Bangor could also be an issue. Condensation could induce several failure mechanisms, such as weakening of barrier layers and the adhesive, absorption of water into the modules as well as higher levels of localised relative humidity. The weakening of the barrier layers, particularly around the electrodes were noted during the tests and warranted further investigation to the causes, as this was not noted in Belo Horizonte.



*Figure 5.3: Variation in environmental conditions during the testing period, with the maximum and minimum daily values of (a) temperature and (b) humidity.*

Although more rainy days were observed in Belo Horizonte, the daily amplitude of relative humidity was similar to Bangor, but the minimum levels of relative humidity were often much lower than Bangor as a result of the higher temperatures in Belo

Horizonte. In Bangor, the maximum and minimum values of relative humidity were very close and constantly high. Combined with low temperature, this could indicate higher condensation, which could have a significant impact on the module degradation.

To evaluate this, an approximate calculation of the dew point was made using the Magnus-Tetens approximation,

$$T_d = \frac{b \cdot \alpha(T, RH)}{a - \alpha(T, RH)} \quad 5.1$$

and

$$\alpha(T, RH) = \ln \frac{RH}{100} + a \cdot \frac{T}{b+T} \quad 5.2$$

where  $T_d$  is the dew point,  $T$  is the temperature,  $RH$  is the relative humidity of the air and  $a$  and  $b$  are coefficients. From the Magnus-Tetens approximation the constants can be determined empirically as,  $a = 17.62$  and  $b = 243.12$  °C [176], [177]. One way to measure the level of condensation at both sites is to consider the average dew point depression (DPD), i.e., the difference between the ambient temperature and the dew point, which was calculated for each day of testing at both locations. Figure 5.4 presents this data in a histogram.

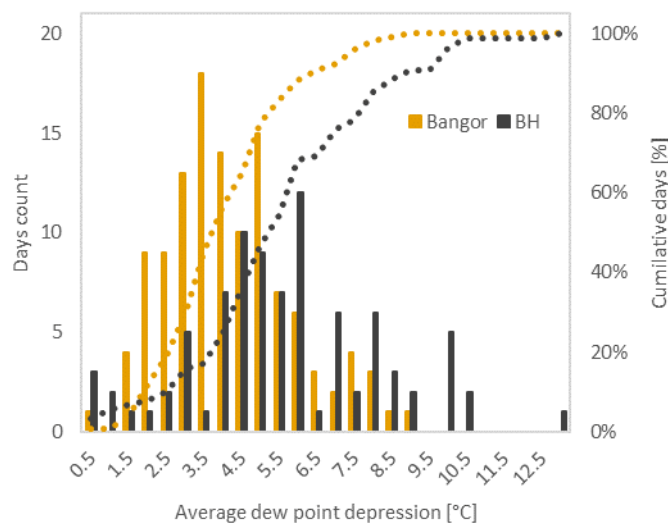


Figure 5.4: Histogram of the average daily dew point depression during the test period.

The graphs show that, in both periods, the average dew point depression in 50 % of the days analyzed was lower than 3.5 °C in Bangor, against 5.5 °C in Belo Horizonte. When raised to 80 %, the numbers change to 5.5 °C and 7.9 °C, respectively. As there are more hours where the ambient (and hence module) temperature is closer to the dew point in Bangor, it can be deduced that condensation levels in Bangor were higher than in Belo Horizonte, which could have increased the water penetration through the encapsulation.

Condensation is rarely studied in the context of PV degradation, possibly because this is unlikely to be a major issue in crystalline silicon modules given the use of strong non-flexible encapsulants. However, flexible OPVs are encapsulated with polymeric films that are prone to water penetration, which can degrade contacts, transport layers and active layers [91], [97]. Therefore, special attention is required in this case. In order to quantify the effect of condensation on the degradation of OPV modules, tests have been performed on six identical modules tested inside an environmental chamber with a controlled environment. The testing was conducted in accordance with ISOS-D-3 conditions where the temperature and relative humidity (RH) of the chamber are set to 65 °C and 85% respectively. Figure 5.5 (a) shows the normalised degradation of the PCE of the modules and Figure 5.5 (b) shows the time to reach  $T_{80}$  and  $T_{50}$  as a function of DPD. Two modules were tested without cooling and two modules were placed on a Peltier cooling device which lowers the module temperature to 60 °C and 57 °C. As the dew point at 65 °C is 61.4 °C, the tests correspond to a dew point depression of 3.6 (for the control sample) and -1.4 and - 4.4 °C for the cooled samples respectively. In practice, a negative dew point is unlikely to occur in operation, but this test allows for an extrapolation of degradation to occur as a result of condensation.

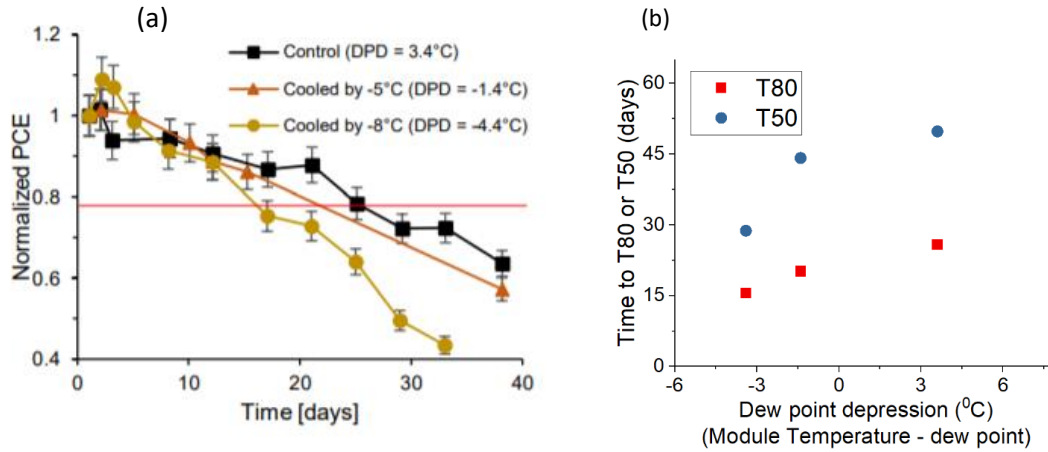
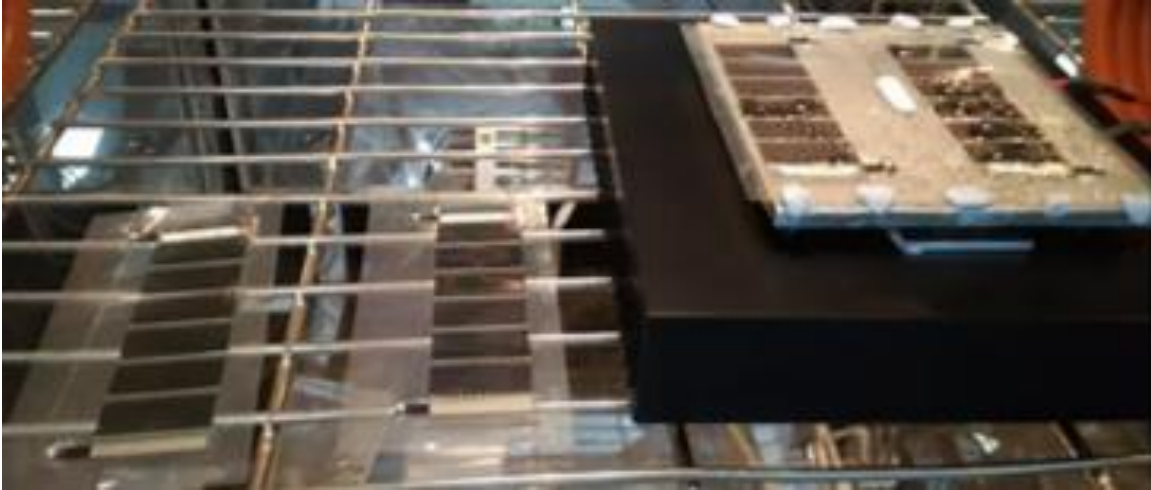


Figure 5.5: (a) PCE of OPV modules tested under ISOS-D-3 conditions (65°C, 85%) with different levels of cooling applied to induce greater condensation on modules, which leads to different dew point depression (DPD) values and (b) effect of dew point depression on Time taken to reach 80% ( $T_{80}$ ) and 50% ( $T_{50}$ ) of the original value.

Figure 5.5 (b) shows the impact of cooling the modules during the ISOS-D-3 tests. It is clear that the modules that have been cooled the most, exhibit the greatest degradation. In doing so, more condensation is formed on the module surface, providing strong evidence that modules, when operated in Bangor, would exhibit faster degradation, with greater periods at lower DPD ranges.

An image of the test is shown in Figure 5.6 with the climate chamber set at 65 °C and 85% RH. Two samples were placed on a Peltier with a heat sink and cooled at 5 °C and 8 °C (test under 5 °C being depicted). The image shows how water droplets formed on the cooled devices whilst not on the samples held at the ambient chamber conditions.



*Figure 5.6: Samples under the indoor test. The climate chamber was set at 65 °C and 85% RH and two samples were placed on a Peltier and cooled at 5 and 8 C (test under 5 °C being depicted). The formation of water drops on top of the samples is evidence of induced condensation.*

### 5.2.2 Machine Learning Methodology

The methods for module installation and acquiring the module data, irradiance levels and weather conditions, in outdoor conditions, are presented in chapter 3, sections 3.2 and 3.5. The ML techniques used to forecast the OPV diurnal performance, daily yield and degradation rates are presented in chapter 3, sections 3.4.1, 3.4.2, 3.4.4, 3.4.5 and 3.4.6. The computational methods have been utilised using Python 3.7 and the principal python packages: Pandas, Numpy, Matplotlib, Datetime, Scipy and Scikitlearn.

The developed python program allows the acquired data to be compiled into a single dataset. In addition to the data gathered from the outdoor monitoring system, the module temperature is derived and included as an attribute in the dataset, using equations 2.6 and 2.7. The acquired plot of module temperature variation as well as the ambient outdoor temperature is shown in Figure 5.7. The plot shows the variation in both parameters over the course of 15 days for clarity. As discussed in section 2.4.3, the module temperature can be calculated from knowing the wind speed and ambient outdoor temperature. When the temperature rise of the module is plotted as a function of the ambient temperature, a straight line will be acquired with gradient equal to the

Ross coefficient. Up to wind speeds of  $5 \text{ ms}^{-1}$  the Ross coefficient scales linearly with wind speed and remains constant above  $5 \text{ ms}^{-1}$ .

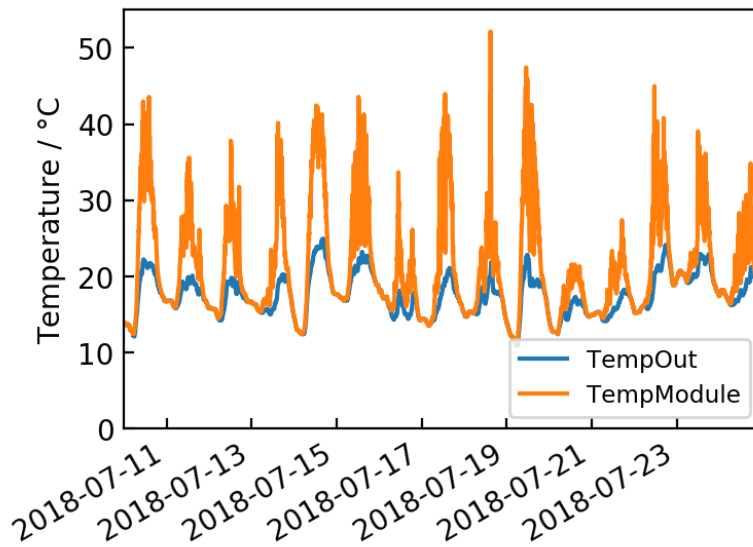


Figure 5.7: Variation in ambient outdoor temperature and derived module temperature.

From Figure 5.7, it is observed how the derived module temperature can be significantly greater than the ambient outdoor temperature. This could suggest that the module temperature would have a more significant impact on the OPV stability as compared with the ambient outdoor temperature.

For training purposes, the ML algorithms are applied to an approximately stable period of 15 days, when 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: irradiance, ambient temperature, module temperature, humidity, dew point, wind speed and UV index.

In the case of forecasting the diurnal performance of OPV modules, the MLP model was first trained *via* supervised learning on a certain set of training day data, based on the climatic conditions. The same model is subsequently applied to unseen climatic conditions, corresponding to days in the future, 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*,



$$Y = \sum_t I_{MPP} V_{MPP} t, \quad 5.3$$

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

Multivariate linear regression is employed to predict the degradation curve for the modules over a period of 6 months as well as the degradation rates of the OPV modules for both summer and winter months, giving two datasets to test. Other ML methods were also tested, however, MLR proved to be the most robust. This prediction is performed with the cumulative climatic effects. Predicting the performance parameters based on the instantaneous climatic conditions is suitable when predicting the diurnal power output of the OPV modules. However, when predicting the degradation, the cumulative effects are required such that the total exposure of the modules to different climatic conditions can be taken into consideration and the historical effect of the total exposure on the degradation state of the module can be identified.

### 5.2.3 Principal Component Analysis of Performance and Degradation Rate

After acquiring the data, ML approaches were used to analyse the data so that predictions on performance and stability could be made. Initially, a qualitative examination of the performance and degradation rates of the modules over the course of the six-month test is presented using principal component analysis (PCA). PCA allows for an enhanced understanding of the variation of the performance and degradation due to the different weather conditions considered. Figure 5.8 (a) – (e) show the score plots for the PCE,  $I_{sc}$ ,  $V_{oc}$ , FF and degradation rate respectively, and Figure 5.8 (f) shows the corresponding loadings plot. The x and y axes for Figure 5.8 (a) – (e) represent the projected score values for principal components 1 and 2 respectively. This is equivalent to representing the high dimensional dataset in a lower dimension coordinate system. The score value of each principal component is determined by projecting each observation in the dataset onto each of the principal components and gives the coordinate value along the principal components.

The PCA analysis allows the factors governing different features of the data to be visually represented. For example, attributes in the loadings plot which spatially align with score values correlate positively and vice versa. The score plot in Figure 5.8 (a), shows that high PCE values reside on the right-hand side of the plot whilst low PCE values reside on the left. By comparing with Figure 5.8 (f), 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. However, for  $I_{sc}$ , a greater spread of low values can be seen, governed by humidity, pressure and wind speed, whilst high  $I_{sc}$  values are very much governed exclusively by irradiance, UV and, to a lesser extent, temperature. The correlation between humidity and low  $I_{sc}$  is likely to be associated with increased rain, introducing a complex variation in the light available for absorption.  $V_{oc}$ , conversely, displays a much greater spread of high values correlating with wind speed (which can cool modules), irradiance, UV and temperature, and a strong correlation is observed between low  $V_{oc}$  values and humidity and dew point.

For the degradation rate, shown in Figure 5.8 (e), 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. High irradiance, UV and pressure correlate negatively with an intermediate degradation rate. This would suggest that as the irradiance, UV and pressure increases the intermediate degradation rates decreases and higher degradation could occur. The PCA analysis serves a purpose for qualitative examination as it allows for identification of which factors should be included in the forecasting 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.

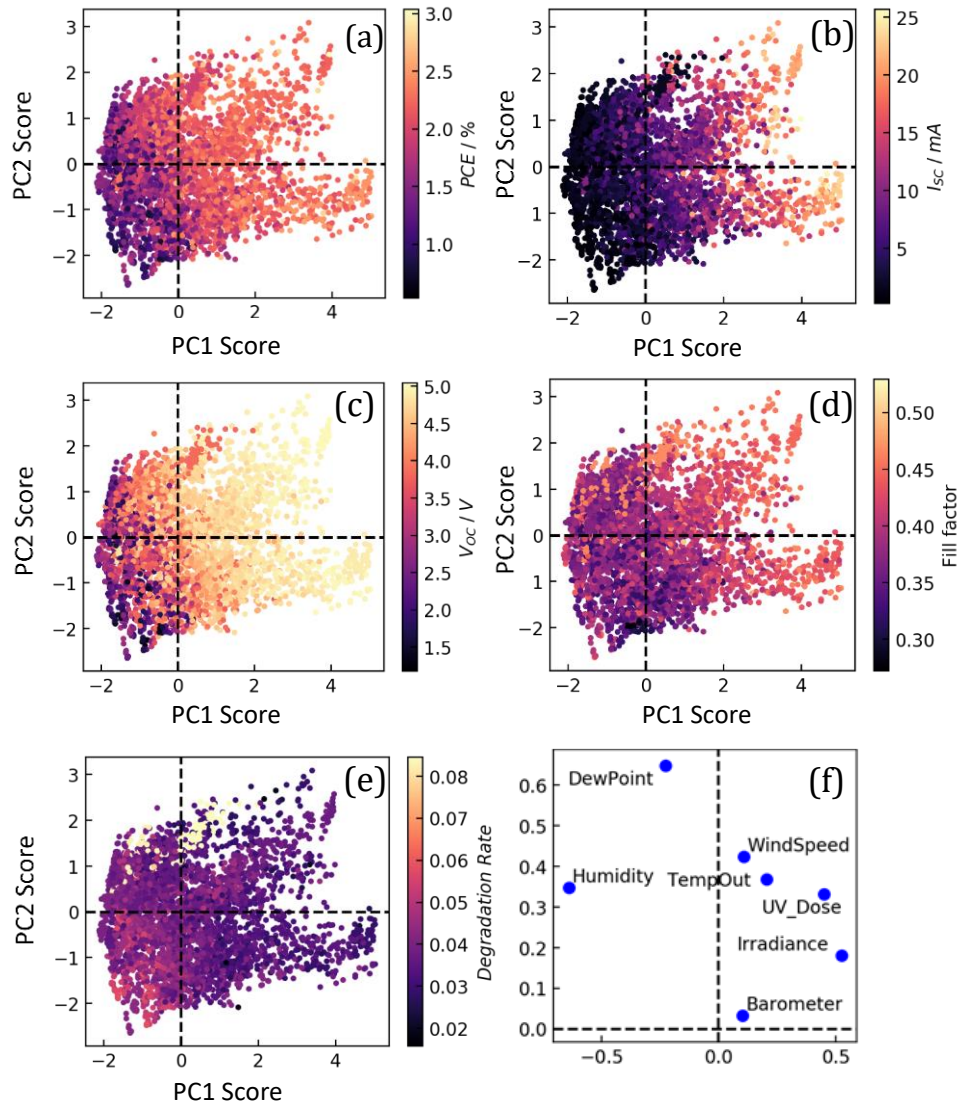


Figure 5.8: Score plot for (a) PCE, (b)  $I_{sc}$ , (c)  $V_{oc}$ , (e) FF and (f) corresponding loadings plot for all modules under test for period of 5 months in outdoor conditions.

To further illustrate how the different weather conditions influence the performance of the OPV modules, Figure 5.9 shows the correlation matrix for each weather conditions and the PCE. As would be expected, the highest correlations are between the PCE, irradiance and UV. However, notable correlations are also observed between PCE and temperature and PCE and wind speed.

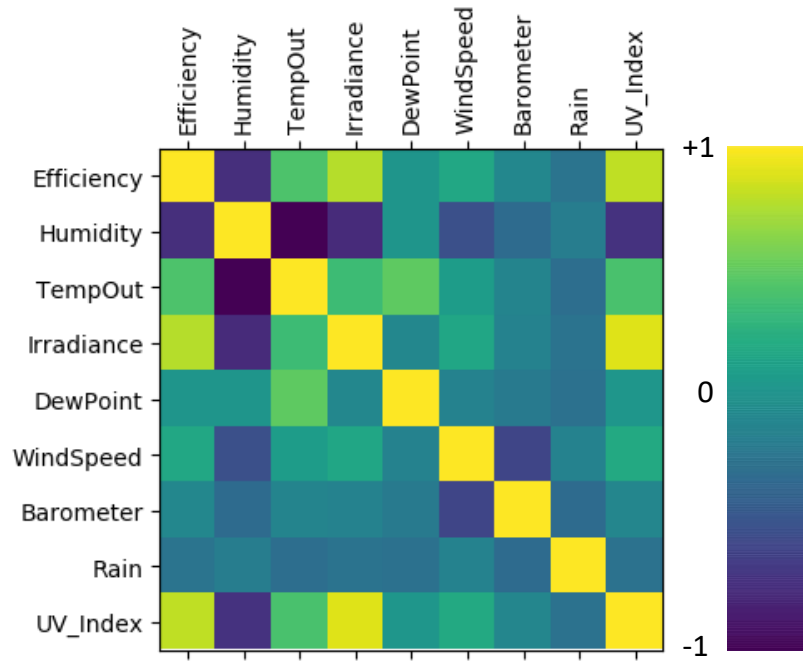


Figure 5.9: Correlation matrix for each weather condition and the PCE.

#### 5.2.4 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, i.e. the number of weeks of data that the model was trained on. Figure 5.10 illustrates the variation in the correlation coefficient (CC) when predicting the PCE of the modules based on different durations of training and testing. The computed CC corresponds 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. 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 two weeks of training data is required. Only marginal improvements in diurnal PCE prediction are achieved beyond the two-week training set.

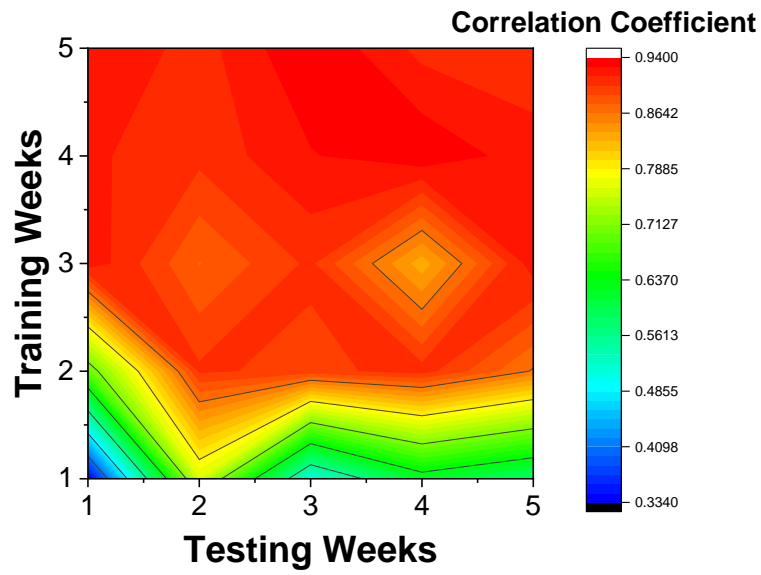


Figure 5.10: Two-dimensional plot showing the variation in correlation coefficient when predicting the PCE based on various durations of training and testing.

Trained Attribute	CC for test set
Irradiance	0.96
Module Temperature	0.76
Outside Temperature	0.45
UV index	0.82
Humidity	0.25
Wind Speed	0.32
Dew Point	0.17
All	0.96
Without Irradiance/UV/Outside Temperature	0.87

Table 5.1: CC for test set for different attribute selection.

The variation in the CC is optimised based on different attribute selections and is shown in Table 5.1, where the CC when testing with each attribute individually is shown as well as for all attributes and when irradiance, UV and outside ambient temperature are not included. The outside temperature is not included such that overfitting is avoided, due to the high correlation between the outside ambient temperature and the module temperature.

The highest CC is obtained when all attributes are employed. However, irradiance and UV measurements are not quantities which would be exactly known when making a prediction. Therefore, the CC when these are ignored is determined, and it is found that a reasonably high CC of 0.87 is still obtained. Therefore, the study forthwith employs this procedure where irradiance, UV and outside temperature are not included in the analysis. The training of the MLP model is conducted on each of the performance parameters.

Using the two-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 5.11 (a) – (d), which shows the predicted and actual diurnal cycle in the performance parameters. The close correlation gives confidence and illustrates that the diurnal OPV 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 5.3. Figure 5.11 (e) illustrates the diurnal power output over the course of the same day and Figure 5.11 (f) shows the diurnal power output for the closest overcast day for comparison.

Table 5.2 quantifies the actual and predicted daily yields for the days shown in Figure 5.11. From inspection of the values obtained in Table 5.2, 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 forecast. 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 5.2 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 5.2.5 reports on how this can be reduced by the introduction of a degradation model.

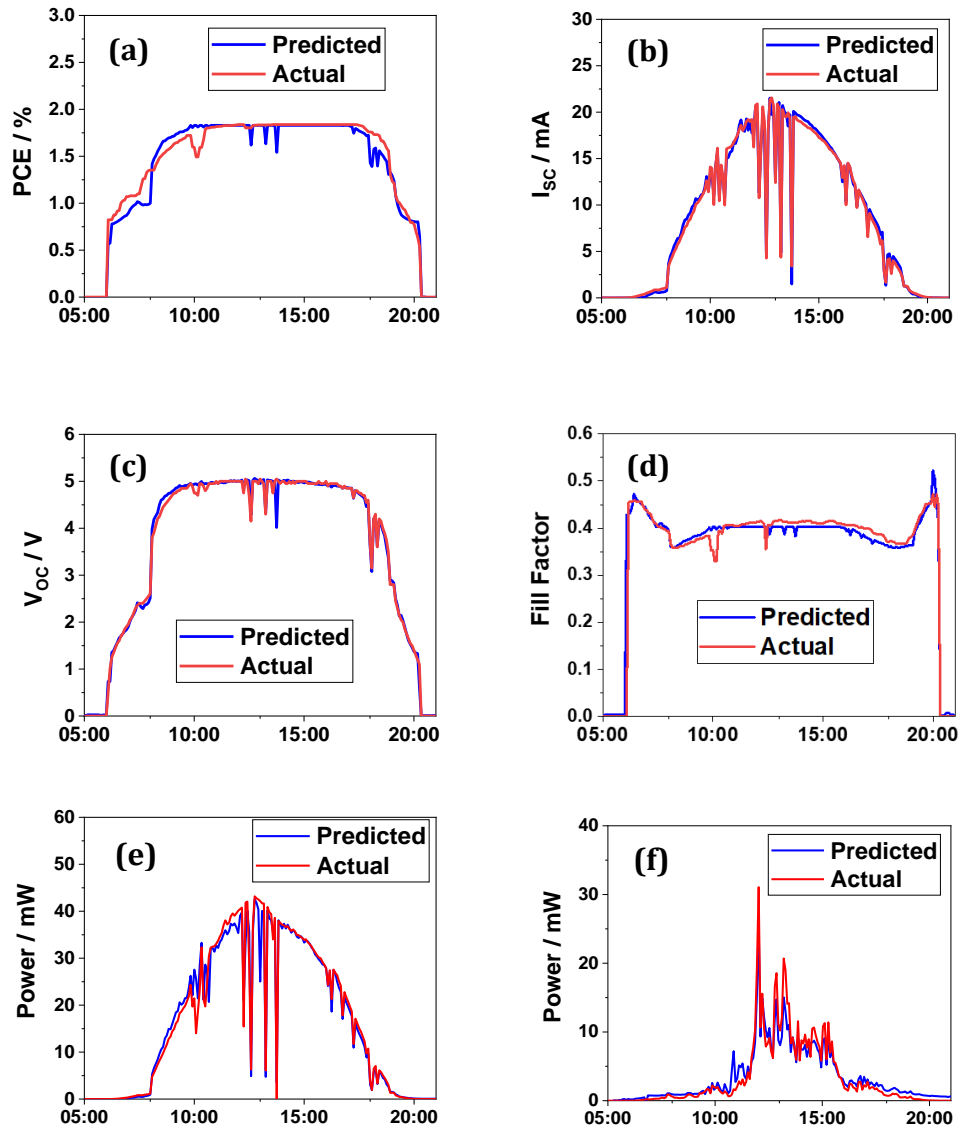


Figure 5.11: Diurnal variation in (a) PCE, (b)  $I_{sc}$ , (c)  $V_{oc}$  and (d) Fill Factor, (e) power, testing on day after training period and (f) power on cloudy day.

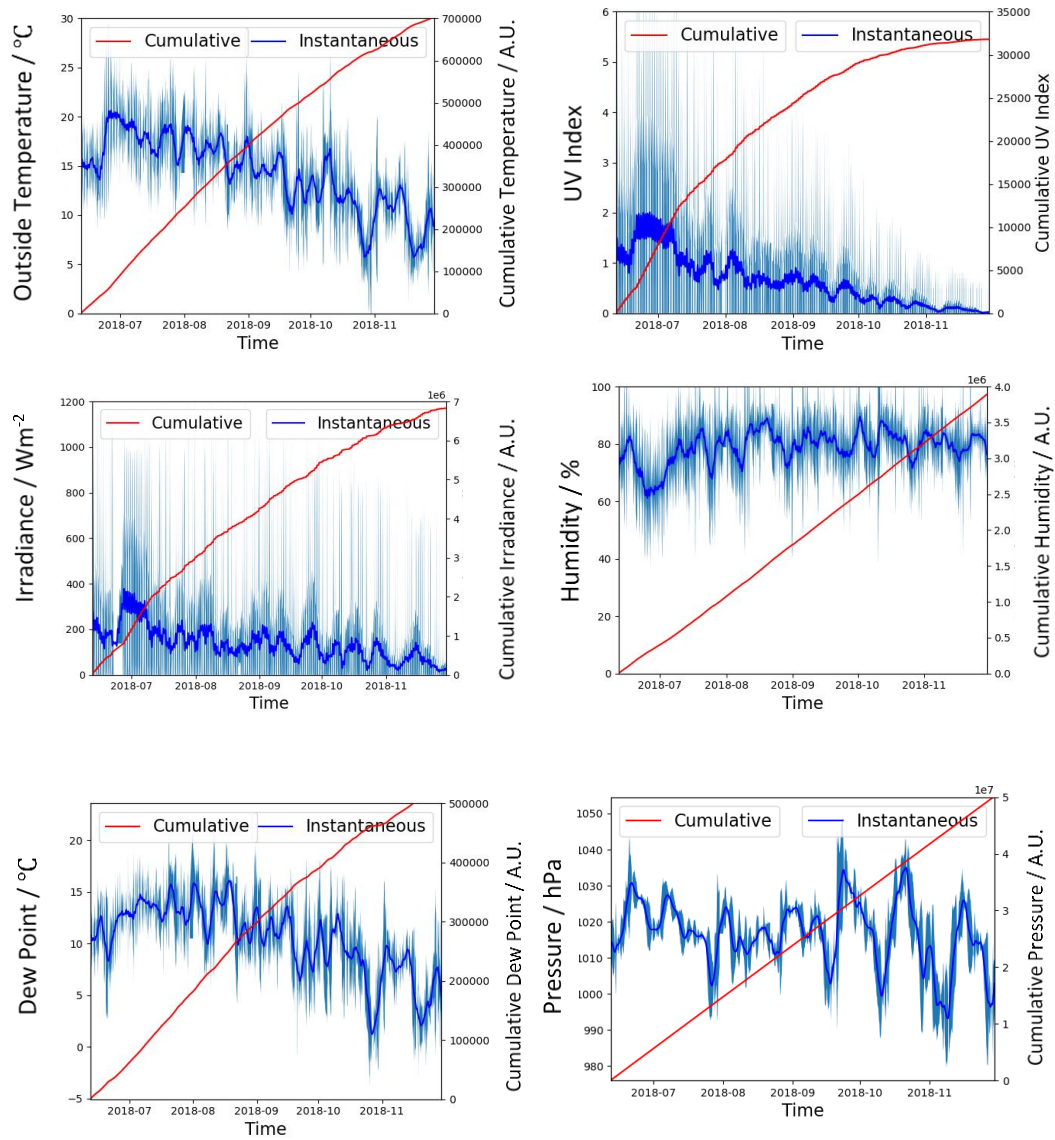
Date	Actual Yield	Predicted Yield	Error
26/8/2018 (cloudy day)	53.2 Wh	60.8 Wh	+ 12.5%
5/9/2018 (sunny day)	276.6 Wh	270.7 Wh	- 2.2%
12/6/2018 – 30/11/2018 (without degradation model)	23.9 kWh	28.9 kWh	+ 20.9%
12/6/2018 – 30/11/2018 (with degradation model)		24.9 kWh	+ 4.1%
26/1/2018 – 26/4/2018 (with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

*Table 5.2: Actual and predicted daily yields for a cloudy day (26/8/2018) and a sunny day (5/9/2018) as well as the total yield for the entire test duration for summer (12/6/2018 – 30/11/2018) and winter (26/1/2018 – 26/4/2018).*

### 5.2.5 Degradation Forecasting of OPV Modules

In order to forecast 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 at any point in time is dependent on the historical exposure of the modules to different stress factors. As a result, a ‘cumulative dose’ of the various weather conditions has been used as a factor for multivariate linear regression (MLR) analysis in order to forecast the degradation of the OPV modules. The cumulative dose was determined by summation of the instantaneous values of each weather condition over time. To illustrate this process, Figure 5.12 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 also illustrated; the thick blue line represents the rolling average value of each weather condition. Since the test progresses from Summer into late Autumn, different stress factors show different rates of change over the measurement period.

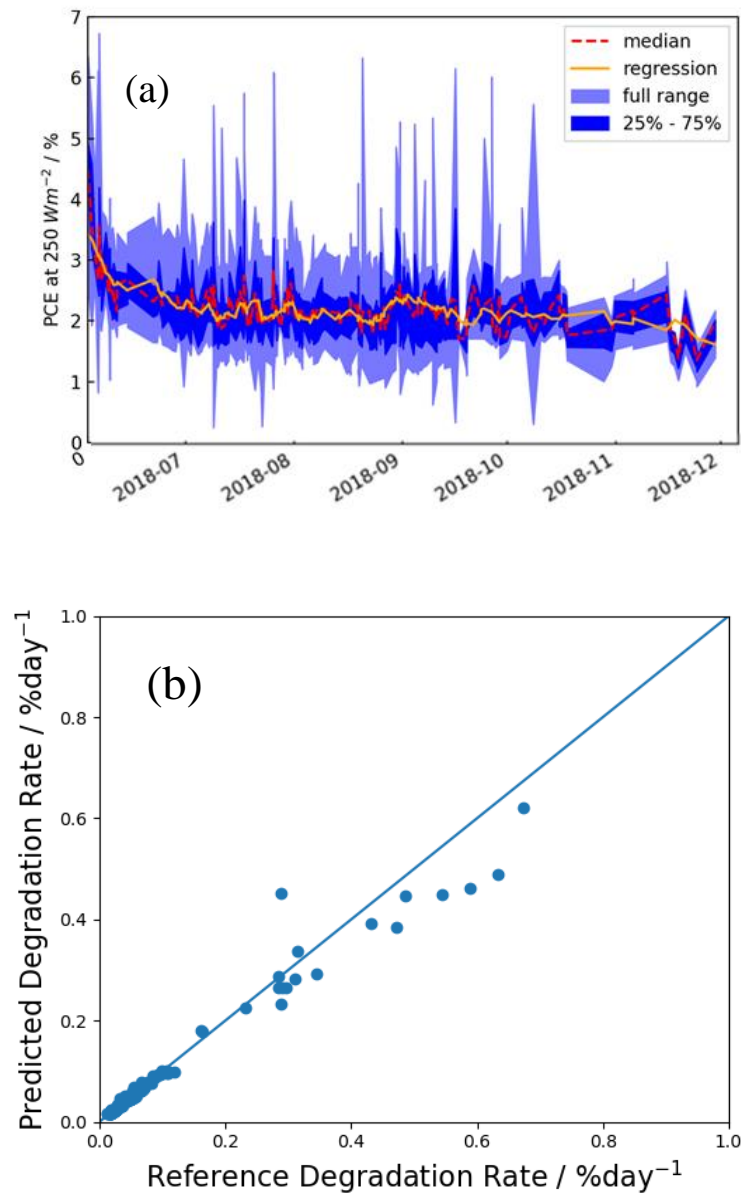




*Figure 5.12: Variation in instantaneous and cumulative weather conditions over the course of 6 months.*

Figure 5.13 (a) illustrates the forecasting results where the median line for the PCE at  $250 \text{ Wm}^{-2}$  is shown along with the full range of data, 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 5.13 (b) shows the regression plot,

illustrating the predicted degradation rates as a function of the reference degradation rate.



*Figure 5.13: (a) Forecasted degradation of OPV modules at  $250 \text{ Wm}^{-2}$ , showing the regression prediction, the rolling median of all the data, the full range of data and the interquartile range. (b) Degradation rate regression, showing the predicted degradation rate as a function of the reference (measured) degradation rate.*

A good fitting can be seen for low degradation rates, however for higher degradation rates (which occur during the burn-in 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 which govern the burn-in. The burn-in is an internal process occurring within the module's active layer [178]. Therefore, this is not surprising since modelling these processes would require consideration of the factors governing the intrinsic processes and not only the extrinsic degradation factors governed by the weather conditions. Therefore, the burn – in period introduces an additional complexity in modelling the degradation; if the model was trained on only the burn – in period, then a significantly faster rate of degradation would be predicted for latter periods since the model would expect the degradation to proceed in the same way after burn – in and would not model the slower linear decay. Similarly, if all the data is used for training, the majority of the test represents periods which do not display burn – in and the process cannot be easily modelled through learning the data structure. This arises due to the different degradation mechanisms which occur during the two periods. However, since the objective here is to model the long – term degradation, understanding the performance variation during burn – in is not as critical.

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 OPVs 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 5.14 (a), which ranks the factors which 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 UK, which

correspond to higher levels of irradiance, UV dose and absolute humidity, whilst low pressure is associated with poorer weather conditions with more cloud cover and rain. Therefore, whilst air pressure itself is not the cause degradation, it will be a governing factor for other weather 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 yield forecasting approach can be conducted. The degradation rate analysis is included into the MLP diurnal prediction so that yield over several months of outdoor operation can be predicted. Figure 5.14 (b) illustrates the actual daily yield variation over the course of the 6-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 to be 24.9 kWh, compared to the originally predicted energy of 28.9 kWh, which did not feature the degradation model. Table 5.2 shows the comparison and it is clear that the error in energy yield has been reduced substantially to only +4.1% (from +20.9%).

The method for degradation modelling was subsequently applied to winter data obtained from fresh modules (from the same manufacturing batch) tested in January to May 2018, as shown in the last row of Table 5.2. 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 yield shows that this approach can be used for predicting outdoor performance of OPV modules accounting for the degradation.

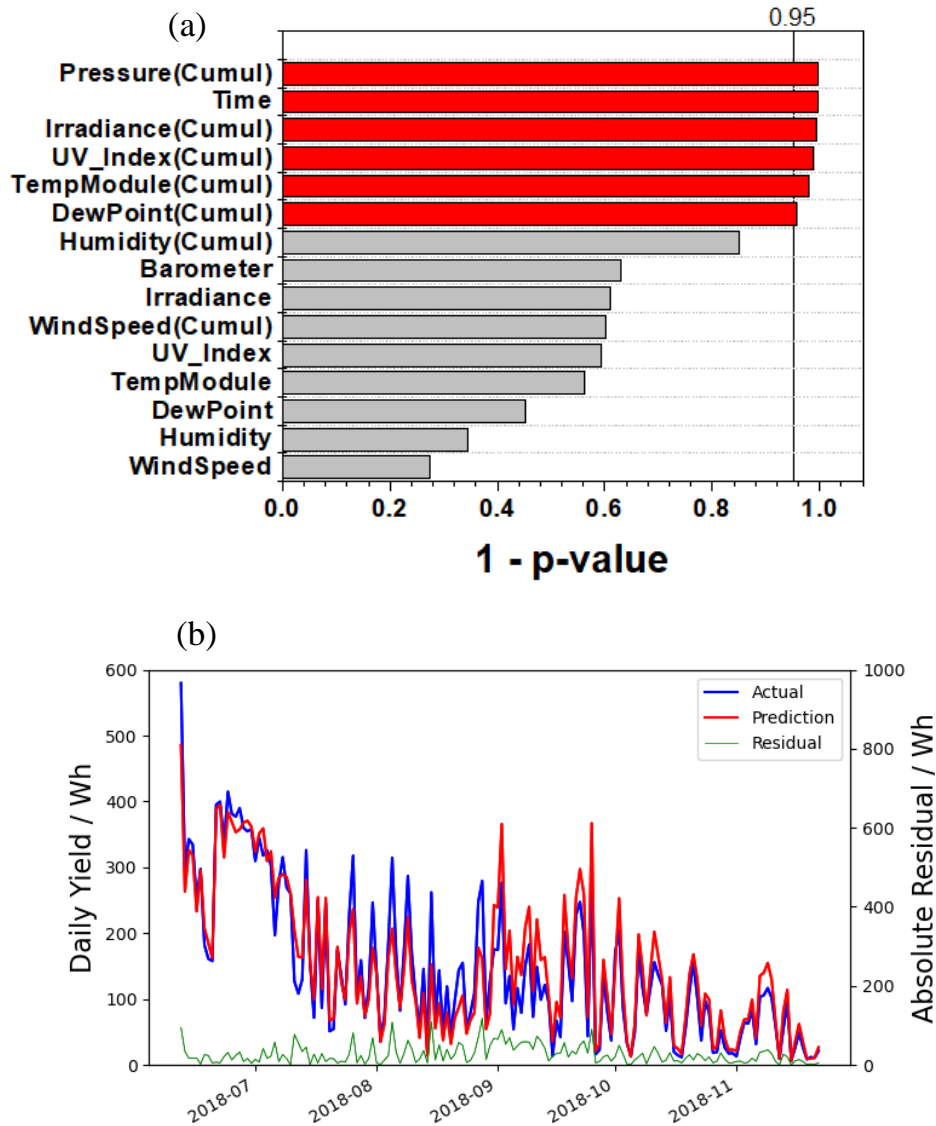


Figure 5.14: Pareto chart showing the significance of each instantaneous and cumulative weather condition, given a 5% confidence interval. For clarity, the statistically significant attributes are shaded in red and the non-significant attributes shaded in grey. The significance level being tested has been represented by the vertical straight line. (b) Variation in actual and predicted daily yield over the course of the 5-month test. The predicted daily yield has been acquired by considering both the prediction made by an MLP model as well as consideration of the degradation rate of the modules. The green line represents the residual between actual and predicted daily yields.

In order to gain a greater understanding of how the different weather conditions vary and how they are governed by each other, the covariance between the different conditions can be analysed. Table 5.3 illustrates the normalised covariance between time

and each of the cumulative conditions. These covariances are analysed since it will allow the conditions which are similar to the intrinsic time variation to be identified. From Table 5.3, it can be seen that the highest covariance is between the atmospheric pressure and time. This arises due to pressure only fluctuating by a small amount, relative to its absolute value, over the course of the test and is typically in the range 970 hPa to 1040 hPa. Other attributes such as outside temperature have a much lower covariance with time due to the large diurnal changes in temperature, relative to the absolute value. These relationships can also be qualitatively seen from inspection of Figure 5.12 where a high covariance leads to a linear variation of the cumulative condition. Therefore, including pressure would lead to some overfitting of the data. However, it was kept as a predictive attribute for completeness and such that its significance could be ranked in relation to the other factors. In addition, atmospheric pressure is a widely reported value and can be measured accurately for a wider area since the condition is not as localised as other conditions. The consequence of the high covariance, however, leads to a similar significance of time and cumulative pressure, as seen in Figure 5.14 (a).

Attributes	Normalised Covariance
Time/Cumulative Humidity	0.077
Time/Cumulative Outside Temperature	0.016
Time/Cumulative Wind Speed	0.0011
Time/Cumulative Pressure	0.99
Time/Cumulative Dew Point	0.012
Time /Cumulative UV Index	0.00083
Time / Cumulative Irradiance	0.16
Time/Cumulative Module Temperature	0.019

*Table 5.3: Covariances between time and each cumulative weather condition.*

Figure 5.15 (a) illustrates the forecasted degradation of OPV modules over the course of winter months. Figure 5.15 (b) shows the regression plot, illustrating the predicted degradation rates as a function of the reference degradation rate.

The forecasted degradation during winter months does not display burn – in during the early stages of testing as seen with summer months. This is as expected since

irradiance levels are lower during winter months. From Figure 5.15 (b), a good fitting is achieved, and it is noteworthy that the rapid, early degradation of approximately  $0.5\% \text{Day}^{-1}$  is accurately modelled during the winter test. The degradation model for winter is derived from fitting to the new data. This rapid degradation is not as high as the burn-in degradation observed during summer months, which proved more difficult to model.

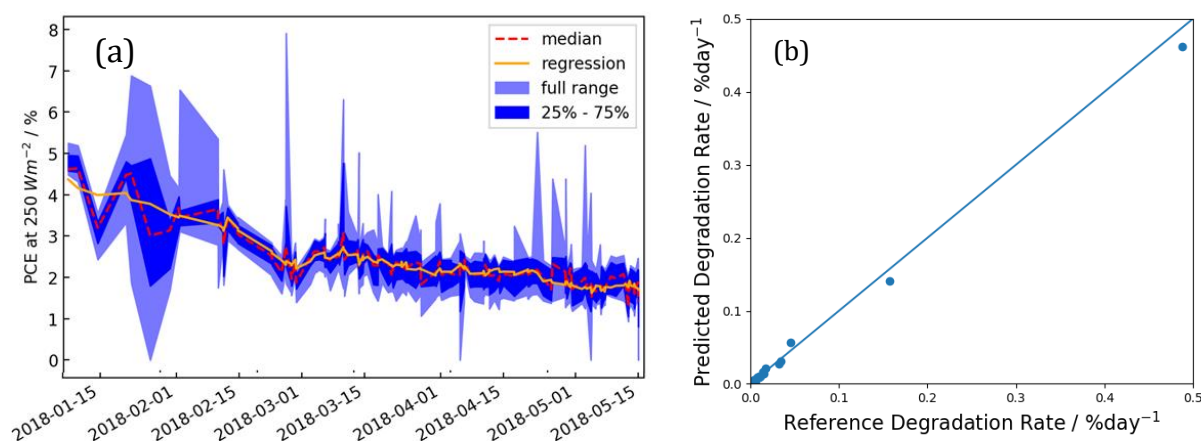


Figure 5.15: (a) Forecasted degradation of OPV modules at  $250 \text{ mWcm}^{-2}$  during winter months, (b) Predicted degradation rate as a function of the reference degradation rate for winter months.

A similar significance test is conducted to determine the significant factors governing the degradation during winter months. Figure 5.16 illustrates the Pareto chart showing the  $1 - p$ -value quantity. When  $1 - p$ -value is greater than 0.95, the attribute is statistically significant. For clarity, the statistically significant attributes have shaded in red and the non-significant attributes shaded in grey. The significance level being tested has been represented by the vertical straight line.

For the winter months, the significant factors governing the degradation are found to be similar to those determined for summer months. In rank order, these are found to be cumulative UV index, cumulative irradiance, humidity, temperature and dew point. However, time is no longer found to be significant and atmospheric pressure is now determined to be the least significant. This could be because atmospheric pressure is less variable during winter months in the UK as compared to summer months. It is also

notable for the winter significance analysis, that not all the significant factors are now cumulative weather conditions; the instantaneous humidity, temperature and dew point now play a significant role. This could be an artifact of the modelling method since these attributes show a greater range of variability during the winter months of testing and, therefore, introduce greater bias in the model.

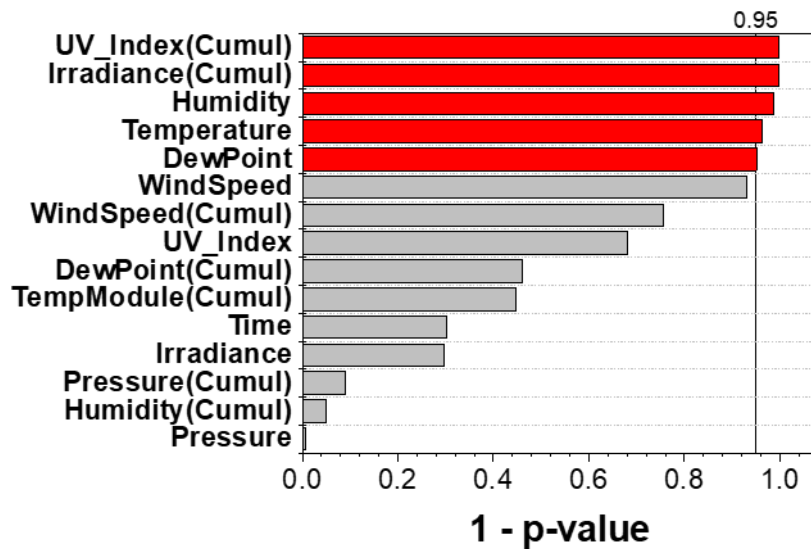


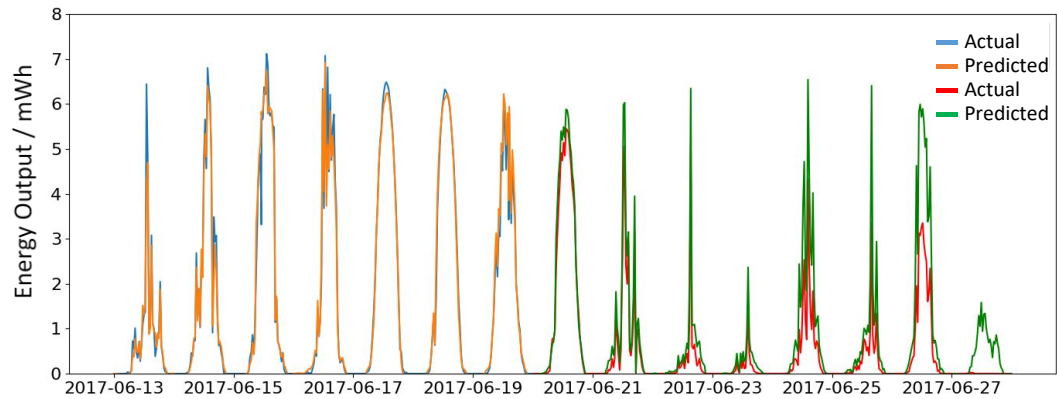
Figure 5.16: Pareto chart showing the significance of each instantaneous and cumulative weather condition, given a 5% confidence interval for winter months.

### 5.2.6 Degradation Forecasting of Other Technologies – Perovskite Solar Cells

As a proof of concept, the forecasting approach is further applied to different emerging technologies, in this case Perovskite solar cell modules, tested in outdoor conditions by Stoichkov *et al.* [179]. Two modules were considered for forecasting the diurnal performance: one using FMC as the active material and one using MAPbI<sub>3</sub> as the active material. This dataset contains fewer records than the OPV outdoor dataset and has, therefore, a significantly shorter test duration; however, it is large enough to perform meaningful analysis. Since the FMC module displayed greater long – term stability in outdoor conditions, this module is used to forecast the diurnal variation. Figure 5.17

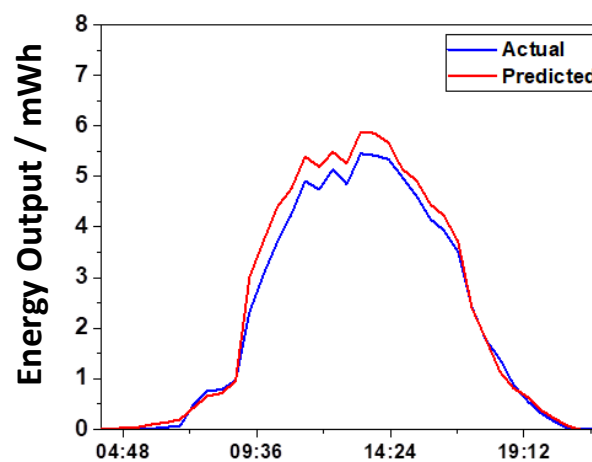


illustrates the actual and predicted diurnal variation in energy output where 1 week has been used for training and one week used for testing.

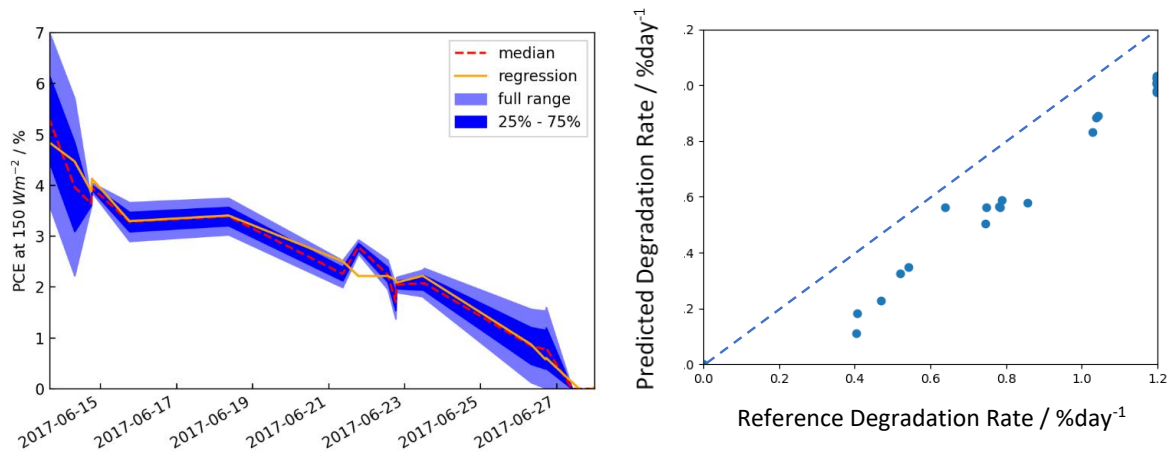


*Figure 5.17: Variation in energy output with one week of training and 1 week of testing of PSC modules in outdoor conditions. Blue is the actual data during the training set, orange is the predicted variation during the training set, red is the actual data during the testing set and green is the predicted variation during the testing set for verification of the model.*

For clarity, the diurnal variation in the energy output for the first day of the testing set is shown in Figure 5.18. A good fitting is achieved for the predicted diurnal variation in energy output for the FMC – based module.



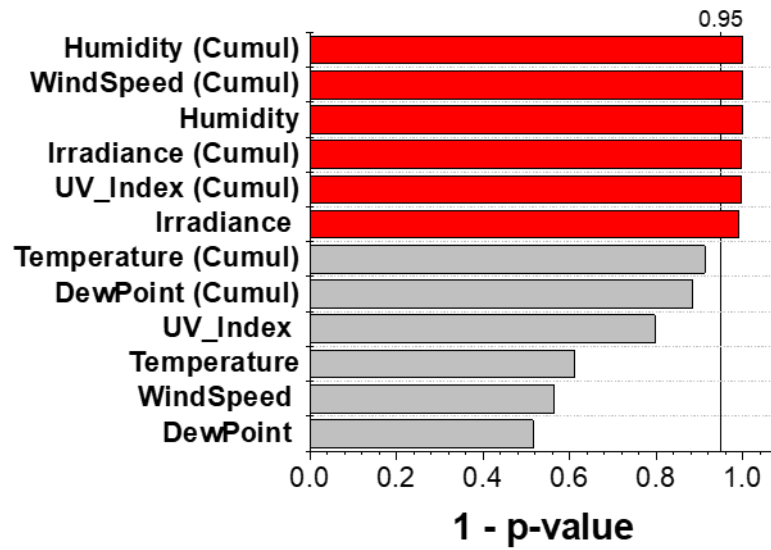
*Figure 5.18: Diurnal variation in energy output for first day in testing set.*



*Figure 5.19: (a) Forecasted degradation of PSC modules at 150 mWcm<sup>-2</sup>, showing the MLR regression prediction. (b) Degradation rate regression, showing the predicted degradation rate as a function of the reference (measured) degradation rate for PSC modules. A CC of 0.97 is achieved.*

Subsequently, degradation analysis is performed for the outdoor tested PSC modules, again using a MLR modelling approach and performed on both the FMC and MAPbI<sub>3</sub> modules. Figure 5.19 (a) illustrates the variation in PCE at 150 Wm<sup>-2</sup> over the course of two weeks. The PSC modules display rapid degradation over the course of two weeks with little or no stabilised degradation. Figure 5.19 (b) illustrates the regression results for the degradation rate. A good fitting is observed for the MLR model with a CC of 0.97; however, the predicted degradation rate was underestimated for all reference degradation rates.

The significant attributes governing the PSC module degradation are determined using hypothesis testing and  $p$  - value analysis. The pareto chart is shown in Figure 5.20. The significance analysis for the PSC module degradation shows that the significant attributes are cumulative humidity, cumulative wind speed, instantaneous humidity, cumulative irradiance, cumulative UV index and instantaneous irradiance. However, it is not possible to reach firm conclusions from this analysis since the devices only lasted for two weeks. Nevertheless, the principle could be applied to better performing devices in the future.



*Figure 5.20: Pareto chart showing the significance of each instantaneous and cumulative weather condition, given a 5% confidence interval for PSC modules.*

It is well known that PSC devices and modules suffer from light and water induced degradation. Therefore, the results from the significance analysis provides reassurance in the application of this approach in modelling the degradation of different solar cell modules in outdoor conditions. Furthermore, the prominence of wind speed as a significant attribute could allude to the susceptibility of PSCs to mechanical stress degrading the material or could suggest the increased ingress of water due to the increased wind speed.

Subsequently, the degradation of both the FMC and MAPbI<sub>3</sub> modules are analysed in terms of their diurnal hysteresis. The performance parameters are extracted for morning and evening periods at an irradiance of 250 Wm<sup>-2</sup>. The variation in both the morning and evening performances are then tracked over the course of two weeks. Figure 5.21 (a) – (d) shows the variation in each performance parameter for the FMC module whilst Figure 5.22 (a) – (d) shows the corresponding plots for the MAPbI<sub>3</sub> module.

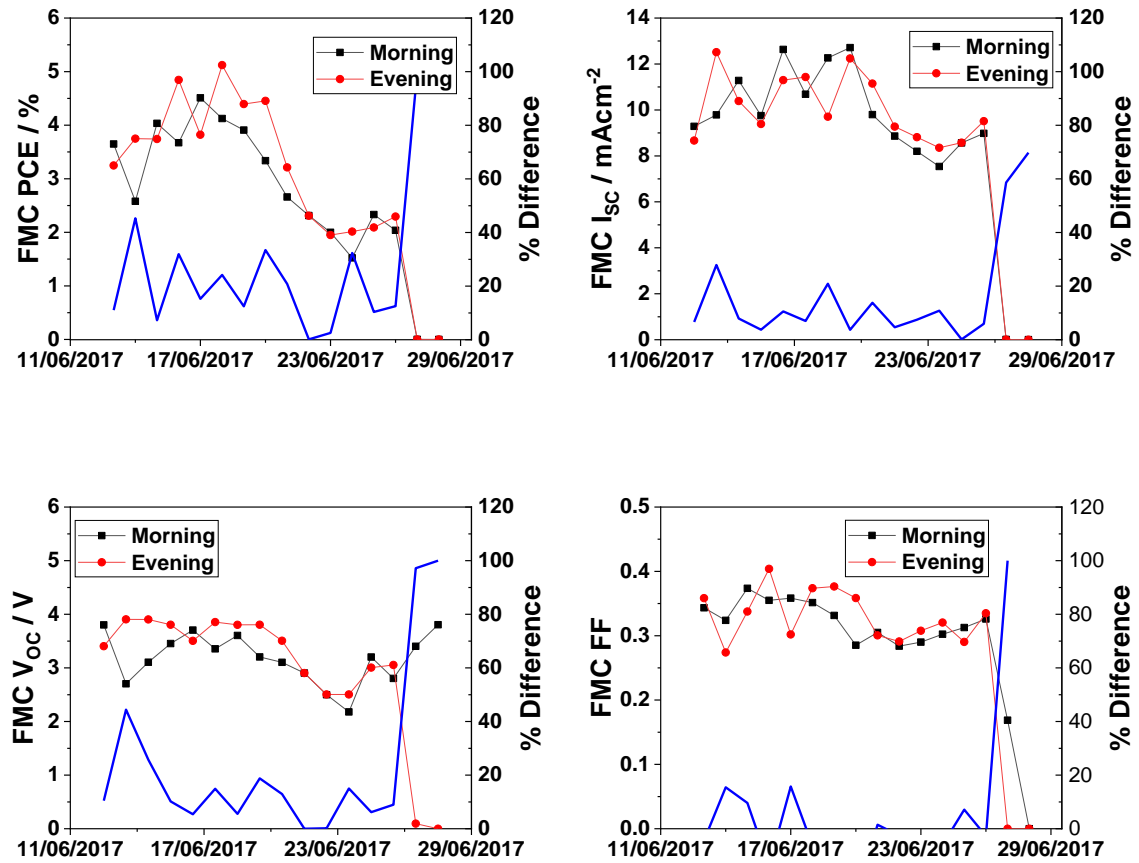


Figure 5.21: Variation in (a) PCE, (b)  $I_{sc}$ , (c)  $V_{oc}$  and (e) FF for the FMC based module over two weeks of testing. Percentage difference between morning and evening is shown in blue.

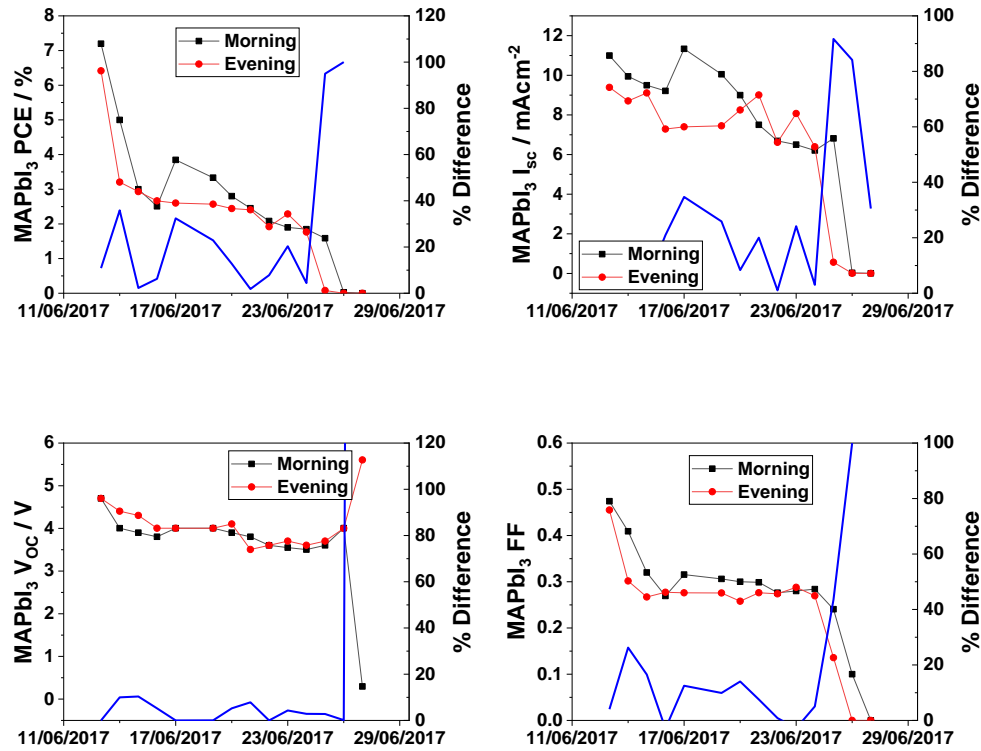


Figure 5.22: Variation in (a) PCE, (b)  $I_{sc}$ , (c)  $V_{oc}$  and (e) FF for the MAPbI<sub>3</sub> based module over two weeks of testing. Percentage difference between morning and evening is shown in blue.

A similar outdoor test was performed by Katz *et al.*, with devices having the architecture: ITO/SnO<sub>2</sub>/Cs<sub>0.05</sub>((CH<sub>3</sub>NH<sub>3</sub>)<sub>0.15</sub>(CH<sub>3</sub>NH<sub>2</sub>)<sub>0.85</sub>)<sub>0.95</sub>PbI<sub>2.55</sub>Br<sub>0.45</sub>/spiro-OMeTAD/Au [180]. The performance of the devices was compared between morning and evening. Two types of degradation were attributed to the device performance variation: 1) decrease in PSC performance parameters partially compensated by night-time recovery and 2) PCE decreased during night time and increased during the day time. Type 1 degradation was observed during early stages of testing whilst type 2 degradation was observed beyond 8 days of testing. The morning/evening variation shown in Figure 5.21 for FMC devices displays type 2 degradation from the outset whilst the MAPbI<sub>3</sub> device display type 1 degradation after 4 days of testing, with an initial rapid burn – in prior to 4 days. Katz *et al.* also note that the decrease in performance during the day is principally governed by  $V_{oc}$  and FF and this behaviour is also observed here.

## 5.3 Conclusion

Machine learning approaches have been applied to Outdoor datasets of OPV and PSC performance data and weather conditions to analyse the effects of different weather conditions on the performance and stability of the 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 forecast the diurnal variation in the OPV performance parameters, where the weather conditions have been used as the attributes for prediction. By deriving the power output of the modules and, therefore, the daily yield, this methodology has successfully been applied to forecast the daily yield of the OPV modules over the course of 6 months with a 20.9% error based on only two weeks of training data.

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 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 forecast for a period of 6 months with high accuracy, and the variation in degradation rate over the course of 6 months also forecasted.

Using the model which accounts for degradation, energy yield was predicted to within 4.1% error. Further tests were conducted on unseen winter 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.

The forecasting methodology is subsequently applied to a dataset representing outdoor tested PSC modules. The degradation is successfully forecasted, and the significant attributes determined. However, this dataset is substantially smaller than the outdoor tested OPV modules. Therefore, more outdoor tested PSC module data is still required in order to determine the validity of this model. This method could easily be made applicable to a wide range of technologies and used by industries and researchers alike to substantially reduce testing times for OPVs in outdoor conditions.

## Chapter 6. Life – Cycle Assessment of Organic Photovoltaics with the application of Genetic Clustering Algorithms.

### 6.1 Introduction

Organic photovoltaic (OPV) technologies are often quoted as being cheaper and easier to manufacture in comparison to their silicon counterparts [181]–[184]. However, one of their main benefits is their ‘energy cost’, which are often noted as being lower than c-Si and other PV technologies. This can be quantified using the ‘embodied energy’ which is defined as the energy consumed in the manufacture of the particular product, in this case the OPV. Embodied energy is dependent on the production of the constituent materials and their processing. As seen in chapter 4, the material selection also has an impact on performance and stability. Therefore, when selecting the optimal materials for an OPV, there are three quantifiable figures of merit to consider: initial efficiency (PCE), stability and the embodied energy. From the literature it would seem that an optimisation of all of these parameters simultaneously has not been performed before.

As a continuation of the previous chapter’s work, the database of OPV structures and matching test data has been extended to include the embodied energies of the solar cell. In this work, only data acquired with ISOS – L standards has been included, as this allows for energy yield to be calculated based on the  $T_{80}$  and  $T_{S80}$  lifetimes. Using the initial PCE, stability and embodied energy of the devices, the net energy produced by the OPV can be determined. Determining the optimum combination of materials can be achieved by adopting machine learning (ML) approaches with the optimisation time significantly reduced. This chapter builds on the results of chapter 4 by extending the modelling of PCE and  $T_{80}$  lifetime using SMOreg to include the energy cost of manufacture and the net energy production during operation. Subsequently, this provides an analysis simultaneously optimising PCE, lifetime, energy costs of manufacture and net energy output. In addition, the results of chapter 4 which provided a ranking of the best materials for performance and stability enhancement have been extended to take into

consideration the optimisation of combining different materials to form an entire device stack. The calculation of net energy output was inspired from the results of chapter 5 where the degradation and daily yield of OPV modules is forecasted, allowing the energy yield of modules to be predicted over the course of their lifetimes. The goal of this chapter is to find new combinations of materials which have a lower environmental footprint whilst maximising the energy output.

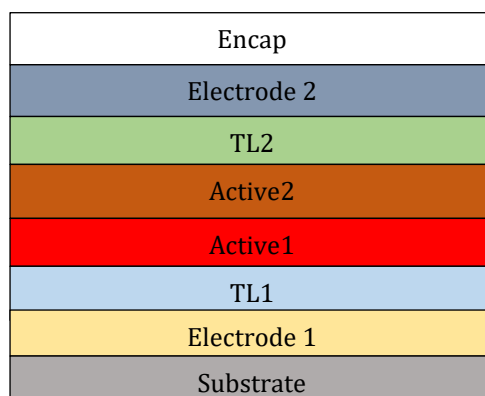
Herein, two principal ML techniques are utilised in order to analyse the net energy output of the OPV devices. Firstly, sequential minimal optimisation regression (SMOreg) is employed using a similar method as in chapter 4 [185]. Here the materials governing the net energy output can be analysed and their relative significance quantified. Secondly, a genetic clustering algorithm is used to find the optimum architecture leading to maximised net energy output. Genetic search methods are used in optimisation problems and is a search heuristic inspired by natural evolution and uses the principal of survival of the fittest relying on mutation, crossover and fitness selection [186].

## 6.2 Methodology

### 6.2.1 Material Embodied Energy Determination

Much work has already been conducted in determining the embodied energy of different OPV materials. For example, [116], [181], [187]–[189] to name but a few examples. Therein, life – cycle assessment (LCA) methods have been used to analyse the cradle – to – gate energy production costs for OPVs and their constituent materials. Cradle – to – gate refers to the manufacturing process from raw material extraction to final finished product, before transportation to the consumer. An LCA methodology considers the total environmental cost of producing different products and materials. This includes the raw material extraction, the production of the material, the usage and finally the disposal of the material. The device architecture considered is shown in Figure 6.1, where each device layer considered is illustrated. These are the substrate, the bottom electrode (Electrode 1), the first transport layer (TL1), the donor and acceptor active materials (Active1 and Active2), the second transport layer (TL2), the top electrode (Electrode 2) and the encapsulation (Encap).





*Figure 6.1: Device architecture considered showing each layer.*

A schematic of the method employed is shown in Figure 6.2, illustrating how the material and energy costs are acquired and the stages of analysis using ML and genetic algorithms.

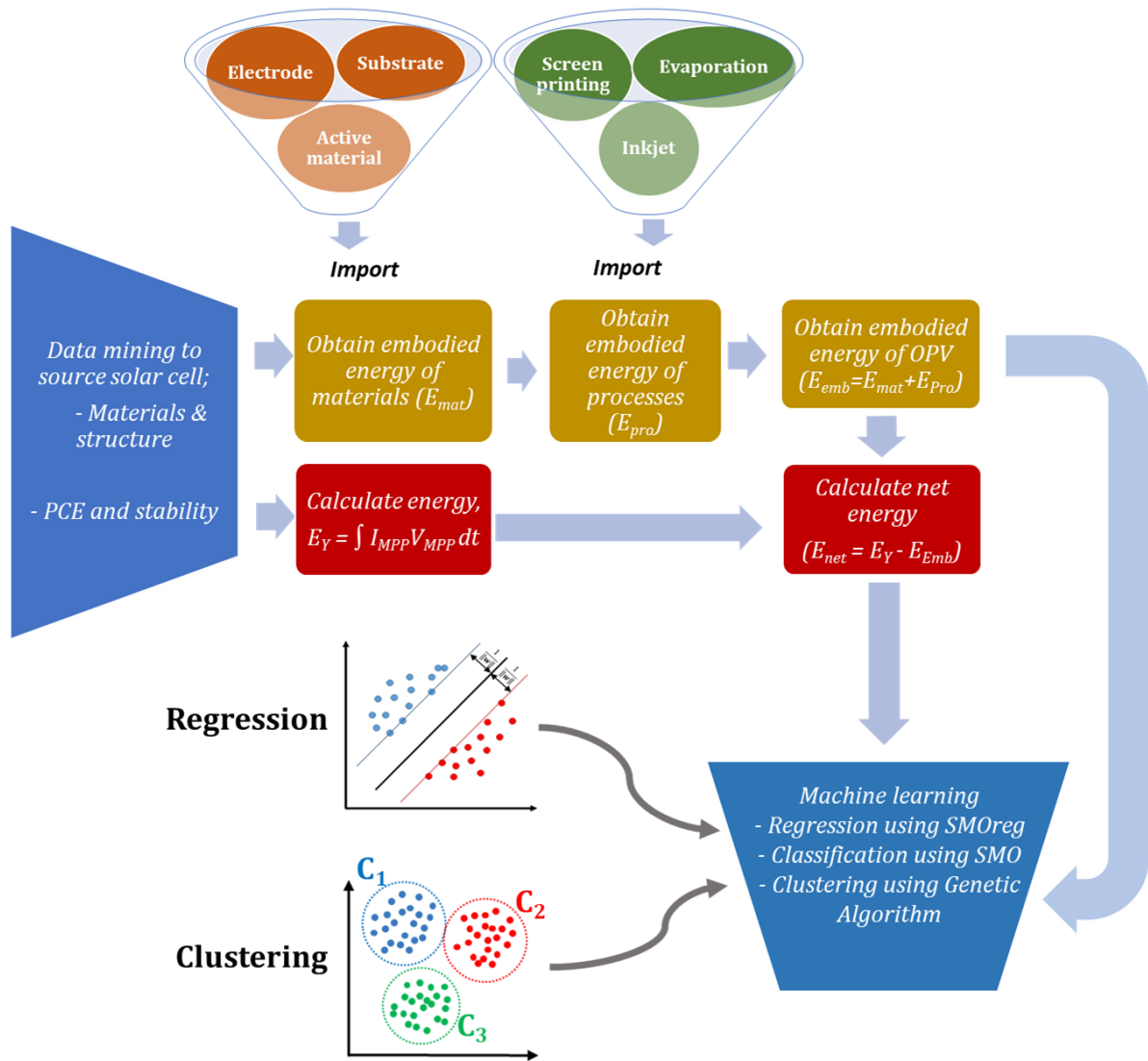


Figure 6.2: Schematic of how material and energy costs are acquired and stages of analysis using ML genetic algorithms.

Using pre-calculated LCA values, the database of OPV structure, performance parameters and embodied energy is subsequently sourced from the literature for all materials which could be found. It is estimated that 70% of data was sourced from other papers, however a small subset of data required embodied energy calculation, which was calculated with Simapro using the same approaches pioneered by Krebs [116], [181], [187], [188]. Where possible, an average of three embodied energy calculations from the literature was taken. One of the main challenges related to inconsistent units or inconsistent scales of productions. Therefore, assumptions were made in order to form a

consistent unit system. For example, many sources quote the embodied energy as either  $\text{MJkg}^{-1}$  or  $\text{MJm}^{-2}$ . The values quoted in  $\text{MJkg}^{-1}$  were converted to  $\text{MJm}^{-2}$  by considering the volume and density of the materials in question; the thickness and active areas are often given, but when unknown a common volume was considered.

Another major challenge is related to the variety of polymers used by the OPV community, many of which have no reports in the literature for their embodied energy. In these circumstances, estimations were made for the embodied energies. For organic materials, the number of synthesis steps was considered. Stiebitz produced a technical report where the embodied energy was considered as a function of the number of synthesis steps [190]. In the study by Stiebitz, three different polymers were studied: P3HT, PTB7 and PCDTBT and the embodied energy determined as a function of the number of synthesis steps. There are 5, 12 and 9 synthesis steps for each polymer, respectively. From this the average embodied energy per step is calculated and used to determine the embodied energies for the different polymers based on the number of synthesis steps for each material. Figure 6.3 illustrates the fitted plot for the embodied energy per mole of polymer as a function of synthesis steps. In addition, the embodied energy per molar mass was determined from several similar materials found in the literature, which were not present in the database, and subsequently converted based on the molar mass of the material and the quantity used. Using this regression model, estimation of the embodied energy of newer polymers and acceptor materials could be made.

In addition, the deposition process of each material is considered and values for the different process are incorporated into the embodied energy value when only the material embodied energy is given in the literature.

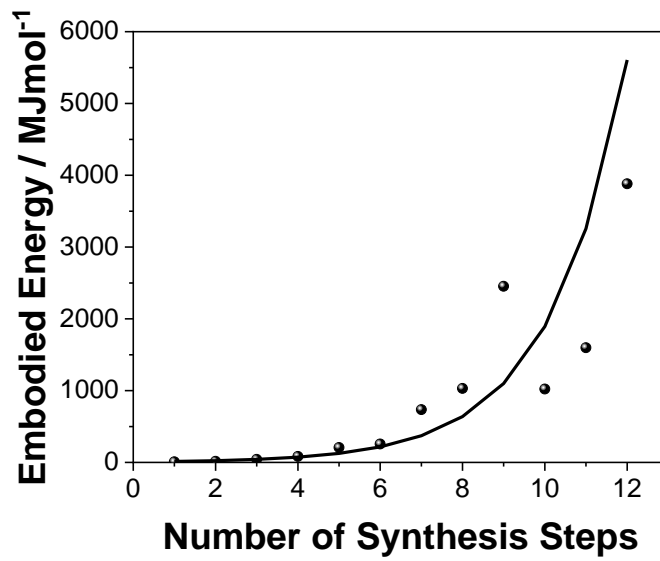


Figure 6.3: Variation in polymer embodied energy as a function of the number of synthesis steps.

Once the embodied energy of each individual material is found or calculated, the total embodied energy for each device in the data set can be calculated using equation 6.1,

$$E_{Emb} = \sum (E_{Mat} + E_{Pro}), \quad 6.1$$

where  $E_{Emb}$  is the total embodied energy of the OPV device including the embodied energy of the materials,  $E_{Mat}$  and the embodied energy of the processing for each material,  $E_{Pro}$ . The embodied energy of the materials can be defined as,

$$E_{Mat} = E_{Sub} + E_{Elec1} + E_{TL1} + E_{Active1} + E_{Active2} + E_{TL2} + E_{Elec2} + E_{Encap} \quad 6.2$$

where  $E_{Sub}$ ,  $E_{Elec1}$ ,  $E_{TL1}$ ,  $E_{Active1}$ ,  $E_{Active2}$ ,  $E_{TL2}$ ,  $E_{Elec2}$  and  $E_{Encap}$  are the embodied energies of each constituent material in each layer (Substrate, Electrode 1, Transport Layer 1, Active 1, Active 2, Transport Layer 2, Electrode 2 and Encapsulation respectively).  $E_{Pro}$  incorporates the energy required for processing each material, for example spin coating, evaporation, screen printing and slot die. An approximate

exponential decay curve, for the maximum power point, is fitted between  $T_0$  (time at  $E_0$ ) and  $T_{80}$ , and a linear degradation between  $T_{80}$  and  $T_{S80}$ . Subsequently,  $NE$  is calculated by integrating the functional power output,  $P(t) = I_{MPP}(t)V_{MPP}(t) = P_0 e^{-\alpha t}$ , over time,

$$NE = \left( \int_0^{T_{80}} P_0 e^{-\alpha_1 t} dt + \int_{T_{80}}^{T_{S80}} P_0 e^{-\alpha_2 t} dt \right) - (E_{Emb}), \quad 6.3$$

where  $I_{MPP}(t)V_{MPP}(t)$  is the time – variation of the power output between time  $t = 0$  and  $T_{S80}$ ,  $P_0$  is the power at  $t = 0$  and  $\alpha_1$  and  $\alpha_2$  are fitted exponential decay constant. The numerical value for  $\alpha_1$  is in the range  $[6 \times 10^{-5}, 5.6]$  and  $\alpha_2$  is in the range  $[7 \times 10^{-5}, 0.092]$ . Since the exponential decay curve is being approximated, the method for determining the  $NE$  will possess some error. A full list of the embodied energy used for each material and process is given in Appendix B.

### 6.2.2 Genetic Algorithm Clustering

In addition to the SMOreg algorithm, genetic algorithm (GA) clustering is employed as a means of determining the optimum material sets for maximising the net energy output of OPVs. The following discussion of Genetic Algorithm Clustering closely follows the theory presented by Maulik *et al.* in [191], which provides an in-depth discussion on the use of genetic algorithms in clustering problems.

GAs are randomised search and optimization techniques inspired by the principals of evolution and natural selection [192]. GA methods are used to search through large, multidimensional spaces in order to find optimal solutions for an objective function. When using GAs, the parameters of the search space are encoded as strings, called chromosomes and a collection of strings is termed a population. A random population is initially created, representing different points in the search space. Objective and fitness functions are associated with each string which represents the “degree of goodness” of the string [191]. Following the principle of “survival of the fittest”, a subset of these strings is used and assigned a number of copies that are passed into a mating pool. New generations of strings are then produced using biologically inspired operators

such as crossover and mutation. The three stages of selection, crossover and mutation continue for a fixed number of generations when the optimum function is achieved.

Clustering in an  $N$ -dimensional space represents the process of partitioning a set of  $n$  points into  $K$  clusters,  $C_1, C_2, \dots, C_k$ , based on a similarity or dissimilarity metric. The steps implemented during GA clustering can be summarised as follows:

1. *String representation*

A string is a sequence of real numbers representing the  $K$  cluster centres (centroids). For the search space, the length of a string will be  $N \times K$ . Here, the first  $N$  positions (genes) represent the  $N$  dimensions of the first cluster centre, the next  $N$  positions represent the second cluster centre and so on [191].

2. *Population initialization*

The  $K$  cluster centres encoded in each chromosome are randomly initialised to  $K$  random points from the data set and repeated for each of the  $P$  chromosomes in the population, where  $P$  is the size of the population [191].

3. *Fitness computation*

The fitness computation process consists of two process. Firstly, the clusters are formed dependent on the centres encoded in the chromosomes. This is achieved by assigning each point,  $\mathbf{x}_i$ ,  $i = 1, 2, \dots, n$ , to the centroids  $\mathbf{z}_j$ , such that,

$$\|\mathbf{x}_i - \mathbf{z}_j\| < \|\mathbf{x}_i - \mathbf{z}_p\|, \quad p = 1, 2, \dots, K, \text{ and } p \neq j. \quad 6.4$$

Secondly, after the clustering has been performed, the centroids, encoded in the chromosomes are replaced by the mean points of the respective clusters; for cluster  $C_i$ , the new centre  $\mathbf{z}_i^*$  is computed as

$$\mathbf{z}_i^* = \frac{1}{n_i} \sum_{x_j \in C_i} x_j \quad i = 1, 2, \dots, K. \quad 6.5$$

Each  $\mathbf{z}_i^*$  replaces the previous  $\mathbf{z}_i$  in the chromosome [191].

#### 4. *Selection*

During the selection process, chromosomes from the mating pool are selected based on the “survival of the fittest” concept. A chromosome is assigned a number of copies proportional to its fitness in the population and are subsequently passed back into the mating pool for genetic operations [191].

#### 5. *Crossover*

The crossover is a probabilistic process where information is exchanged between two parent chromosomes to generate two child chromosomes. For chromosomes of length  $l$ , a random integer, called the crossover point, is generated in the range  $[1, l-1]$ . The portions to the right of the crossover point are exchanged to produce two offspring [191].

#### 6. *Mutation*

Each chromosome undergoes mutation with a fixed probability  $\mu_m$ . A bit position (gene) is mutated by flipping its value. A number,  $\delta$ , is generated in the range  $[0,1]$  with uniform distribution. For a value at a gene position,  $v$ , after mutation the value becomes,

$$v \pm 2 * \delta * v, \quad v \neq 0$$

$$v \pm 2 * \delta, \quad v = 0$$

The positive (+) and negative (-) cases occur with equal probability [191].

#### 7. *Termination Criterion*

The processes of fitness computation, selection, crossover and mutation are executed for a maximum number of iterations. The best chromosome observed up to the last generation provides the solution to the clustering task [191].

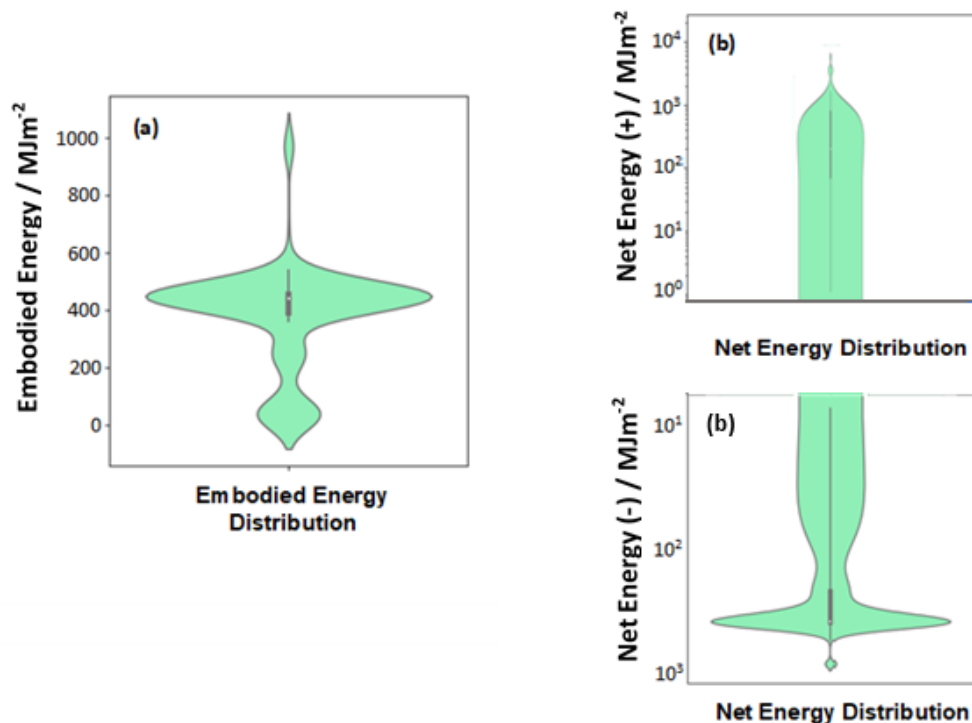
Each cluster is determined based on a similarity metric which uses the Manhattan Distance Class, where the distance is determined based on a square grid formation and the shortest path is determined from the grid lines. For categorical values, the strings are compared by considering the number of bit positions in which the two strings differ.

When applying the GA algorithm, GenClust++, the optimum material for each layer is determined, whereas when using the SMOreg ML algorithm, the best attributes for enhancing the net energy are found; these could all belong to a single attribute class (for example the electrode) and not take into consideration the optimisation for combining each device layer together. Therefore, even if each device layer is determined by considering the full list of weights derived using SMOreg, the identified device architecture would not be an optimised solution, as in the case of using the GenClust++ GA algorithm.

## 6.3 Results and Discussion

### 6.3.1 Initial Data Exploration

The data was initially visualised in several ways in order to better understand the data structure. Figure 6.4 shows the categorical violin plots for (a) Embodied Energy and (b) Net Energy of the devices in the dataset.



*Figure 6.4: Categorical distribution for (a) embodied energy and (b) net energy for the dataset. The negative net energies are shown in the lower plot and positive net energies are shown in upper plot.*



These plots intuitively show the median, interquartile range and  $1.5 \times$  interquartile range. In addition, the plot shows the kernel density estimation to illustrate the distribution of each parameter. For the Net Energy, the data was split for positive and negative values and plotted on separate axes to allow for a logarithmic scale to be used. The negative net energies are shown in the lower half and positive net energies are shown in the upper half.

In addition, Figure 6.5 illustrates scatter graphs between (a) embodied energy and PCE, (b) net energy and PCE, (c) embodied energy and  $T_{80}$  and (d) net energy and  $T_{80}$ . Little or no correlation can be observed for (a), (b) and (c), however, (d) does show a positive correlation indicating that as  $T_{80}$  increase, so does the net energy of the device; this is as expected.

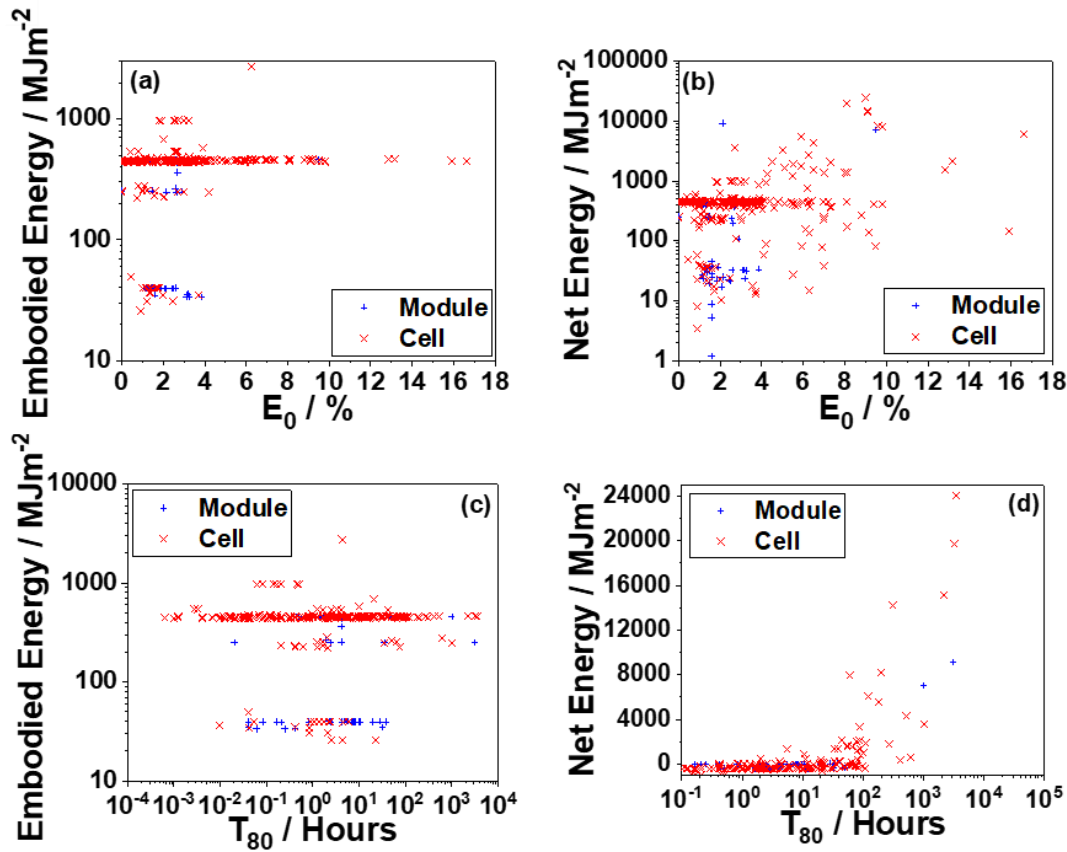


Figure 6.5: Distributions of (a) embodied energy as a function of  $E_0$ , (b) net energy as a function of  $E_0$ , (c) embodied energy as a function of  $T_{80}$  and (d) net energy as a function of  $T_{80}$ . For net energy, the absolute value is plotted in order to use a logarithmic scale and for clarity.

### 6.3.2 Classification using SMO algorithm

In the first instance, the dataset embodied and net energies were classified such that there are approximately 20 instances in each class and each class is assigned a categorical variable (A, B, C,...). The SMO classifier is then applied in order to predict the embodied and net energy classes based on the structural components. This will allow the performance of the algorithm to be tested as well as verify the integrity of the dataset. Table 6.1 shows the classifier statistics for both embodied energy and net energy.

This can be visualised using confusion matrices as shown in Table 6.2 for embodied energy and Table 6.3 for net energy. A confusion matrix shows how many instances have been assigned to the correct class where each column and row represent the classes. Therefore, for a classifier which operates effectively, all instances should lie on the diagonal and represent correctly classified instances whilst off-diagonal instances represent incorrectly classified instances and contribute to the ‘confusion’ of the algorithm. A correct classification occurs when the algorithm assigns a record to the same class as the label for that record. From Table 6.2, it is apparent that the SMO classifier effectively classifies the dataset with a 96.7% accuracy. For most incorrectly classified cases, for example classes B, C, and G, a small degree of confusion is observed with the algorithm assigning some cases to adjacent classes. There is one instance, however, where the classifier assigns an instance in class E to class R. Nevertheless, the results for embodied energy demonstrates that the SMO classifier is able to effectively understand the structure of the dataset. This gives us confidence that new device structures can be predicted where one can have confidence in the results.

	Correctly classified	Incorrectly classified	Kappa statistic	MAE	RMSE	RAE	RRSE
<b>Embodied Energy</b>	96.7%	3.31%	0.965	0.0988	0.218	94.1%	95.1%
<b>Net Energy</b>	77.7%	22.3%	0.764	0.0991	0.219	94.4%	95.4%

*Table 6.1: SMO classifier statistics for embodied energy and net energy.*

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
A	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	17	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	1	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	2	18	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	1	0	18	1	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	1	19	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	1	19	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	1	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0
R	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	22

Table 6.2: Confusion matrix for embodied energy using SMO classifier.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
A	12	3	4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
B	0	14	1	1	0	0	3	0	0	0	0	0	1	0	0	0	0	0
C	1	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	1	3	12	0	0	2	0	0	1	0	0	1	0	0	0	0	0
E	0	3	0	1	15	0	1	0	0	0	0	0	0	0	0	0	0	0
F	0	1	0	1	0	20	1	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	1	17	0	0	0	0	0	0	0	0	0	1	1
H	0	0	0	0	0	0	0	16	0	1	0	0	0	2	0	0	0	1
I	0	0	0	0	0	0	1	0	11	2	3	0	0	2	0	0	1	0
J	0	0	0	0	0	0	0	0	0	14	4	0	0	1	1	0	0	0
K	0	0	0	0	0	0	0	0	0	0	19	1	0	0	0	0	0	0
L	0	0	0	0	0	0	1	0	2	1	3	12	0	1	0	0	0	0
M	0	0	0	0	0	0	2	0	1	0	0	0	16	1	0	0	0	0
N	0	0	0	0	0	2	0	0	0	0	0	0	1	15	1	1	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	1	16	3	0	0
P	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	17	1	0
Q	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	17	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20

Table 6.3: Confusion matrix for net energy using SMO classifier.

The confusion matrix for net energy (Table 6.3) displays a good classification accuracy of 77.7%. However, significantly more confusion is observed with multiple occurrences of incorrect classification. Furthermore, a broader distribution of incorrectly classified instances is seen with multiple non-adjacent classifications. For example, instances in class A are assigned to classes B, C, E, and instances in class L are assigned to G, I, J, K, N. However, an increased confusion is expected for the classification of net energy since it represents a derived quantity dependent on uncertainty in the PCE and stability calculations.

### 6.3.3 Application of SMOreg algorithm

Subsequently, the SMOreg ML algorithm has been applied to the dataset and the structural components of the OPVs are used to predict the embodied energy of new OPV structures. This provides additional information over the classification approach since it allows the attributes governing the embodied energy to be ranked in terms of their significance on the environmental footprint of the OPV. The analysis was performed using cross – validation approach to ensure no overfitting of the data. Three regimes were chosen for the analysis: (a) using all structural components as attributes, (b) excluding the first electrode (Electrode1) and (c) analysing only the active layer influence. The regression results are shown in Figure 6.6 (a)-(c). The fitting parameters for the three subsets are shown in Table 6.4. From both Figure 6.5 and Figure 6.6 (a) – (c), it can be seen that many devices display similar embodied energies. This is due to many devices being fabricated from similar materials. For example, 42% of devices use a P3HT:PCBM active layer and 79% of devices use an ITO bottom electrode. This means the analysed devices will only display small variations in embodied energies, due to different materials used for other layers. This implies that the SMOreg model will be primarily fitting to a small range of embodied energy values with smaller groups of data instances observed over a larger range. Consequently, this could lead to some overfitting of the data. The cross – validation approach was adopted such that overfitting was minimised. Inspection of Figure 6.6 (a) demonstrates that the instances having total embodied energies different to the main cluster have been accurately predicted and do not represent outliers. Therefore, despite the model principally fitting to the main central cluster, representing devices of similar architecture, the model is still capable of predicting other

device embodied energies. However, the effect of many P3HT:PCBM active layer devices can be seen in Figure 6.6 (c) where there is a central cluster of instances (corresponding to P3HT:PCBM devices) and other devices not being predicted as accurately. This is due to less attributes being employed such that the model cannot account for the variation in active material used.

In all cases, high CCs of 0.988, 0.947 and 0.724 were found for the three subsets of data respectively. This is not surprising as the embodied energy is based on multiple calculations from the literature and can be accurately calculated.

The weights obtained from the SMOreg model can be analysed and ranked as the ten most beneficial and ten most detrimental for decreasing the embodied energy, as shown in Table 6.5 for the three subsets. For the beneficial weights, a more negative value corresponds to a larger decrease in embodied energy, whilst for the detrimental weights, a more positive value corresponds to a larger increase in embodied energy.

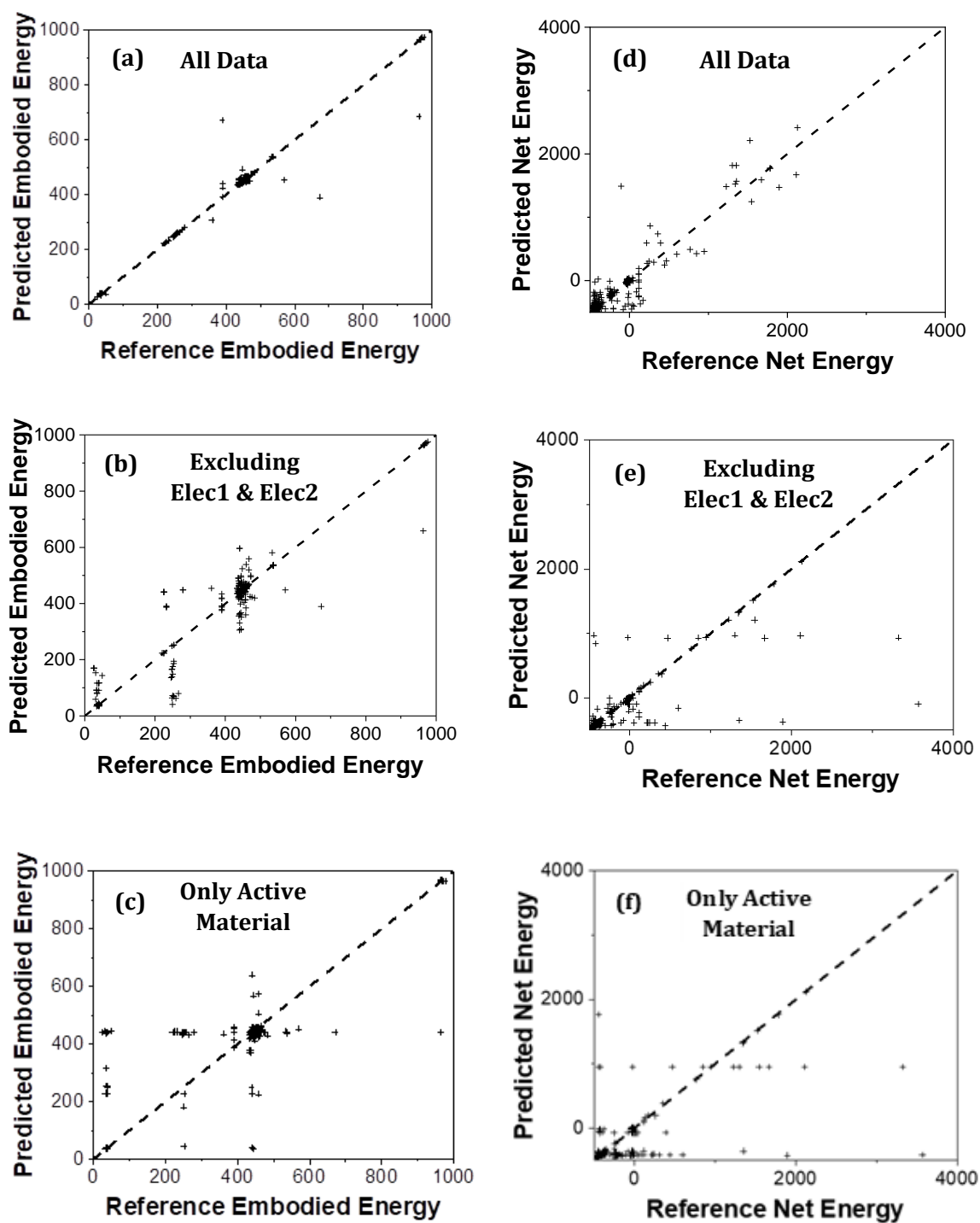


Figure 6.6: Predicted Embodied Energy and net energy as a function of reference Embodied Energy and net energy using SMOreg algorithm for (a, d) all structural components, (b, e) excluding Electrode 1,2 and (c, f) using only active layer components.

Inspection of the weights for all structural components show that the most beneficial attributes are substrate and electrode 1 materials with PET being the most beneficial. The most detrimental attribute is to use a quartz substrate. Furthermore, silver is identified as being beneficial for lowering the embodied energy whilst using composite electrodes such as ITO|Ag|ITO and Cr|Aluminium are detrimental. Considering TL1, only PEDOT:PSS is identified as being beneficial whilst  $\text{CrO}_2$ ,  $\text{TiO}_2$  and  $\text{NiO}_x$  are found to be detrimental. This is due to the high processing costs of metal oxide materials. For the active layer, only a few materials are identified, thus making any conclusions about these difficult. The same is true for TL2, however, ZnO nanoparticles are found to be beneficial whilst BPhen: $\text{MoO}_3$  is found to be detrimental. Electrode 2 is not identified as having any significant effect on the embodied energy. Many of the identified attributes correspond to electrode 1 for both the positive and negative attributes. This provides the justification for subsequently reapplying the algorithm without including electrode 1, as well as electrode 2 as attributes.

<b>Metric</b>	<b>All attributes</b>	<b>Excluding Electrode1,2</b>	<b>Only Active layer components</b>
Correlation coefficient	0.988	0.947	0.724
Mean Absolute Error	5.19	26.1	60.8
Root Mean Squared Error	27.2	57.1	125.6
Relative Absolute Error	4.04%	20.3%	47.5%
Root Relative Squared Error	15.3%	32.2%	70.8%
Number of Instances	363	363	363

*Table 6.4: Fitting parameters for SMOreg applied to the dataset to predict embodied energy with a cross – validation approach for all structural components, excluding electrode 1 and considering only active layer components.*

When electrodes are not included, inspection of the weights show that again using a PET or a PSA|Barrier|PET substrate are the most beneficial for decreasing embodied

energy whilst quartz increases the embodied energy. For TL1, high conductivity PEDOT:PSS lowers the embodied energy whilst  $C_{60}F_{36}$ ,  $NiO_x$ ,  $CrO_2$  and slot die ZnO increase the embodied energy. For the active layer, IDTBRIDFBT, PDTSTTz-4 and PV-D4610 decrease the embodied energy whilst  $F_{16}CuPc$ ,  $C_{60}$ , P3OT and functionalised P3HT increase the embodied energy. For TL2, ZnO, AZO and HTL-Solar decrease the embodied energy whilst no TL2 materials are found to increase the embodied energy.

When only active layer components are considered in the analysis, PBDB-T, PDTSTTz-4 and PV-D4610 are found to lower the embodied energy the most, whilst C6PcH2, PCE12 and PPE-PPV:PCBM increase the embodied energy the most. The analysis for only the active materials allows for a rapid assessment of the relative benefits of different materials and provides direction for which materials experimental verification should focus on.



All structural attributes						Excluding Electrode 1,2						Active Only					
Best 10 Attributes			Worst 10 Attributes			Best 10 Attributes			Worst 10 Attributes			Best 10 Attributes			Worst 10 Attributes		
Name	Weight		Name	Weight		Name	Weight		Name	Weight		Name	Weight		Name	Weight	
Substrate=PET	-0.2359		TL2=Bphen:MoO <sub>3</sub>	0.018		Substrate=PET	-0.2984		Active2=F16CuPc	0.0721		Active1=PBDB-T	-0.2813		Active2=ICBA	0.0249	
Substrate=PSA_Barrier_PET	-0.2343		TL1=TiO <sub>2</sub>	0.0231		TL1=PEDOT:PSS Agfa5010 High Conductivity	-0.258		Active2=C60	0.0779		Active1=PD TSTTz-4(Slotdie)	-0.2241		Active2=CuPc	0.0251	
Elec1=Ag(Evaporation)	-0.1046		Elec1=Unknown	0.0484		Substrate=PSA_Barrier_PET	-0.2126		Active1=P3OT(Spincoated)	0.0952		Active1=PVD4610(Slotdie)	-0.2217		Active2=IDTB <sub>R</sub>	0.0266	
Elec1=Ag(ScreenPrinting)	-0.1034		Active2=IE(CO-4FP2FBTT-H	0.0516		TL2=ZnO	-0.1186		TL1=C60F36	0.1019		Active1=P3HT(Slotdie)	-0.2216		Active2=IT-4F	0.027	
Elec1=Ag(Inkjet)	-0.099		Elec1=ITO Ag ITO	0.074		TL2=AZO	-0.1179		Active1=P3HT (functionalised)	0.1024		Active1=PB DTTTz-4(Slotdie)	-0.2199		Active2=ZnPc	0.0285	
Active2=IDTB <sub>R</sub> DFBR	-0.0674		Elec1=Cr Aluminum	0.1073		TL2=HTL Solar	-0.1101		TL2=Alq3	0.1032		Active1=P3OT(Spincoated)	-0.1972		Active2=IDTB <sub>R</sub>	0.0372	
TL2=ZnO NP	-0.0559		TL1=CrO <sub>2</sub>	0.1073		TL2=ZnO(Slotdie)	-0.1077		TL1=NiOx	0.1046		Active2=PCBM(Slotdie)	-0.1742		Active1=Tetracene	0.0393	
Elec1=FTO	-0.0497		TL1=NiOx	0.1164		Active2=IDTB <sub>R</sub> DFBR	-0.0764		TL1=CrO <sub>2</sub>	0.126		Active2=ITIC	-0.0465		Active1=PPE-PPV;PCBM	0.0586	
Active1=AnE-PVstat	-0.0493		Elec1=ITO	0.1271		Active1=PDTSTTz-4(Slotdie)	-0.0692		TL1=ZnO(SlotDie)	0.1942		Active1=AnE-Pvstat	-0.0465		Active1=PCE1 <sub>2</sub>	0.1626	
Substrate=Glass	-0.0341		Substrate=Quartz	0.5044		Active1=PVD4610(Slotdie)	-0.0682		Substrate=Quartz	0.5239		Active2=IDTB <sub>R</sub> DFBR	-0.0322		Active1=C6Pc <sub>H2</sub>	0.5597	

Table 6.5: Best 10 attributes and worst 10 attributes for Embodied Energy minimisation considering the three attribute subsets.

The SMOreg algorithm has been applied to the predicted net energy output of the OPVs. Similarly, this was performed for all attributes, without electrodes and for active material only. The regression results are shown in Figure 6.6 (d)-(f) and the fitting parameters are shown in Table 6.6.

Using the 'net energy' parameter (this showed energy yield with the embodied energy subtracted), a high CC is achieved for the test set with values of 0.838, 0.743 and 0.719 for the three subsets of data respectively. Subsequently, the weights obtained from the SMOreg model can be analysed and ranked as the ten most beneficial and ten most detrimental for increasing the net energy, as shown in Table 6.7. For the beneficial weights, a more positive value indicates a greater improvement whilst a more negative value indicates a greater reduction in net energy.

<b>Metric</b>	<b>All attributes</b>	<b>Excluding Electrode1,2</b>	<b>Only Active layer components</b>
Correlation coefficient	0.838	0.743	0.719
Mean Absolute Error	175	166	218
Root Mean Squared Error	681	825	859
Relative Absolute Error	36.6%	34.6%	45.6%
Root Relative Squared Error	55.6%	67.5%	70.2%
Number of Instances	363	363	363

*Table 6.6: Fitting parameters for SMOreg applied to the dataset to predict net energy with a cross – validation approach.*

All structural attributes				Excluding Electrode 1,2				Active Only			
Best 10 Attributes		Worst 10 Attributes		Best 10 Attributes		Worst 10 Attributes		Best 10 Attributes		Worst 10 Attributes	
Name	Weight	Name	Weight	Name	Weight	Name	Weight	Name	Weight	Name	Weight
Active1=DRCN7T	0.0423	TL2=V2O5	-0.0166	Active1=DRCN7T	0.9242	TL1=AlQ3	-0.0395	Active1=DRCN7T	0.9206	Active1=P3HT	-0.0399
Active1=DR3TSBDT	0.0299	Elec1=Ag (Evaporation)	-0.0176	Active1=PBDTTT	0.2667	Active1=PININE:DTBT	-0.0401	Active1=PBDTTT	0.2667	Active1=PININE:DTBT	-0.0414
Active1=PBDTTT-EFT	0.0281	Active1=PBDTTTPD	-0.0215	TL1=P-type/HTL	0.0944	Active1=NDP2 (Novaled) dopedDINPB	-0.0402	Active1=PM6	0.0914	Active1=PSBTBT	-0.0421
TL2=ZnO NP	0.026	Active1=PDPP4T-2F	-0.0226	Active1=PM6	0.0916	Active1=MEH-PPV	-0.0404	Active1=PBDTTT PD	0.0739	Active1=PCPDTBT	-0.0423
Active2=IT-4F	0.026	Active2=IEICO- 4FP2FBTT-H	-0.0268	Active1=PBDTTT PD	0.0741	Active1=P3HT (functionalised)	-0.0413	Active1=PDPP4T- 2F	0.0729	Active1=CuPc	-0.0426
Active2=IDFBR	0.0205	TL2=ZrO	-0.0389	Active1=PM7	0.0764	Active1=P3OT(Spincoa ted)	-0.0449	Active1=PM7	0.0519	Active1=PPE- PPV:PCBM	-0.0426
Active2=IDTBRIDFBR	0.0182	Active1=PM7	-0.0443	Active1=PDPP4T -2F	0.0729	Active1=CN- P3HT(Slortdie)	-0.0453	Active1=PCDTBT	0.0498	Active1=SubPc	-0.0428
Substrate=PSA Barrier PET	0.0146	Substrate=Quartz	-0.0448	Active1=PCDTBT	0.0527	TL1=BF-DPB	-0.0502	Active1=IDTT	0.0378	Active1=Tetracene	-0.0428
Active1=PCDTBT	0.014	Embodied Energy	-0.156	Active2=IDTBRI DFBR	0.0293	Active1=PSBTBT	-0.058	Active2=IT-4F	0.0305	Active1=CN- P3HT(Slortdie)	-0.0439
TL1=NiOx	0.0122	Active1=PBDTTT	-0.2277	TL2=ZnONP	0.0275	Active1=PPE- PPV:PCBM	-0.0605	Active2=IDTBRID FBR	0.0294	Active1=C6PcH2	-0.0747

Table 6.7: Best 10 attributes and worst 10 attributes for Net Energy maximisation considering the three attribute subsets.

In the first dataset, all structural components are used as attributes for the predictive model. Considering the substrate, using a composite substrate consisting of PSA, barrier layer and PET is found to be beneficial for improving the net energy whilst using a quartz substrate proves to be detrimental, largely due to the high embodied energy. The former shows the benefit of using a barrier film; whilst the embodied energy of the OPV increases, the net energy produced increases more significantly since it is able to protect the cell better, ensuring that the OPV remains operational and generating energy for a longer period of time.

For Electrode 1 (Elec1), only evaporated silver is found to be detrimental for net energy. The likely reason for this is that there are a number of poor stability small devices (i.e. cells) in the literature which lowers the energy yield. Since they survive for <100 hours during light soaking, they can never recover the energy needed to manufacture the electrodes in the first instance. Despite causing an increase in the embodied energy,  $\text{NiO}_x$  is found to improve the net energy output. This is because  $\text{NiO}_x$  produces highly stable and efficient devices which operate for long enough to overcome the high embodied energy costs. No TL1s are identified as being detrimental for net energy.

For both improving and reducing the net energy, several different active layer materials have been identified, and these results could be used to optimise combinations of different active materials for improved net energy output. Notably, the three most beneficial materials are all active layer components being DRCN7T, DR3TSBDT and PBDTTT-EFT, which are all materials employed in more recent studies since 2019. In contrast, the most detrimental active material is PBDTTT, which is a more common material and studied extensively. PBDTTTPD is also identified as having a detrimental effect; in both cases due to poor stability. However, other novel materials investigated recently are identified as having a detrimental effect. These include PM7, IEICO-4FP2FBTT-H, and PDPP4T-2F. Experimental verification of this would be highly desirable by testing the identified active layer materials for improved energy output. Notably, the second transport layer (TL2) plays a more prominent role governing the net energy output as compared with the embodied energy.

ZnO nanoparticles are found to improve the net energy when used as TL2 whilst  $V_2O_5$  and ZrO are found to be detrimental. However, this could be symptomatic of the fact that ZnO has been tested far more widely than  $V_2O_5$  and ZrO and introduces some bias into the analysis. No Electrode 2 materials were identified when analysing the net energy output, implying that the choice of this material has minimal effect. It should also be noted that the embodied energy was used as a predictive attribute when analysing the net energy and found to be a negative influence. This means that as the embodied energy increases, the net energy decreases and vice versa. This demonstrates the need to use materials with low manufacturing costs whilst also maintaining a balance between using materials with a high energy output potential.

The analysis when excluding the electrodes and only considering the active materials is also performed for the net energy for completeness; these analyses are not as essential in this case since the analysis using all attributes is not dominated by the electrode. This is verified from inspection of the weights for the two other regimes, which show comparable results to the initial analysis. Nevertheless, the further analyses do provide a means of focussing on which materials, particularly the active material, provide the greatest enhancement. Notably, in all three analyses, DRCN7T is identified as the best active material for net energy enhancement. Interestingly, whilst PBDTTT is identified as a detrimental attribute for all structural attributes, it is identified as being beneficial when the electrodes are excluded, possibly indicating how the electrodes bias the analysis. IDTBRIDFBR is found to be beneficial when considering all structural attributes as well as when excluding the electrodes. In addition, P3HT derivatives are found to be detrimental when excluding the electrodes. PM7, which was found to be detrimental when considering all structural attributes, was found to be beneficial when excluding the electrodes. This illustrates how different aspects of the device can be focused on and important materials identified through filtering the dataset and considering different subsets of attributes.

#### 6.3.4 Clustering using Genetic Algorithms

In addition to using regression ML approaches, clustering has been undertaken using the genetic clustering algorithm, GenClust++. The SMOreg algorithm has proven successful at

identifying materials for enhancing the net energy. However, it does not optimise the combination of different materials in each layer to form the complete device stack; this is overcome by using the genetic clustering algorithms which determines the optimised combination of device layers. The net energies are now categorised in groups of equal distribution such that each class contains 20 instances and labelled from “A” to “R”. Where the same net energy values overlap across different classes, modifications are made to the distribution of instances whilst maintaining as equal a distribution as possible. Subsequently, the clustering algorithm was applied to the dataset and the classes to cluster evaluation is employed. This means that the attributes related to energy factors such as the materials and encapsulation are clustered, and each cluster is assigned to a particular net energy class. The results of the clustering algorithm are shown in Table 6.8.

Type	Encapsulation	Substrate	Electrode 1	Transport Layer 1	Active Layer 1	Active Layer 2	Transport Layer 2	Electrode 2	Embodied Energy	Net Energy	Net Energy (Class)	Net Energy (Magnitude)
Cell	None	Quartz	ITO	MoO <sub>3</sub>	C6PcH <sub>2</sub>	PCBM-61	None	Al	967.025	-966.589	A	
Cell	Rigid	Glass	Cr/Aluminum	CrO <sub>2</sub>	P3HT	PCBM-61	PEDOT:PSS	Au	536.61	-514.915	A	
Cell	None	Glass	ITO	MoO <sub>3</sub>	P3HT	PCBM-61	LiF	Al	459.76	-454.625	B	
Cell	Rigid	Glass	ITO	None	CuPc	C60	Bathocuproine	Al	450.88	-449.426	C	
Cell	None	Glass	ITO	PEDOT:PSS	P3HT	PCBM-61	Ca	Ag	446.09	-445.799	D	
Cell	None	Glass	ITO	ZnO	PTB7	PCBM-71	MoO <sub>3</sub>	Ag	456.355	-440.313	C	
Cell	Rigid	Glass	ITO	PEDOT:PSS	P3HT	PCBM-61	LiF	Al	442.68	-438.433	E	
Cell	None	Glass	ITO	PEDOT:PSS	P3HT	PCBM-61	None	Al	438.16	-437.52	G	
Cell	None	Glass	ITO	None	P3HT	PCBM	None	Al	434.74	-434.769	H	
Cell	Rigid	Glass	ITO	PEDOT:PSS	P3HT	PCBM-61	PEDOT:PSS	Al	441.87	-423.047	I	
Cell	Rigid	Glass	ITO	ZnO	P3HT	PCBM-61	PEDOT:PSS	Ag	440.58	-401.959	J	
Module	Flexible	PET	ITO	ZnO(SlotDie)	P3HT(Slotdie)	PCBM-61	HTLSolar	Ag	253.168	-250.22	L	
Cell	Rigid	Glass	Ag(Inkjet)	PEDOT:PSS	P3HT	PCBM-61	LiF	Al	225.218	-221.719	M	
Cell	Rigid	PET	Ag(Inkjet)	PEDOT:PSS	P3HT(Slotdie)	PCBM(Slotdie)	PEDOT:PSS	Ag(inkjet)	39.586	-37.44	N	
Module	Yes	PET	Ag(Evaporation)	ZnO	PBDB-T	ITIC	HTLSolar	CPP:PEDOT:PSS	33.44	-32.343	N	
Cell	None	PET	Ag(Inkjet)	PEDOT:PSS	P3HT	ICBA	AZO	Al	30.868	-27.52	O	
Module	Flexible	PET	Ag(Inkjet)	PEDOT:PSS	P3HT(Slotdie)	PCBM(Slotdie)	PEDOT:PSS	Ag(inkjet)	39.586	-21.151	O	
Cell	Rigid	Glass	ITO	ZnO	P3HT	PCBM-61	MoO <sub>3</sub>	Ag	455.89	117.3728	Q	
Cell	Rigid	Glass	ITO	PEDOT:PSS	PCDTBT	PCBM-71	Ca	Al	444.91	897.9285	R	

Increasing Net Energy

Table 6.8: Clustered attributes based on net energy output. Clusters ranked from low (A) to high (R).

In Table 6.8, each row corresponds to a single cluster and the column headings represent the attributes used for clustering. The rows have been ranked from low net

energy at the top to high net energy at the bottom. The column containing the embodied energy and net energy values give the raw calculated or derived values. These were not included as attributes when applying the clustering algorithm but are included in the table to aid with understanding the data. The distribution of instances in each cluster are shown in Table 6.9.

Cluster	Size
0	6 (8%)
2	7 (10%)
3	3 (4%)
4	11 (15%)
5	4 (5%)
6	3 (4%)
7	5 (7%)
8	4 (5%)
9	5 (7%)
12	3 (4%)
13	3 (4%)
14	8 (11%)
15	2 (3%)
16	2 (3%)
17	7 (10%)

*Table 6.9: Distribution of instances across clusters.*

The results shown in Table 6.8 show how different combinations of materials can be used in conjunction with each other to tune the net energy output and which combinations lead to the greatest net energy output. In this case, a cell having the structure: **Glass|ITO|PEDOT:PSS|PCDTBT|PC<sub>71</sub>BM|Ca|Al**, would lead to the greatest net energy output. For a module, the highest net energy is obtained with the structure: **PET|Ag(Inkjet)|PEDOT:PSS|P3HT(Slotdie)|PCBM(Slotdie)|PEDOT:PSS|Ag(Inkjet)**. Clustering allows us to identify groups of high performing devices. Trends are observed

where modules have better stability and best performing devices correspond to 'new' data.

It should be noted that these combinations of materials do not necessarily occur in the dataset and are derived based on the clustering algorithm, employing genetic search methods. Therefore, these combinations correspond to optimised combinations based on the analysed dataset. It can be seen that the majority of materials identified in the clusters are common materials such as ITO, PEDOT:PSS, P3HT and PCBM. However, interesting results can be seen when looking at other materials identified and where they reside in the overall ranking.

Even though the highest net energy is achieved using cells, modules are generally displaying improved net energies in comparison to cells, with the net energies ranging between class 'L' and class 'O'. This could be related to the encapsulation since all clusters for modules were encapsulated whereas many cell devices were not. This is reflected in Table 6.4 which shows that 43% of the clusters for cell devices were unencapsulated and 86% of these belong to the lower half of net energy classes. The cell devices which display the highest net energy are most likely associated with well-made devices which can outperform modules, even when they use evaporation of the electrodes [81].

For the substrate material, a trend can be seen where glass and quartz lead to low net energies whilst PET leads to higher embodied energies, although this trend is not absolute. A similar trend can be seen for Electrode 1 with ITO leading to lower net energies and silver leading to higher net energies. Interestingly, metal oxides do not appear as often as one would expect in high net energy class, apart from ZnO. This could be due to the large embodied energies which they possess, as compared with PEDOT:PSS. However, significant variance is present in the data. A greater prevalence of slot die coated active layer materials can be observed for higher net energy classes, with spin coated layers occupying low net energy classes. Non-fullerene acceptors (NFAs) also display some promise. However, similarly to transport layers, there is significant variance in the data with some NFAs displaying good stability and others displaying poor stability. However, ICBA and ITIC are identified for classes 'N' and 'O', indicating they relate to higher net energy clusters. For modules, all-inkjet processing appears as the most promising deposition technique for silver electrodes, as well as using conductive PEDOT:PSS. However, the model doesn't account for production speed; inkjet printing is



not as energy intensive as other printing/coating approaches, but probably has a lower throughput. Further work should consider this practical issue.

## 6.4 Conclusion

In this chapter, a database of performance, stability, and embodied energies for OPV devices and materials has been analysed using machine learning approaches. The methodology for producing the database and deriving the embodied energies is outlined. The method for subsequently deriving the net energy output of OPV devices tested under ISOS – L protocols is explained. An LCA methodology has been applied here in order to develop a comprehensive database of OPV material embodied energies. The embodied energy values have been found from literature sources, which have previously performed LCA on the materials, or have been derived using SimaPro software.

After acquiring the embodied energies of the constituent materials, the total embodied energy of each OPV device in the dataset was calculated. Subsequently, the net energy output of each device was calculated based on their initial efficiency,  $T_{80}$  and  $T_{S80}$  by calculating the energy output over the course of each stability test and subtracting the embodied energy. Machine learning techniques are then applied to analyse the net energy by applying the SMOREG algorithm using the structural components of each device as the predictive attributes and the net energy as the target variable for prediction. This provided a model giving the weights, and therefore relative significance of the different attributes in the improvement or deterioration of the net energy.

Finally, clustering algorithms, using a genetic search method, are employed in order to determine the optimum device configuration for maximising the total net energy output. This analysis provides a means of identifying optimum device configurations and, as material sets grow in complexity and scope, this method could prove invaluable in rapidly identifying the best device architectures for development without the need for exhaustive, time – intensive experimentation.

## Chapter 7. Conclusions and Avenues for Future Research

### 7.1 Conclusions from Results Chapters

Although accelerated lifetime testing has been used extensively for determining new materials for OPV development, the body of work is now vast and identifying the trends within the literature is becoming increasingly difficult. Machine learning was determined to be key in order to analyse the literature and provide a means of identifying materials which can enhance the PCE and stability of devices. In this thesis, machine learning has been used to successfully identify materials which can enhance the PCE and stability of devices. This has allowed for a deeper understanding and extraction of information from the scientific literature and subsequently expedite the development of new technologies at reduced costs and time requirements.

An OPV dataset consisting of 1850 data entries was utilised in this thesis in order to ascertain the materials which govern and enhance the dark intrinsic stability and photostability of OPV devices. Initially, multivariate linear regression was employed in order to find the principal trends within the dataset which governed the  $T_{80}$  and  $E_{80}$  lifetime metrics. Several important results were obtained using this method; namely, the importance of the electrode selection for improving the  $T_{80}$  response, the active material choice as well as the HTL and ETL selection. Furthermore, the multivariate linear regression analysis demonstrated the importance of a rigid encapsulation for improving the device stability, as opposed to a flexible one. However, the degree of knowledge that can be acquired *via* the multivariate approach is limited and largely qualitative. Therefore, a SMOREG machine learning algorithm was applied to the dataset in order to acquire a deeper understanding of the factors governing the performance and stability of OPVs. The SMOREG algorithm was used to model  $T_{80}$  and  $E_0$  and the model was analysed in terms of the attribute weights. This allowed the most significant OPV design properties to be identified and ranked in terms of their importance. This, therefore, allowed for a rapid assessment and identification of the materials governing the OPV stability and performance as well as identify which device components dictate the performance most prominently.

Another aspect of OPV and PSC development is the time requirements to test the technologies under real – world operation, which is difficult to simulate fully under indoor conditions. Machine learning offers the possibility of forecasting device performance with the outdoor environment providing an ideal platform to conduct this test. During outdoor testing the multitude of different stress conditions are, by default, applied simultaneously without the need for expensive environmental test chambers.

In this thesis, OPV modules were exposed to outdoor conditions for several months and seasons and their performance was monitored in the outdoors. The weather conditions were additionally monitored simultaneously along with the irradiance level. The OPV modules were tested at two sites: Bangor, North Wales and Belo Horizonte, Brasil. A remarkable difference in performance was observed between the two sites with devices degrading at a faster rate in Bangor. Indoor condensation tests subsequently revealed that this heightened degradation could be the result of a lower dew point depression in Bangor as compared with Belo Horizonte, thus leading to increased condensation and resultant water infiltration.

The outdoor dataset acquired was subsequently analysed using machine learning approaches; these included the multilayer perceptron and multivariate linear regression. The analysis was conducted using the environmental conditions as the predictive attributes for the models. The multilayer perceptron algorithm was employed to forecast the diurnal variation in the performance parameters and multivariate linear regression was employed to model the degradation of the modules over the course of summer and winter months. A significance test was performed in order to rank the factors governing the degradation. Subsequently, the energy yield of the modules was forecasted using each of the methods and, for summer, predictions with a 20.9% and 15.2% error were achieved for multilayer perceptron and multivariate linear regression, respectively. A novel approach was employed whereby the multilayer perceptron algorithm and the multivariate linear regression algorithms were used in conjunction as a means of reducing the relatively large error which was successfully reduced to 4.1%. This methodology was additionally applied to an ‘unseen’ winter dataset and a prediction with a 3.8% error was achieved. As a proof of concept, and to highlight the applicability of the method to other technologies, the same approach was applied to a dataset consisting of outdoor tested PSC modules. The method proved to be successful in predicting both the

diurnal performance of the PSC modules as well as the degradation over the course of two weeks.

In order to design more environmentally friendly OPVs, the energy costs of manufacturing must be balanced against the energy output from the devices. This was achieved by considering the embodied energy of the constituent materials alongside the net energy output from the OPV devices under test. This approach incorporated the LCA methodology with OPV lifetime testing. A dataset of embodied energies for OPV materials was collated and used to determine the total embodied energies of the OPV devices in the stability and performance database. Using the initial PCE, stability and embodied energy of the devices, the net energy produced by the OPV was determined.

Machine learning techniques were applied using SMOREg to predict both the embodied energy and the net energy with associated analysis of the attributes which lead to enhancement and reduction in both parameters as well as the relative significance of the attributes. This allowed several interesting materials to be identified which could lead to enhanced net energies despite having high embodied energy. An analysis of this type, which considers the holistic impact of devices was demonstrated to be highly relevant in balancing the cost versus output of this new technology.

Genetic clustering algorithms were applied to classified net energy data, with the structural components of the devices used as the predictive attributes. This algorithm clusters the structural components based on their similarity and associates each cluster with a particular net energy class. This allows the optimum combination of materials which lead to different net energies to be determined. The significance of this approach lies in its ability to identify material combinations which may not have been previously tested. As material sets develop and grow, this method could prove invaluable in rapidly identifying device architectures for enhanced net energy outputs before exhaustive experimental fabrication and testing has been performed.

## 7.2 Future Work

The research conducted herein demonstrated how acquiring reliable and consistent data is paramount to being able to arrive at meaningful and useful conclusions, and, ultimately, identify the optimum methods and materials for performance enhancement. It is also

noted that acquiring datasets dedicated to specific device features could lead to enhanced knowledge extraction and deeper understanding. It is important to have dedicated datasets since this reduces noise as well as computational cost. Given more time, future research could focus on analysis of detailed datasets dedicated to different device properties. For example, a dataset considering only the active materials used in different devices could include much greater detail on the material properties and electronic structures and could lead to an enhanced understanding of how the physics and chemistry of different materials lead to enhanced performance in OPV. Furthermore, the same methodology described herein should be applied to PSC datasets in order to understand the factors leading to degradation and subsequently identify factors and materials which can lead to stable PSC devices and modules.

For forecasting of the outdoor performance of OPVs and PSCs, the next task should be to acquire more datasets of outdoor tested OPV and PSC. These datasets should be broad and varied in terms of location and season. For example, for OPV modules, forecasting should be applied to data from different countries corresponding to different climates, thus allowing for a comparison of how the factors governing the performance and degradation vary as a function of environmental conditions; this would give a much greater understanding of the operational stabilities of the technology. Furthermore, the forecasting approach could be applied to other OPV technologies. This would show how different technologies, manufactured using different materials, respond to the environmental conditions. Once more datasets have been acquired the subsequent task would be to apply the different acquired models on the different climates and technologies, similar to what was tested in sections 5.2.5 and 5.2.6. This would yield a highly relevant and impactful method for rapidly assessing technologies and determine their viability in different climates across the globe.

PSC technologies have not yet reached the level of development which OPV technologies have attained and real – world testing of PSC is limited in scope, primarily due to their intrinsic low stabilities. However, the resource for developing a PSC database is rapidly emerging and would pave the way for identifying the factors which govern PSC stability and performance, in much the same way as shown in Chapter 4. Additionally, more outdoor tests could be conducted using emerging PSC technologies and applied to different environments to attain a greater understanding of the real – world operation of

PSCs as well as forecast their expected performance based on the environmental conditions. This would additionally give much greater insight into the dark – storage recovery observed in PSC technologies and determine how this property can influence the long – term stability when exposed to outdoor diurnal variation in irradiance.

## REFERENCES

- [1] R. K. Pachauri *et al.*, "Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change," IPCC, Geneva, Switzerland, Geneva.
- [2] Unfccc, "ADOPTION OF THE PARIS AGREEMENT - Paris Agreement text English. Accessed: Jul. 28, 2020. [Online]."
- [3] M. Perez and R. Perez, "UPDATE 2015-A FUNDAMENTAL LOOK AT SUPPLY SIDE ENERGY RESERVES FOR THE PLANET," 2015.
- [4] A. Detollenaere - Becquerel Institute, *Snapshot of Global PV Markets 2020 PVPS Task 1 Strategic PV Analysis and Outreach.* .
- [5] "Best Research-Cell Efficiency Chart | Photovoltaic Research | NREL." [Online]. Available: <https://www.nrel.gov/pv/cell-efficiency.html>. [Accessed: 28-Jul-2020].
- [6] J. U. Tsao, N. Lewis, and G. Crabtree Argonne, "Solar FAQs."
- [7] E. J. McCartney, McCartney, and E. J., "Optics of the atmosphere: Scattering by molecules and particles," *nyjw*, 1976.
- [8] "Direct Normal Irradiation - an overview | ScienceDirect Topics." [Online]. Available: <https://www.sciencedirect.com/topics/engineering/direct-normal-irradiation>. [Accessed: 26-Aug-2020].
- [9] "Diffuse Radiation - an overview | ScienceDirect Topics." [Online]. Available: <https://www.sciencedirect.com/topics/engineering/diffuse-radiation>. [Accessed: 26-Aug-2020].
- [10] "Standard Solar Spectra | PVEducation." [Online]. Available: <https://www.pveducation.org/pvcdrom/appendices/standard-solar-spectra>. [Accessed: 26-Aug-2020].
- [11] "Solar Spectra | Grid Modernization | NREL." [Online]. Available: <https://www.nrel.gov/grid/solar-resource/spectra.html>. [Accessed: 26-Aug-2020].
- [12] D. R. Myers, K. Emery, and C. Gueymard, "Proposed reference spectral irradiance standards to improve concentrating photovoltaic system design and performance evaluation," in *Conference Record of the IEEE Photovoltaic Specialists Conference*, 2002, pp. 923–926.
- [13] T. Saga, "Advances in crystalline silicon solar cell technology for industrial mass production," *NPG Asia Materials*, vol. 2, no. 3. Nature Publishing Group, pp. 96–102, 22-Jul-2010.
- [14] L. C. Andreani, A. Bozzola, P. Kowalczewski, M. Liscidini, and L. Redorici, "Silicon solar cells: Toward the efficiency limits," *Advances in Physics: X*, vol. 4, no. 1. Taylor and Francis Ltd., p. 1548305, 01-Jan-2019.
- [15] "How a Solar Cell Works - American Chemical Society." [Online]. Available:

- <https://www.acs.org/content/acs/en/education/resources/highschool/chemmatters/past-issues/archive-2013-2014/how-a-solar-cell-works.html>. [Accessed: 17-May-2021].
- [16] T. Markvart and L. Castañer, "Principles of solar cell operation," in *McEvoy's Handbook of Photovoltaics: Fundamentals and Applications*, Elsevier Inc., 2018, pp. 3–28.
  - [17] B. Qi and J. Wang, "Fill factor in organic solar cells," *Phys. Chem. Chem. Phys.*, vol. 15, p. 8972, 2013.
  - [18] M. A. Green, "Silicon photovoltaic modules: a brief history of the first 50 years," *Prog. Photovoltaics Res. Appl.*, vol. 13, no. 5, pp. 447–455, Aug. 2005.
  - [19] "Fraunhofer Institute for Solar Energy Systems, 'Photovoltaics Report,' 2016. [Online]. Available: <https://www.ise.fraunhofer.de/en/downloads-englisch/pdf-files-englisch/photovoltaics-report-slides.pdf>."
  - [20] "Exploring Alternatives to Silicon Photovoltaic Cells - ADI Analytics." [Online]. Available: <https://adi-analytics.com/2016/07/11/exploring-alternatives-to-silicon-photovoltaic-cells/>. [Accessed: 11-Sep-2020].
  - [21] C. W. Tang, "Two-layer organic photovoltaic cell," *Appl. Phys. Lett.*, vol. 48, no. 2, pp. 183–185, Jan. 1986.
  - [22] G. Yu, J. Gao, J. C. Hummelen, F. Wudl, and A. J. Heeger, "Polymer photovoltaic cells: Enhanced efficiencies via a network of internal donor-acceptor heterojunctions," *Science (80-. )*, vol. 270, no. 5243, p. 1789, Dec. 1995.
  - [23] M. M. Lee, J. Teuscher, T. Miyasaka, T. N. Murakami, and H. J. Snaith, "Efficient hybrid solar cells based on meso-superstructured organometal halide perovskites," *Science (80-. )*, vol. 338, no. 6107, pp. 643–647, Nov. 2012.
  - [24] C. Kittel, "Introduction to Solid State Physics, 8th edition," *Wiley & Sons, New York, NY*. 2004.
  - [25] J. M. Nunzi, "Organic photovoltaic materials and devices," *Comptes Rendus Phys.*, vol. 3, no. 4, pp. 523–542, Jan. 2002.
  - [26] M. Fox, "Optical properties of solids."
  - [27] N. Marinova, S. Valero, and J. L. Delgado, "Organic and perovskite solar cells: Working principles, materials and interfaces," *Journal of Colloid and Interface Science*, vol. 488. Academic Press Inc., pp. 373–389, 15-Feb-2017.
  - [28] D. Caruso and A. Troisi, "Long-range exciton dissociation in organic solar cells."
  - [29] A. Dreuw, J. L. Weisman, and M. Head-Gordon, "Long-range charge-transfer excited states in time-dependent density functional theory require non-local exchange," *J. Chem. Phys.*, vol. 119, p. 2943, 2003.
  - [30] V. I. Arkhipov, E. V Emelianova, and H. Bässler, "Charge carrier transport and recombination at the interface between disordered organic dielectrics," *J. Appl. Phys.*, vol. 90, p. 2352, 2001.
  - [31] D. Credgington, F. C. Jamieson, B. Walker, T.-Q. Nguyen, and J. R. Durrant,



- "Quantification of Geminate and Non-Geminate Recombination Losses within a Solution-Processed Small-Molecule Bulk Heterojunction Solar Cell," *Adv. Mater.*, vol. 24, no. 16, pp. 2135–2141, Apr. 2012.
- [32] J. Nelson, S. A. Choulis, and J. R. Durrant, "Charge recombination in polymer/fullerene photovoltaic devices," in *Thin Solid Films*, 2004, vol. 451–452, pp. 508–514.
- [33] C. M. Proctor, M. Kuik, and T. Q. Nguyen, "Charge carrier recombination in organic solar cells," *Progress in Polymer Science*, vol. 38, no. 12, Pergamon, pp. 1941–1960, 01-Dec-2013.
- [34] S. M. Menke, N. A. Ran, G. C. Bazan, and R. H. Friend, "Understanding Energy Loss in Organic Solar Cells: Toward a New Efficiency Regime," *Joule*, vol. 2, no. 1, pp. 25–35, 2018.
- [35] J. Nelson, "Organic photovoltaic films," *Materials Today*, vol. 5, no. 5, Elsevier, pp. 20–27, 01-May-2002.
- [36] C. W. Tang, "Two-layer organic photovoltaic cell," *Appl. Phys. Lett.*, vol. 48, no. 2, pp. 183–185, 1986.
- [37] S. S. Sun and L. R. Dalton, *Introduction to organic electronic and optoelectronic materials and devices: Second edition*. CRC Press, 2016.
- [38] "Diffusion Length - an overview | ScienceDirect Topics." [Online]. Available: <https://www.sciencedirect.com/topics/earth-and-planetary-sciences/diffusion-length>. [Accessed: 14-Jul-2020].
- [39] Y. Yang, K. Mielczarek, M. Aryal, A. Zakhidov, and W. Hu, "Nanoimprinted polymer solar cell," *ACS Nano*, vol. 6, no. 4, pp. 2877–2892, Apr. 2012.
- [40] J. E. Slota, X. He, and W. T. S. Huck, "Controlling nanoscale morphology in polymer photovoltaic devices," *Nano Today*, vol. 5, no. 3, Elsevier, pp. 231–242, 01-Jun-2010.
- [41] T. H. Lai, S. W. Tsang, J. R. Manders, S. Chen, and F. So, "Properties of interlayer for organic photovoltaics," *Materials Today*, vol. 16, no. 11, Elsevier B.V., pp. 424–432, 01-Nov-2013.
- [42] V. Gowrishankar, S. R. Scully, M. D. Mcgehee, Q. Wang, and H. M. Branz, "Exciton splitting and carrier transport across the amorphous-silicon/polymer solar cell interface," *Appl. Phys. Lett.*, vol. 89, p. 252102, 2006.
- [43] J. Drechsel, B. Männig, F. Kozlowski, M. Pfeiffer, K. Leo, and H. Hoppe, "Efficient organic solar cells based on a double architecture using doped wide-gap transport layers ARTICLES YOU MAY BE INTERESTED IN," *Appl. Phys. Lett.*, vol. 86, p. 244102, 2005.
- [44] X. Xi *et al.*, "The characteristics of the small molecule organic solar cells with PEDOT:PSS/LiF double anode buffer layer system," *Sol. Energy Mater. Sol. Cells*, vol. 94, no. 3, pp. 623–628, Mar. 2010.
- [45] D. Cubicciotti, "The Oxidation of Calcium at Elevated Temperatures," *J. Am. Chem. Soc.*, vol. 74, no. 2, pp. 557–558, Jan. 1952.

- [46] D. A. Nissen, "The low-temperature oxidation of calcium by water vapor," *Oxid. Met.*, vol. 11, no. 5, pp. 241–261, Oct. 1977.
- [47] G. E. Thompson, Y. Xu, P. Skeldon, K. Shimizu, S. H. Han, and G. C. Wood, "Anodic oxidation of aluminium," *Philos. Mag. B Phys. Condens. Matter; Stat. Mech. Electron. Opt. Magn. Prop.*, vol. 55, no. 6, pp. 651–667, 1987.
- [48] "The oxidation of aluminium in dry and humid oxygen atmospheres," *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.*, vol. 236, no. 1204, pp. 68–88, Jul. 1956.
- [49] J. K. Kim, "PEG-assisted Sol-gel Synthesis of Compact Nickel Oxide Hole-Selective Layer with Modified Interfacial Properties for Organic Solar Cells," *Polymers (Basel)*, vol. 11, no. 1, p. 120, Jan. 2019.
- [50] N. Chaudhary, R. Chaudhary, J. P. Kesari, A. Patra, and S. Chand, "Copper thiocyanate (CuSCN): an efficient solution-processable hole transporting layer in organic solar cells," *J. Mater. Chem. C*, vol. 3, no. 45, pp. 11886–11892, Oct. 2015.
- [51] M. Aatif *et al.*, "Graphene oxide-molybdenum oxide composite with improved hole transport in bulk heterojunction solar cells ARTICLES YOU MAY BE INTERESTED IN," 2019.
- [52] S. Schumann *et al.*, "Inverted organic photovoltaic devices with high efficiency and stability based on metal oxide charge extraction layers," *J. Mater. Chem.*, vol. 21, no. 7, pp. 2381–2386, Feb. 2011.
- [53] B. A. Macleod *et al.*, "Stability of inverted organic solar cells with ZnO contact layers deposited from precursor solutions," *Energy Environ. Sci.*, vol. 8, no. 2, pp. 592–601, Feb. 2015.
- [54] S. Kundu *et al.*, "Device stability of inverted and conventional bulk heterojunction solar cells with MoO<sub>3</sub> and ZnO nanoparticles as charge transport layers," *Org. Electron.*, vol. 14, no. 11, pp. 3083–3088, Nov. 2013.
- [55] D. H. Kim, D. J. Lee, B. Kim, C. Yun, and M. H. Kang, "Tailoring PEDOT:PSS polymer electrode for solution-processed inverted organic solar cells," *Solid. State. Electron.*, vol. 169, p. 107808, Jul. 2020.
- [56] M. Ruscello *et al.*, "Nanocomposite of nickel oxide nanoparticles and polyethylene oxide as printable hole transport layer for organic solar cells †," 2019.
- [57] M. Zafar, J. Y. Yun, and D. H. Kim, "Highly stable inverted organic photovoltaic cells with a V<sub>2</sub>O<sub>5</sub> hole transport layer," *Korean J. Chem. Eng.*, vol. 34, no. 5, pp. 1504–1508, May 2017.
- [58] M. Zafar, J. Y. Yun, and D. H. Kim, "Performance of inverted organic photovoltaic cells with nitrogen doped TiO<sub>2</sub> films by atomic layer deposition," *Korean J. Chem. Eng.*, vol. 35, no. 2, pp. 567–573, Feb. 2018.
- [59] M. F. M. Noh *et al.*, "Effect of temperature on the properties of SnO<sub>2</sub> layer fabricated via AACVD and its application in photoelectrochemical cells and organic photovoltaic devices," *Sol. Energy*, vol. 158, pp. 474–482, 2017.
- [60] A. W. Hains, H. Y. Chen, T. H. Reilly, and B. A. Gregg, "Cross-linked perylene diimide-based n-type interfacial layer for inverted organic photovoltaic devices,"

- ACS Appl. Mater. Interfaces*, vol. 3, no. 11, pp. 4381–4387, Nov. 2011.
- [61] J. Subbiah, C. M. Amb, I. Irfan, Y. Gao, J. R. Reynolds, and F. So, “High-efficiency inverted polymer solar cells with double interlayer,” *ACS Appl. Mater. Interfaces*, vol. 4, no. 2, pp. 866–870, Feb. 2012.
  - [62] M. S. White, D. C. Olson, S. E. Shaheen, N. Kopidakis, and D. S. Ginley, “Inverted bulk-heterojunction organic photovoltaic device using a solution-derived ZnO underlayer,” *Appl. Phys. Lett.*, vol. 89, no. 14, p. 143517, Oct. 2006.
  - [63] S. A. Gevorgyan *et al.*, “Lifetime of Organic Photovoltaics: Status and Predictions,” *Adv. Energy Mater.*, vol. 6, no. 2, p. 1501208, Jan. 2016.
  - [64] J. Ahmad, K. Bazaka, L. J. Anderson, R. D. White, and M. V. Jacob, “Materials and methods for encapsulation of OPV: A review,” *Renewable and Sustainable Energy Reviews*, vol. 27, Pergamon, pp. 104–117, 01-Nov-2013.
  - [65] H. J. Lee *et al.*, “Solution processed encapsulation for organic photovoltaics,” *Sol. Energy Mater. Sol. Cells*, vol. 111, pp. 97–101, Apr. 2013.
  - [66] A. Uddin, M. Upama, H. Yi, and L. Duan, “Encapsulation of Organic and Perovskite Solar Cells: A Review,” *Coatings*, vol. 9, no. 2, p. 65, Jan. 2019.
  - [67] M. D. Clark, M. L. Jespersen, R. J. Patel, and B. J. Leever, “Ultra-thin alumina layer encapsulation of bulk heterojunction organic photovoltaics for enhanced device lifetime,” *Org. Electron.*, vol. 15, no. 1, pp. 1–8, Jan. 2014.
  - [68] J. Kettle *et al.*, “Printable luminescent down shifter for enhancing efficiency and stability of organic photovoltaics,” *Sol. Energy Mater. Sol. Cells*, vol. 144, pp. 481–487, Jan. 2016.
  - [69] D. M. Tanenbaum, H. F. Dam, R. Rösch, M. Jørgensen, H. Hoppe, and F. C. Krebs, “Edge sealing for low cost stability enhancement of roll-to-roll processed flexible polymer solar cell modules,” in *Solar Energy Materials and Solar Cells*, 2012, vol. 97, pp. 157–163.
  - [70] B. Paci *et al.*, “Stability enhancement of organic photovoltaic devices utilizing partially reduced graphene oxide as the hole transport layer: Nanoscale insight into structural/interfacial properties and aging effects,” *RSC Adv.*, vol. 5, no. 129, pp. 106930–106940, 2015.
  - [71] K. X. Steirer *et al.*, “Solution deposited NiO thin-films as hole transport layers in organic photovoltaics,” *Org. Electron.*, vol. 11, no. 8, pp. 1414–1418, Aug. 2010.
  - [72] M. Li *et al.*, “Graphene quantum dots as the hole transport layer material for high-performance organic solar cells,” *Phys. Chem. Chem. Phys.*, vol. 15, no. 43, pp. 18973–18978, Nov. 2013.
  - [73] M. H. Chen, Y. C. Kuo, H. H. Lin, Y. P. Chao, and M. S. Wong, “Highly stable inverted organic photovoltaics using aluminum-doped zinc oxide as electron transport layers,” *J. Power Sources*, vol. 275, pp. 274–278, Feb. 2015.
  - [74] Y.-J. Lee *et al.*, “Low-Temperature Solution-Processed Molybdenum Oxide Nanoparticle Hole Transport Layers for Organic Photovoltaic Devices,” *Adv. Energy Mater.*, vol. 2, no. 10, pp. 1193–1197, Oct. 2012.

- [75] N. Ahmad *et al.*, "Polydopamine/ZnO electron transport layers enhance charge extraction in inverted non-fullerene organic solar cells," *J. Mater. Chem. C*, vol. 7, no. 35, pp. 10795–10801, 2019.
- [76] S. Saylan and C. T. Howells, "Boosting the short-circuit current density of organic photovoltaics using a composite electrode," *Synth. Met.*, vol. 265, p. 116379, Jul. 2020.
- [77] G. Madhaiyan *et al.*, "A 0.05 v driven ammonia gas sensor based on an organic diode with a top porous layered electrode and an air-stable sensing film," *J. Mater. Chem. C*, vol. 7, no. 21, pp. 6440–6447, May 2019.
- [78] H. J. Son, H. K. Park, J. Y. Moon, B. K. Ju, and S. H. Kim, "Thermal degradation related to the PEDOT:PSS hole transport layer and back electrode of the flexible inverted organic photovoltaic module," *Sustain. Energy Fuels*, vol. 4, no. 4, pp. 1974–1983, Apr. 2020.
- [79] M. A. Saeed, S. H. Kim, S. Y. Lee, and J. W. Shim, "High indoor performance of flexible organic photovoltaics using polymer electrodes," *Thin Solid Films*, vol. 704, p. 138006, Jun. 2020.
- [80] D. J. Lee, D. K. Heo, C. Yun, Y. H. Kim, and M. H. Kang, "Solution-Processed Semitransparent Inverted Organic Solar Cells from a Transparent Conductive Polymer Electrode," *ECS J. Solid State Sci. Technol.*, vol. 8, no. 2, pp. Q32–Q37, Mar. 2019.
- [81] C. H. Peters, I. T. Sachs-Quintana, J. P. Kastrop, S. Beaupré, M. Leclerc, and M. D. McGehee, "High Efficiency Polymer Solar Cells with Long Operating Lifetimes," *Adv. Energy Mater.*, vol. 1, no. 4, pp. 491–494, Jul. 2011.
- [82] D. Baran *et al.*, "Reducing the efficiency-stability-cost gap of organic photovoltaics with highly efficient and stable small molecule acceptor ternary solar cells," *Nat. Mater.*, vol. 16, no. 3, pp. 363–369, Mar. 2017.
- [83] N. Gasparini *et al.*, "Burn-in Free Nonfullerene-Based Organic Solar Cells," *Adv. Energy Mater.*, vol. 7, no. 19, 2017.
- [84] Q. Burlingame *et al.*, "Reliability of Small Molecule Organic Photovoltaics with Electron-Filtering Compound Buffer Layers," *Adv. Energy Mater.*, vol. 6, no. 21, p. 1601094, Nov. 2016.
- [85] M. Ylikunnari *et al.*, "Flexible and Printed Electronics Flexible OPV modules for highly efficient indoor applications," 2020.
- [86] Y. Cui *et al.*, "Over 16% efficiency organic photovoltaic cells enabled by a chlorinated acceptor with increased open-circuit voltages," *Nat. Commun.*, vol. 10, no. 1, pp. 1–8, Dec. 2019.
- [87] Q. Liu *et al.*, "18% Efficiency organic solar cells," *Sci. Bull.*, vol. 65, no. 4, pp. 272–275, Feb. 2020.
- [88] M. O. Reese *et al.*, "Consensus stability testing protocols for organic photovoltaic materials and devices," *Sol. Energy Mater. Sol. Cells*, vol. 95, no. 5, pp. 1253–1267, 2011.

- [89] M. V. Khenkin *et al.*, "Consensus statement for stability assessment and reporting for perovskite photovoltaics based on ISOS procedures," *Nat. Energy*, vol. 5, no. 1, pp. 35–49, Jan. 2020.
- [90] W. R. Mateker and M. D. McGehee, "Progress in Understanding Degradation Mechanisms and Improving Stability in Organic Photovoltaics," *Advanced Materials*, vol. 29, no. 10. Wiley-VCH Verlag, 14-Mar-2017.
- [91] S. A. Gevorgyan *et al.*, "Improving, characterizing and predicting the lifetime of organic photovoltaics: Topical Review," *J. Phys. D. Appl. Phys.*, vol. 50, no. 10, pp. 1–35, 2017.
- [92] M. Corazza, F. C. Krebs, and S. A. Gevorgyan, "Predicting, categorizing and intercomparing the lifetime of OPVs for different ageing tests," *Sol. Energy Mater. Sol. Cells*, vol. 130, pp. 99–106, 2014.
- [93] V. Stoichkov, D. Kumar, P. Tyagi, and J. Kettle, "Multistress testing of OPV modules for accurate predictive aging and reliability predictions," *IEEE J. Photovoltaics*, vol. 8, no. 4, pp. 1058–1065, Jul. 2018.
- [94] A. K. K. Kyaw *et al.*, "Efficient solution-processed small-molecule solar cells with inverted structure," *Adv. Mater.*, vol. 25, no. 17, pp. 2397–2402, May 2013.
- [95] A. Guerrero and G. Garcia-Belmonte, "Recent advances to understand morphology stability of organic photovoltaics," *Nano-Micro Letters*, vol. 9, no. 1. SpringerOpen, pp. 1–16, 01-Jan-2017.
- [96] F. C. Krebs and K. Norrman, "Analysis of the failure mechanism for a stable organic photovoltaic during 10 000 h of testing," *Prog. Photovoltaics Res. Appl.*, vol. 15, no. 8, pp. 697–712, Dec. 2007.
- [97] P. Cheng and X. Zhan, "Stability of organic solar cells: Challenges and strategies," *Chemical Society Reviews*, vol. 45, no. 9. Royal Society of Chemistry, pp. 2544–2582, 07-May-2016.
- [98] M. Y. Ameen, P. Shamjid, T. Abhijith, T. Radhakrishnan, and V. S. Reddy, "Stability enhancement of P3HT:PCBM polymer solar cells using thermally evaporated MoO<sub>3</sub> anode buffer layer," *Phys. B Condens. Matter*, vol. 530, pp. 201–207, Feb. 2018.
- [99] N. Y. Doumon *et al.*, "Improved photostability in ternary blend organic solar cells: The role of [70]PCBM," *J. Mater. Chem. C*, vol. 7, no. 17, pp. 5104–5111, 2019.
- [100] Z. P. Yu *et al.*, "Simple non-fused electron acceptors for efficient and stable organic solar cells," *Nat. Commun.*, vol. 10, no. 1, pp. 1–9, Dec. 2019.
- [101] Y. Zhang *et al.*, "Comparative indoor and outdoor stability measurements of polymer based solar cells OPEN."
- [102] M. Karakawa *et al.*, "Factors contributing to degradation of organic photovoltaic cells," *Org. Electron.*, vol. 76, Jan. 2020.
- [103] V. I. Madogni, B. Kounouhéwa, A. Akpo, M. Agbomahéna, S. A. Hounkpatin, and C. N. Awanou, "Comparison of degradation mechanisms in organic photovoltaic devices upon exposure to a temperate and a subequatorial climate," *Chem. Phys.*

- Lett.*, vol. 640, pp. 201–214, Nov. 2015.
- [104] H. C. Weerasinghe, S. E. Watkins, N. Duffy, D. J. Jones, and A. D. Scully, “Influence of moisture out-gassing from encapsulant materials on the lifetime of organic solar cells,” *Sol. Energy Mater. Sol. Cells*, vol. 132, pp. 485–491, Jan. 2015.
- [105] T. S. Glen *et al.*, “Dependence on material choice of degradation of organic solar cells following exposure to humid air,” *J. Polym. Sci. Part B Polym. Phys.*, vol. 54, no. 2, pp. 216–224, Jan. 2016.
- [106] R. Roesch, K. R. Eberhardt, S. Engmann, G. Gobsch, and H. Hoppe, “Polymer solar cells with enhanced lifetime by improved electrode stability and sealing,” *Sol. Energy Mater. Sol. Cells*, vol. 117, pp. 59–66, 2013.
- [107] S. Bertho *et al.*, “Effect of temperature on the morphological and photovoltaic stability of bulk heterojunction polymer:fullerene solar cells,” *Sol. Energy Mater. Sol. Cells*, vol. 92, no. 7, pp. 753–760, Jul. 2008.
- [108] Y. Jang, Y. Ju Cho, M. Kim, J. Seok, H. Ahn, and K. Kim, “Formation of Thermally Stable Bulk Heterojunction by Reducing the Polymer and Fullerene Intermixing,” *Sci. Rep.*, vol. 7, no. 1, pp. 1–9, Dec. 2017.
- [109] F. Yang, J.-H. Kim, Z. Ge, and Y.-S. Kim, “Enhanced Thermal Stability of Inverted Polymer Solar Cells with Pentacene,” *Isr. J. Chem.*, vol. 55, no. 9, pp. 1028–1033, Sep. 2015.
- [110] W. Yang *et al.*, “Simultaneous enhanced efficiency and thermal stability in organic solar cells from a polymer acceptor additive,” *Nat. Commun.*, vol. 11, no. 1, pp. 1–10, Dec. 2020.
- [111] X. Xu *et al.*, “Thermally stable, highly efficient, ultraflexible organic photovoltaics.”
- [112] Z.-G. Zhang *et al.*, “Constructing a Strongly Absorbing Low-Bandgap Polymer Acceptor for High-Performance All-Polymer Solar Cells,” *Angew. Chemie Int. Ed.*, vol. 56, no. 43, pp. 13503–13507, Oct. 2017.
- [113] Y. J. Kim, D. S. Chung, and C. E. Park, “Highly thermally stable non-fullerene organic solar cells: P-DTS(FBTTh2)2: P(NDI2OD-T2) bulk heterojunction,” *Nano Energy*, vol. 15, pp. 343–352, Jul. 2015.
- [114] W. Greenbank, L. Hirsch, and S. Chambon, “Electrode de-wetting as a failure mechanism in thermally-aged OPV devices,” *Sol. Energy Mater. Sol. Cells*, vol. 178, pp. 8–14, May 2018.
- [115] E. A. Katz, S. Gevorgyan, and F. C. Krebs, “Out-door testing and long-term stability of plastic solar cells,” *Eur. Phys. J. Appl. Phys.*, vol. 36, pp. 307–311, 2007.
- [116] F. C. Krebs, N. Espinosa, M. Hösel, R. R. Søndergaard, and M. Jørgensen, “25th anniversary article: Rise to power - OPV-based solar parks,” *Advanced Materials*, vol. 26, no. 1. John Wiley & Sons, Ltd, pp. 29–39, 08-Jan-2014.
- [117] R. K. Garner *et al.*, “Boron subphthalocyanines as electron donors in outdoor lifetime monitored organic photovoltaic cells,” *Sol. Energy Mater. Sol. Cells*, vol. 176, pp. 331–335, Mar. 2018.
- [118] D. S. Josey *et al.*, “Outdoor Performance and Stability of Boron

Subphthalocyanines Applied as Electron Acceptors in Fullerene-Free Organic Photovoltaics,” 2017.

- [119] N. Bristow and J. Kettle, “Outdoor performance of organic photovoltaics: Diurnal analysis, dependence on temperature, irradiance, and degradation,” *J. Renew. Sustain. Energy*, vol. 7, no. 1, p. 013111, Jan. 2015.
- [120] R. Saive, “S-Shaped Current-Voltage Characteristics in Solar Cells: A Review,” *IEEE J. Photovoltaics*, vol. 9, no. 6, pp. 1477–1484, Nov. 2019.
- [121] N. Bristow and J. Kettle, “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*, vol. 175, pp. 52–59, Feb. 2018.
- [122] M. Koehl, M. Heck, S. Wiesmeier, and J. Wirth, “Modeling of the nominal operating cell temperature based on outdoor weathering,” *Sol. Energy Mater. Sol. Cells*, vol. 95, no. 7, pp. 1638–1646, Jul. 2011.
- [123] L. Maturi, G. Belluardo, D. Moser, and M. Del Buono, “BiPV system performance and efficiency drops: Overview on PV module temperature conditions of different module types,” in *Energy Procedia*, 2014, vol. 48, pp. 1311–1319.
- [124] Y. Zhang, I. D. W. Samuel, T. Wang, and D. G. Lidzey, “Current Status of Outdoor Lifetime Testing of Organic Photovoltaics,” *Adv. Sci.*, vol. 5, no. 8, p. 1800434, Aug. 2018.
- [125] Y. M. Sung, Y. C. Huang, F. S. Sen Chien, and C. S. Tsao, “Mechanism and Analysis of Thermal Burn-In Degradation of OPVs Induced by Evaporated HTL,” *IEEE J. Photovoltaics*, vol. 9, no. 3, pp. 694–699, May 2019.
- [126] W. Sun *et al.*, “The Use of Deep Learning to Fast Evaluate Organic Photovoltaic Materials,” *Adv. Theory Simulations*, vol. 2, no. 1, p. 1800116, Jan. 2019.
- [127] E. O. Pyzer-Knapp, K. Li, and A. Aspuru-Guzik, “Learning from the Harvard Clean Energy Project: The Use of Neural Networks to Accelerate Materials Discovery,” *Adv. Funct. Mater.*, vol. 25, no. 41, pp. 6495–6502, Nov. 2015.
- [128] Z. Li, Q. Xu, Q. Sun, Z. Hou, and W.-J. Yin, “Stability Engineering of Halide Perovskite via Machine Learning,” Mar. 2018.
- [129] Ç. Odabaşı Özer and R. Yıldırım, “Performance analysis of perovskite solar cells in 2013–2018 using machine-learning tools,” *Nano Energy*, vol. 56, pp. 770–791, Feb. 2019.
- [130] H. Sahu, W. Rao, A. Troisi, and H. Ma, “Toward Predicting Efficiency of Organic Solar Cells via Machine Learning and Improved Descriptors,” *Adv. Energy Mater.*, vol. 8, no. 24, p. 1801032, Aug. 2018.
- [131] R. Fazai *et al.*, “Machine learning-based statistical testing hypothesis for fault detection in photovoltaic systems,” *Sol. Energy*, vol. 190, pp. 405–413, Sep. 2019.
- [132] E. Garoudja, A. Chouder, K. Kara, and S. Silvestre, “An enhanced machine learning based approach for failures detection and diagnosis of PV systems,” *Energy Convers. Manag.*, vol. 151, pp. 496–513, Nov. 2017.

- [133] D. Assouline, N. Mohajeri, and J.-L. Scartezzini, "Quantifying rooftop photovoltaic solar energy potential: A machine learning approach," 2016.
- [134] M. P. Almeida, O. Perpiñán, and L. Narvarte, "PV power forecast using a nonparametric PV model," *Sol. Energy*, vol. 115, pp. 354–368, May 2015.
- [135] M. Hossain, S. Mekhilef, M. Danesh, L. Olatomiwa, and S. Shamshirband, "Application of extreme learning machine for short term output power forecasting of three grid-connected PV systems," *J. Clean. Prod.*, vol. 167, pp. 395–405, Nov. 2017.
- [136] G. Perveen, M. Rizwan, and N. Goel, "Short-term PV power forecasting based on sky-conditions using intelligent modelling techniques," *Int. J. Eng. Sci. Technol.*, vol. 11, no. 4, p. 49, Oct. 2019.
- [137] H. T. Yang, C. M. Huang, Y. C. Huang, and Y. S. Pai, "A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output," *IEEE Trans. Sustain. Energy*, vol. 5, no. 3, pp. 917–926, 2014.
- [138] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments."
- [139] E. Oja, "Principal components, minor components, and linear neural networks," *Neural Networks*, vol. 5, no. 6, pp. 927–935, 1992.
- [140] R. Palaniappan and K. V. R. Ravi, "Improving visual evoked potential feature classification for person recognition using PCA and normalization," *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 726–733, May 2006.
- [141] C. Chester and H. T. Maecker, "High-Dimensional Cytometry Data Algorithmic Tools for Mining," 2015.
- [142] K. H. Esbensen and B. Swarbrick, "Multivariate Data Analysis 6th Edition An introduction to Multivariate Analysis, Process Analytical Technology and Quality by Design," 2018.
- [143] "Assumptions of Linear Regression - Statistics Solutions." [Online]. Available: <https://www.statisticssolutions.com/assumptions-of-linear-regression/>. [Accessed: 11-Feb-2021].
- [144] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
- [145] M. T. Puth, M. Neuhäuser, and G. D. Ruxton, "Effective use of Pearson's product-moment correlation coefficient," *Animal Behaviour*, vol. 93. Academic Press, pp. 183–189, 01-Jul-2014.
- [146] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer New York, 1995.
- [147] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3. Kluwer Academic Publishers, pp. 199–222, Aug-2004.
- [148] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, Nov. 1958.



- [149] H. Ramchoun, M. Amine, and J. Idrissi, "Multilayer Perceptron: Architecture Optimization and Training multi-criteria learning and nonlinear optimization View project," *Artic. Int. J. Interact. Multimed. Artif. Intell.*, vol. 4, pp. 1–26, 2016.
- [150] B. H. Menze, B. M. Kelm, D. N. Splitthoff, U. Koethe, and F. A. Hamprecht, "On oblique random forests," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011, vol. 6912 LNAI, no. PART 2, pp. 453–469.
- [151] L. Breiman, "Random Forests," 2001.
- [152] D. L. Olson and D. Wu, "Regression Tree Models," Springer, Singapore, 2020, pp. 57–77.
- [153] T. Hastie, R. Tibshirani, and J. Friedman, "Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction."
- [154] M. Bramer, *Principles of Data Mining*. London: Springer London, 2020.
- [155] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.
- [156] C. Sutton, M. Sindelar, and A. McCallum, "Feature Bagging: Preventing Weight Undertraining in Structured Discriminative Learning," 2005.
- [157] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, 1998.
- [158] "dataists » A Taxonomy of Data Science." [Online]. Available: <http://www.dataists.com/2010/09/a-taxonomy-of-data-science/>. [Accessed: 13-Aug-2020].
- [159] J. L. Schafer, "Multiple imputation: a primer," *Stat. Methods Med. Res.*, vol. 8, no. 1, pp. 3–15, Feb. 1999.
- [160] R. J. A. Little and D. B. Rubin, "STATISTICAL ANALYSIS WITH MISSING DATA."
- [161] S. Zhang, "Nearest neighbor selection for iteratively kNN imputation," *J. Syst. Softw.*, vol. 85, no. 11, pp. 2541–2552, Nov. 2012.
- [162] S. van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in R," *J. Stat. Softw.*, vol. 45, no. 3, pp. 1–67, Dec. 2011.
- [163] T. De Waal, J. Pannekoek, and S. Scholtus, "Statistical Data Editing and Imputation."
- [164] R. Alex Marsh, J. M. Hodgkiss, S. Albert-Seifried, and R. H. Friend, "Effect of annealing on P3HT:PCBM charge transfer and nanoscale morphology probed by ultrafast spectroscopy," *Nano Lett.*, vol. 10, no. 3, pp. 923–930, Mar. 2010.
- [165] C. Zhong, F. Huang, Y. Cao, D. Moses, and A. J. Heeger, "Role of Localized States on Carrier Transport in Bulk Heterojunction Materials Comprised of Organic Small Molecule Donors," *Adv. Mater.*, vol. 26, no. 15, pp. 2341–2345, Apr. 2014.
- [166] J. Kettle, V. Stoichkov, D. Kumar, M. Corazza, S. A. Gevorgyan, and F. C. Krebs, "Using ISOS consensus test protocols for development of quantitative life test

- models in ageing of organic solar cells," *Sol. Energy Mater. Sol. Cells*, vol. 167, pp. 53–59, Aug. 2017.
- [167] M. Kaltenbrunner *et al.*, "Flexible high power-per-weight perovskite solar cells with chromium oxide-metal contacts for improved stability in air," *Nat. Mater.*, vol. 14, no. 10, pp. 1032–1039, Oct. 2015.
- [168] T.-M. Kim, J. W. Kim, H.-S. Shim, and J.-J. Kim, "High efficiency and high photo-stability zinc-phthalocyanine based planar heterojunction solar cells with a double interfacial layer ARTICLES YOU MAY BE INTERESTED IN," *Cite as Appl. Phys. Lett.*, vol. 101, p. 113301, 2012.
- [169] A. Hayakawa, M. Yukawa, and T. Sagawa, "Stability Improvement of Photovoltaic Performance in Antimony Sulfide-Based Hybrid Solar Cells," *ECS J. Solid State Sci. Technol.*, vol. 6, no. 4, pp. Q35–Q38, Feb. 2017.
- [170] S. Alem, J. Lu, R. Movileanu, T. Kololuoma, A. Dadvand, and Y. Tao, "Solution-processed annealing-free ZnO nanoparticles for stable inverted organic solar cells," *Org. Electron.*, vol. 15, no. 5, pp. 1035–1042, May 2014.
- [171] M. S. Salem *et al.*, "Physically Based Analytical Model of Heavily Doped Silicon Wafers Based Proposed Solar Cell Microstructure," *IEEE Access*, vol. 8, pp. 138898–138906, 2020.
- [172] A. Sharma, S. Zivanovic, S. R. Animilli, and D. A. Genov, "Analytical Model of Thin-Film Silicon Solar Cell," 2015.
- [173] M. Corazza, F. C. Krebs, and S. A. Gevorgyan, "Lifetime of organic photovoltaics: Linking outdoor and indoor tests," *Sol. Energy Mater. Sol. Cells*, vol. 143, pp. 467–472, Aug. 2015.
- [174] M. K. Behera, I. Majumder, and N. Nayak, "Solar photovoltaic power forecasting using optimized modified extreme learning machine technique," *Eng. Sci. Technol. an Int. J.*, vol. 21, no. 3, pp. 428–438, Jun. 2018.
- [175] Z. Zhou and M. Carbajales-Dale, "Assessing the photovoltaic technology landscape: Efficiency and energy return on investment (EROI)," *Energy Environ. Sci.*, vol. 11, no. 3, pp. 603–608, Mar. 2018.
- [176] O. A. Alduchov and R. E. Eskridge, "Improved Magnus Form Approximation of Saturation Vapor Pressure," *J. Appl. Meteorol.*, vol. 35, no. 4, pp. 601–609, Apr. 1996.
- [177] "SHTxx Application Note Dew-point Calculation."
- [178] C. H. Peters *et al.*, "The mechanism of burn-in loss in a high efficiency polymer solar cell," *Adv. Mater.*, vol. 24, no. 5, pp. 663–668, Feb. 2012.
- [179] V. Stoichkov, N. Bristow, J. Troughton, F. De Rossi, T. M. Watson, and J. Kettle, "Outdoor performance monitoring of perovskite solar cell mini-modules: Diurnal performance, observance of reversible degradation and variation with climatic performance," *Sol. Energy*, vol. 170, pp. 549–556, Aug. 2018.
- [180] M. V Khenkin *et al.*, "Dynamics of Photoinduced Degradation of Perovskite Photovoltaics: From Reversible to Irreversible Processes," 2018.

- [181] B. Azzopardi, C. J. M. Emmott, A. Urbina, F. C. Krebs, J. Mutale, and J. Nelson, "Economic assessment of solar electricity production from organic-based photovoltaic modules in a domestic environment," *Energy Environ. Sci.*, vol. 4, no. 10, pp. 3741–3753, Oct. 2011.
- [182] C. J. Mulligan, C. Bilen, X. Zhou, W. J. Belcher, and P. C. Dastoor, "Levelised cost of electricity for organic photovoltaics," *Sol. Energy Mater. Sol. Cells*, vol. 133, pp. 26–31, Feb. 2015.
- [183] F. Machui *et al.*, "Cost analysis of roll-to-roll fabricated ITO free single and tandem organic solar modules based on data from manufacture," *Energy Environ. Sci.*, vol. 7, no. 9, pp. 2792–2802, Aug. 2014.
- [184] M. Mrinalini, N. Islavath, S. Prasanthkumar, and L. Giribabu, "Stipulating Low Production Cost Solar Cells All Set to Retail...!", *Chem. Rec.*, vol. 19, no. 2–3, pp. 661–674, Feb. 2019.
- [185] T. W. David, H. Anizelli, T. J. Jacobsson, C. Gray, W. Teahan, and J. Kettle, "Enhancing the stability of organic photovoltaics through machine learning," *Nano Energy*, vol. 78, p. 105342, Dec. 2020.
- [186] M. Kumar, M. Husain, N. Upreti, and D. Gupta, "Genetic Algorithm: Review and Application," *SSRN Electron. J.*, Mar. 2020.
- [187] N. Espinosa and F. C. Krebs, "Life cycle analysis of organic tandem solar cells: When are they warranted?," *Sol. Energy Mater. Sol. Cells*, vol. 120, no. PART B, pp. 692–700, Jan. 2014.
- [188] N. Espinosa *et al.*, "Life cycle assessment of ITO-free flexible polymer solar cells prepared by roll-to-roll coating and printing," in *Solar Energy Materials and Solar Cells*, 2012, vol. 97, pp. 3–13.
- [189] M. Krebs-Moberg, M. Pitz, T. L. Dorsette, and S. H. Gheewala, "Third generation of photovoltaic panels: A life cycle assessment," *Renew. Energy*, vol. 164, pp. 556–565, Feb. 2021.
- [190] P. Stiebitz, "Hyperspectral Polymer Solar Cells, Integrated Power for Microsystems," Golden, CO (United States), May 2014.
- [191] U. Maulik and S. Bandyopadhyay, "Genetic algorithm-based clustering technique," *Pattern Recognit.*, vol. 33, no. 9, pp. 1455–1465, Sep. 2000.
- [192] M. A. Albadr, S. Tiun, M. Ayob, and F. AL-Dhief, "Genetic Algorithm Based on Natural Selection Theory for Optimization Problems," *Symmetry (Basel)*, vol. 12, no. 11, p. 1758, Oct. 2020.
- [193] M. Konstantakou, D. Perganti, P. Falaras, and T. Stergiopoulos, "Anti-solvent crystallization strategies for highly efficient perovskite solar cells," *Crystals*, vol. 7, no. 10. MDPI AG, p. 291, 01-Oct-2017.
- [194] X. Jiang *et al.*, "Efficient Compact-Layer-Free, Hole-Conductor-Free, Fully Printable Mesoscopic Perovskite Solar Cell," *J. Phys. Chem. Lett.*, vol. 7, no. 20, pp. 4142–4146, Oct. 2016.
- [195] F. De Rossi *et al.*, "An Interlaboratory Study on the Stability of All-Printable Hole

- Transport Material-Free Perovskite Solar Cells," *Energy Technol.*, p. 2000134, May 2020.
- [196] A. Pockett, D. Raptis, S. M. P. Meroni, J. Baker, T. Watson, and M. Carnie, "Origin of Exceptionally Slow Light Soaking Effect in Mesoporous Carbon Perovskite Solar Cells with AVA Additive," *J. Phys. Chem. C*, vol. 123, no. 18, pp. 11414–11421, May 2019.
- [197] H. Lakhiani *et al.*, "Variations of Infiltration and Electronic Contact in Mesoscopic Perovskite Solar Cells Revealed by High-Resolution Multi-Mapping Techniques," *Adv. Funct. Mater.*, vol. 29, no. 25, p. 1900885, Jun. 2019.
- [198] D. Yang *et al.*, "High efficiency planar-type perovskite solar cells with negligible hysteresis using EDTA-complexed SnO<sub>2</sub>," *Nat. Commun.*, vol. 9, no. 1, pp. 1–11, Dec. 2018.
- [199] J. Kniepert, M. Schubert, J. C. Blakesley, and D. Neher, "Photogeneration and recombination in P3HT/PCBM solar cells probed by time-delayed collection field experiments," *J. Phys. Chem. Lett.*, vol. 2, no. 7, pp. 700–705, Apr. 2011.
- [200] A. K. K. Kyaw *et al.*, "Intensity dependence of current-voltage characteristics and recombination in high-efficiency solution-processed small-molecule solar cells," *ACS Nano*, vol. 7, no. 5, pp. 4569–4577, 2013.
- [201] D. Bogachuk *et al.*, "Low-temperature carbon-based electrodes in perovskite solar cells," *Energy and Environmental Science*, vol. 13, no. 11. Royal Society of Chemistry, pp. 3880–3916, 01-Nov-2020.
- [202] E. V. Péan *et al.*, "Investigating the Superoxide Formation and Stability in Mesoporous Carbon Perovskite Solar Cells with an Aminovaleric Acid Additive," *Adv. Funct. Mater.*, vol. 30, no. 12, p. 1909839, Mar. 2020.

# Appendix A Enhancing the Indoor Performance and Stability of Emerging Photovoltaics through Machine Learning – Additional Information

In total, the dataset contains 17 attributes, each of which contains a number of categories. These are detailed in Table A.1.

Type	Materials	Configuration	Encapsulation	ISOS compatability	Conditions	Intensity	Temperature	Light type	Structure
OPV	pol	Normal	none	ISOS-D-1	indoor	0	30	N/A	Glass
Module	Small molecule	inverted	flexible	ISOS-D-2	inert	80	60	HL	PET
		tandem	rigid	ISOS-D-3	vacuum	100	110	XE	Quartz
			N/A	ISOS-L-1	N/A	300	40	LED	Unknown
				ISOS-L-2	outdoor	500	85	FL	Other
				ISOS-L-3	low_light	600	22.5	Sun	PEN
				ISOS-O-1		800		SPL	
				ISOS-O-2		1000			
				ISOS-O-3		1630			
				ISOS-T-3		1710			
				ISOS not compatible		2000			
						2670			
						2700			
						2810			
						4000			
						6580			
						8700			
						9990			
Elec1	TL1	Active1			Active2	TL2	Elec2	Encap material	
ITO	PEDOT:PSS (Clevios)			P3HT	PCBM	Unknown	Al	None	
Ag grid	PEDOT:PSS			CuPc	C60	Ca	Ag grid	PET	
Ag grid inkjet	PEDOT:PSSBitronLowConductive			MDMO-PPV	Other	LiF	Ag	glass	
Ag	None			P3HT-Br	PCBM-71	Yb	Unknown	Barrier layer	
Ag grid-flexo	PEDOT:PSS (Baytron)			Other	OXCBA	PEDOT:PSSAgfa5010	Au		
Cr/Al	ZnO			Unknown	Unknown	AlQ3	Other		
Other	PEDOT:PSS (Bayer)			PCDTBT	None	SAM			
Ag-evaporated	MoO3			C6PcH2	ICBA	Other			
FTO	PEDOT:PSS (H.C. Starck)			PTB7	PCBM-61	MoO3			
	AZO			Pentacene	PCBM-slot-die	None			
	ZnO- spray			P3HT-slot die	PTCDI	PEDOT:PSS			
	Unknown			PBDTTT-C		TiOx			
	CF4			PV-D4610		TiO2 NPs			
	Metal Oxide			Benzodithiophene terthiophene rhodanine (BTR)		V2O5			
	Graphene Oxide			N(Ph-2T-DCN-Et)3		PEDOT:PSS (Baytron P VP A14083)			
	Cs2CO3			P3HT+additive		PEDOT:PSS (Clevios P VP A1 4083)			
	HPS-Au (nanoparticles)			ClInPc		BCP			
	Other			ZnPc		ZnO			
	TiOx			ZnPc+NDP2(Novaled) doped DiNPB		BPhen			
	ZnO - slot die			Bithiophene Imide and Benzodithiophene based Copolymer		CrOx			
	PEDOT:PSS (Heraeus PH1000)			PBDTTTz-4		ZnO - slot die			
	TiO2			PECz-DTQx		CuOx			
	V2O5			PDTSTTz-4		PEDOT:PSS			
	PEDOT:PSSAgfa5010HighConductivity			PDTTBT-3		ZnO			
	CrO2			P3HT - doctor		ZnO NPs			
	PS nanospheres			PC-TBT-TQ		BaF2			
	Au								
	NiOx								
	PEIE								
	LZO								
	Ammonium heptamolybdate								
	WO3								
	Ca								
	BF-DPB								
	NDP2(Novaled) doped DiNPB								
	F6TCNNQ								
	Zn1-xMgxO								
	NDP2(Novaled) doped PV-TPD								

Table A.1: Features within each attribute in dataset.

## Appendix B Life Cycle Assessment of Organic Photovoltaics with the application of Genetic Clustering Algorithms – Additional Information

Table B.1 gives the embodied energy of each material and Table B.2 gives the embodied energy of each process.

Material	Embodied Energy (MJm <sup>-2</sup> )
Glass	208.6076667
PET	16.08466667
Quartz	720
Silicon	2488
Ag	2.08
Ag	7.488
Ag	2.450777
ITO	224.3233333
Cr Aluminum	320.24
FTO	54.02
AlQ3	2.325294118
BF-DPB	2.325294118
C60F36	2.325294118
Cr	0.09
CrO2	0.65
CuI	0.65
Ethanolamine	2.325294118
Graphene Oxide	2.759
MeO-TPD	0.00041234
MoO3	17.56
NiOx	131
PANI/PSS	3.54
PEDOT:PSS	3.54
PEDOT:PSS (Baytron)	3.54
PEDOT:PSS (Clevios)	3.54
PEDOT:PSS (H.C. Starck)	3.54
PEDOT:PSS (Heraeus PH1000)	3.54
PEDOT:PSS(Bayer)	3.54
PEDOT:PSSAgfa5010HighConductivity	3.54
PEDOT:PSSAldrichLowConductivity	3.54
PEDOT:PSSBitronLowConductive	3.54
Plexcore HTL	3.54

PSS-g-PANI/PFI	3.54
P-type/HTL	2.325294118
s-MoO <sub>x</sub>	18.89658333
TiO <sub>2</sub>	35.9
TiO <sub>2</sub>	35.9
W <sub>ox</sub>	18.89658333
X	2.325294118
ZnO NP	0.58
ZnO	0.435
ZnO	0.29
ZnO	0.435
AZO	0.58
2PACz	0.058654415
ATO	0.464
n-type semiconducting oxide	0.58
IZO	224.3233333
Bithiophene Imide and Benzodithiophene based Copolymer P2	0.967867907
C6PcH <sub>2</sub>	0.967867907
CN-P3HT	0.38
Crosslinked PTM21	0.02418339
Cr-PCPDTBT:PC71BM:ODT	4.675163398
LBP	0.967867907
MEH-PPV	0.967867907
MEH-PPV	0.967867907
NDP2(Novaled) doped DiNPB	0.967867907
P3HT (15% of ester functionalised side chain thiophene units)	0.38
P3HT	0.38
P3HT	0.38
P3HT	0.38
P3HT-Au (doped with Au nanoparticles)	0.38
P3MHOCT	0.276597427
P3OT	0.38
PBDTTTz-4	0.079333928
PCPDTBT:C71-X	0.967867907
PCPDTBT:PC71BM:ODT	2.51120225
PCPDTBT	1.3425
PDTSTTz-4	0.908833697
PDTTDABT	1.012225177
Pentacene	0.908833697
PININE-DTBT	1.012225177
POHT	0.38
polymer donor	1.012225177
PPE-PPV	0.276597427

PTB7:PC71BM	1.036006515
PTM10	1.012225177
PV-D4610	0.276597427
RISO GREEN 1 (LBP)	0.967867907
RISO GREEN 2 (LBP)	0.967867907
SubPc (subphthalocyanine chloride)	1.012225177
Tetracene	1.012225177
TDCV-TPA	0.967867907
ZnPc	1.012225177
Al NPs	0.24
DPPTPTA	0.967867907
DR3TSBDT	0.276597427
DR3TSBDT-y	0.276597427
DRCN5T	0.276597427
DRCN5T-p	0.276597427
DRCN7T	0.276597427
IDTT	1.012225177
NF3000-N	0.276597427
PBDB-7	1.012225177
PBDB-T	1.012225177
PBDTTT	1.012225177
PBDTTT-EFT	1.012225177
PBDTTTPD	1.012225177
PCDTBT	1.012225177
PCE10	1.012225177
PCE11	1.012225177
PCE12	1.012225177
PCPDTBT	1.3425
PCPDTBT:C71-X	1.3425
PDPP4T-2F	1.3425
PFDT2BT	1.3425
pffBt4T	1.3425
PM6	1.3425
PM7	1.3425
P-SBTBDT	1.3425
PSBTBT	1.3425
PTB7-Th	1.036006515
C60	8.7
C60	8.7
C71-X	3.18375
CuPc	3.05
F16CuPc (Copper hexadecafluorophthalocyanine)	0.00357392
ICBA	0.011385
PBDTTTz-4	0.010363386



PC61BM	1.07
PCBM	1.07
PCBM	1.07
PCBM[70]	1.07
SiNp	0.1332
SiNp-OA (modified with octanoic acid)	0.1332
ZnPc	0.0267375
Alq3	3.18375
SubPc	1.012225177
CFO:PCBM	1.07
ICBA	2.56355939
ICxA	2.56355939
IDFBR	2.56355939
IDTBR	2.56355939
IDTBR:IDFBR	2.56355939
IEICO-4F:Cross-linked P2FBTT-Br	2.56355939
IEICO-4F:P2FBTT-H	2.56355939
IT-4F	2.56355939
ITDBR	2.56355939
ITIC	2.56355939
N3	2.56355939
NF3000-N	2.56355939
PC71BM:Cross-linked P2FBTT-Br	0.74
PC71BM:P2FBTT-H	0.74
Alq3	3.18375
BaF2	6.670630895
BaF2	6.670630895
Bathocuproine	6.670630895
Bphen	6.13
Ca	6.09
Cr	3.12
LiF	3.89
MEH-PPV:PCBM	0.967867907
MoO3	18.84825
PDTSTTz-4	6.670630895
PEDOT:PSS	4.46
PEDOT:PSS	4.46
PEDOT:PSS (Baytron P)	4.46
PEDOT:PSS (Clevios P VP A1 4083)	4.46
PEDOT:PSS (EL-P 5010 from Agfa)	4.46
PEDOT:PSS (H.C. Starck)	4.46
PEDOT:PSSAgfa5010HighConductivity	4.46
PEDOT:PSSAgfa5010HighConductivity	4.46
p-type organic semiconductor	6.670630895

SnCl <sub>2</sub> Pc	6.670630895
SPP01	6.670630895
TiO <sub>x</sub>	18.84825
V2O5	12.89
V2O5	12.89
ZnO	0.58
HTL Solar	6.670630895
PEI	6.670630895
PEIE	6.670630895
PFn	6.670630895
TaSiW-12	18.84825
ZrO	18.84825
AZO	0.58
Ag	4.9
Ag	2.452900485
Al	0.238085
Au	2.122746371
Carbon	0.9
CPP PEDOT:PSS	4.46
UV encapsulant	8.19
PET back sheet	5.89

*Table B.1: Embodied Energies of materials in dataset.*

<b>Deposition Process</b>	<b>Embodied Energy (MJm<sup>-2</sup>)</b>
Sintering metal oxide	5.5
Screen printing	3.85E-02
Spray of blocking layer	3.44E-03
Spin coating	8.08E-01
Encapsulation	1.48E-01
Ultrasonic cleaning	1.74
Sintering organic	1.28
Metal evaporation	1.19
Plasma cleaning	0.08
Photonic sintering	0.125

*Table B.1: Embodied Energies of processing methods.*

## Appendix C Additional Studies – HTL and Carbon-based PSCs.

### C.1 Methods: Inverted PSC fabrication

The following sections outline the methods for fabricating inverted PSC devices.

#### *Substrate Cleaning*

Process performed in cleanroom conditions. All PSC devices were prepared on glass/ITO (15  $\Omega$ /square) substrates which were first washed in Deacon90 detergent and then sequentially cleaned *via* ultrasonication in deionised water, acetone and isopropyl alcohol (IPA) for 10 minutes each. The substrates were subsequently treated with oxygen plasma for 5 minutes to remove any residual organic compounds from the surface and to enhance the wettability by modifying the surface energy of the ITO.

#### *Solution Preparation*

Processes was performed in cleanroom conditions and formulations are as described below.

**PEDOT:PSS:** Heraeus Clevios™ AI 4083, bought commercially from Ossila, was filtered using a 1  $\mu$ m filter and stirred using magnetic beads for a minimum of 30 minutes prior to deposition.

**Copper Iodide (CuI):** 20 mg of solid CuI was dissolved in 1 ml of acetonitrile and stirred using magnetic beads overnight in the glovebox prior to deposition.

**Copper(I) Thiocyanate (CuSCN):** 20 mg of solid CuSCN was dissolved in 1 ml of aqueous ammonia and stirred using magnetic beads for 1 hour at 50°C prior to deposition.

**Nickel Oxide (NiO<sub>x</sub>):** 182.7 mg of solid Nickel(II) Nitrate Hexahydrate was dissolved in 1 ml of ethylene glycol with 60  $\mu$ l of ethylenediamine and stirred using magnetic beads for 3 hours at 70°C prior to deposition.

**Phenyl-C61-butyric acid methyl ester (PCBM):** 20 mg of PCBM was dissolved in 1 ml of chlorobenzene and filtered using a 0.45  $\mu\text{m}$  filter and stirred using magnetic beads overnight prior to deposition.

**Perovskite Solution:**  $\text{FA}_{0.7}\text{MA}_{0.2}\text{Cs}_{0.1}\text{Pb}(\text{I}_{5/6}\text{Br}_{1/6})_3$  (FMC) was prepared by dissolving 26 mg of Caesium Iodide (CsI), 32 mg of Methylammonium Iodide (MAI), 92 mg of lead bromide ( $\text{PbBr}_2$ ), 121 mg of formamidinium iodide (FAI) and 346 mg of lead iodide ( $\text{PbI}_2$ ) in 1 ml of N,N-Dimethylformamide (DMF). The solution was subsequently stirred using magnetic beads overnight and at 70°C for 20 minutes prior to deposition.

#### *Deposition Techniques*

Solution/Material	Method	Rate	Annealing
PEDOT:PSS	Spin – coating	4000rpm for 40s	150°C for 20min
CuI	Spin – coating	3500rpm for 60s	100°C for 20min
CuSCN	Spin – coating	3500rpm for 60s	120°C for 20min
$\text{NiO}_x$	Spin – coating	2000rpm for 90s	300°C for 60min
PCBM	Spin – coating	2000rpm for 30s	No
Perovskite	Spin – coating	5000rpm for 20s	100°C for 60min
LiF	Thermal Evaporation	$0.1 \text{ \AA s}^{-1}$	No
$\text{V}_2\text{O}_5$	Thermal Evaporation	$1 \text{ \AA s}^{-1}$	No
$\text{MoO}_3$	Thermal Evaporation	$1 \text{ \AA s}^{-1}$	No
Ag	Thermal Evaporation	$1 \text{ \AA s}^{-1}$	No

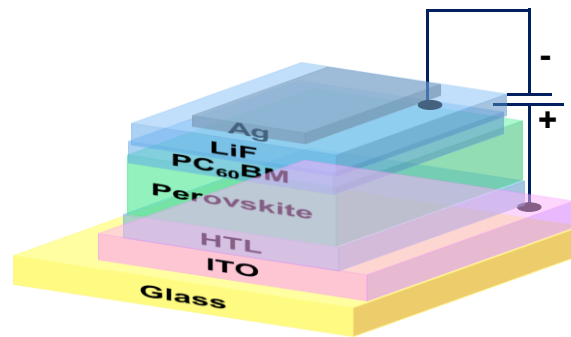
*Table C.1: Solutions and materials used for PSC fabrication showing deposition methods.*

Each layer was deposited as described in Table C.1 which shows the deposition method, and deposition parameters. In addition, for the perovskite deposition, the antisolvent treatment method was employed [193]. Toluene was dispersed on the spinning substrate

4s after the deposition of the perovskite precursor in order to expediate and seed the crystal growth of the perovskite crystal.

### *Device Architecture*

PSC devices were fabricated with the architecture shown in Figure C.1. This corresponds to the inverted architecture for PSCs. Each layer was deposited as described in Table C.1 and using the formulations described above. Thermal evaporation of the metal contacts was performed using a Kurt J. Lesker thermal evaporator.



*Figure C.1: Device architecture for inverted PSC with varying HTL materials.*

## C.2 Methods: Carbon Based PSC fabrication

Intercalated carbon – based PSCs were fabricated by de Rossi *et al.* at Swansea university. The fabrication was performed as follows. FTO glass substrates were initially etched with a Rofin Nd:YVO<sub>4</sub> laser at a rate of 150 mms<sup>-1</sup>, washed using Hellmanex solution in deionised water, pure deionised water and rinsed using acetone and isopropanol. The substrates were subsequently treated with O<sub>2</sub> plasma. 50 nm of compact TiO<sub>2</sub> was deposited *via* spray – pyrolysis at 300°C from a 1:9 solution of titanium di-isopropoxide bis(acetylacetonate) (Sigma) in isopropanol. The triple mesoporous stack was deposited *via* screen printing of commercially bought pastes: 1) TiO<sub>2</sub> layer (30 NRD Dyesol diluted 1:1 by weight with terpineol, and sintered at 550°C for 30 minutes, 2) ZrO<sub>2</sub> layer (Solaronix), sintered at 400°C for 30 minutes, 3) carbon layer (Gwent Electronic Materials), sintered at 400°C for 30 minutes. A solution of AVAI, MAI and PbI<sub>2</sub> in gamma-

butyrolactone (GBL) was prepared using the method presented by Jiang *et al.* [194] and infiltrated through the top carbon electrode, percolated throughout the mesoporous stack to the compact  $\text{TiO}_2$ . The perovskite solution subsequently filled the pores and crystallised within the stack when annealed at  $50^\circ\text{C}$  for 1 hour. Silver paint was applied to the contacts and black tape applied to the glass side giving an aperture of  $0.5\text{cm}^2$  [195]. A schematic of the device configuration is shown in Figure C.2. Once received, flat, low resistance contacts were ultrasonically soldered to allow the metal contacts to be soldered to the non-metal surface and is performed flux-free. Flat contacts were used to improve the bonding of the wires with the electrodes.

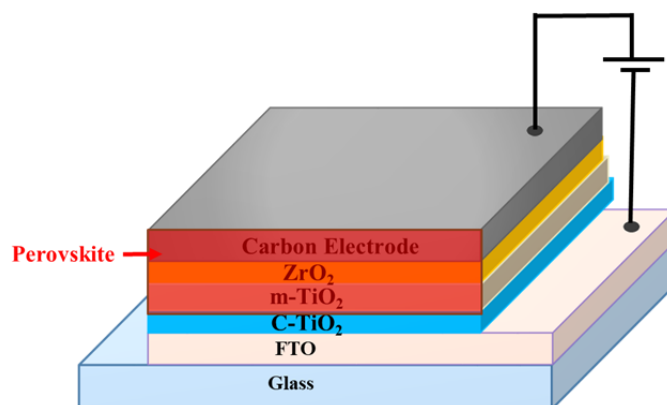


Figure C.2: Device architecture for intercalated carbon perovskite solar cell.

### C.3 Results and Discussion – HTL Screening

Initially the morphology and optical properties of the FMC perovskite film deposited on each of the six different HTLs is identified. This is achieved through the use of SEM imaging and X-ray diffraction spectroscopy (XRD) of perovskite films deposited on each of the six HTLs. Figure C.3 shows the SEM images, taken over  $10\text{ }\mu\text{m}$ , for each FMC film deposited on the six HTLs: (a) PEDOT:PSS, (b)  $\text{MoO}_3$ /PEDOT:PSS, (c)  $\text{CuSCN}$ , (d)  $\text{CuI}$ , (e)  $\text{NiO}_x$  and (f)  $\text{V}_2\text{O}_5$ .

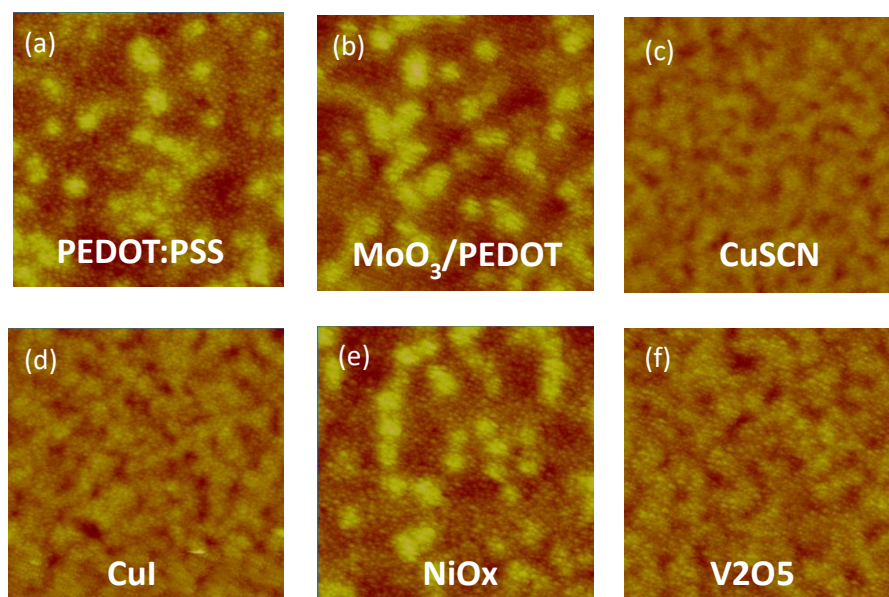


Figure C.3: SEM images for FMC films deposited on (a) PEDOT:PSS, (b)  $\text{MoO}_3/\text{PEDOT}$ , (c)  $\text{CuSCN}$ , (d)  $\text{CuI}$ , (e)  $\text{NiO}_x$  and (f)  $\text{V}_2\text{O}_5$ .

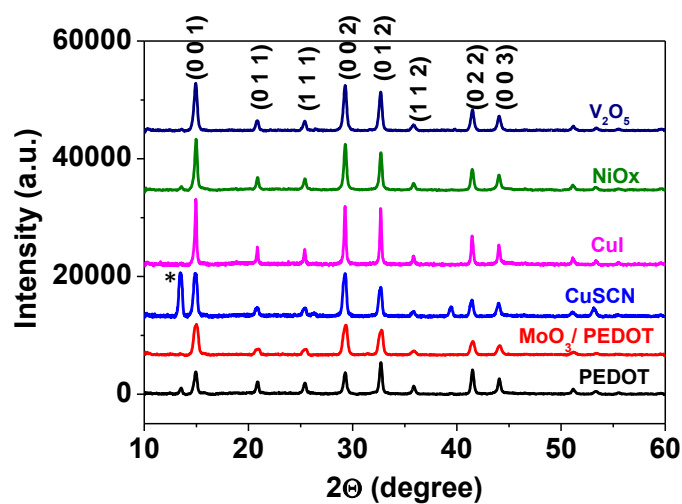


Figure C.4: XRD spectra for FMC films deposited on PEDOT:PSS,  $\text{MoO}_3/\text{PEDOT}$ ,  $\text{CuSCN}$ ,  $\text{CuI}$ ,  $\text{NiO}_x$  and  $\text{V}_2\text{O}_5$ .

The SEM images illustrate the film quality when utilising each of the HTLs. As can be seen, FMC films deposited on  $\text{CuSCN}$ ,  $\text{CuI}$  and  $\text{V}_2\text{O}_5$  display a smoother film morphology

as compared to films deposited on PEDOT:PSS,  $\text{MoO}_3/\text{PEDOT:PSS}$  and  $\text{NiO}_x$ . Furthermore, the XRD spectra are illustrated in Figure C.4 for each of the six HTLs. All peaks associated with the FMC film are identified. However, for the film deposited on CuSCN, a prominent peak (labelled by \*) can be observed. This peak corresponds to the presence of non-stoichiometric, excess  $\text{PbI}_2$  which has not been incorporated into the Perovskite structure. This excess  $\text{PbI}_2$  can also be observed, to a significantly lesser degree, in the XRD spectra of PEDOT:PSS and  $\text{NiO}_x$ . Interestingly, the inclusion of  $\text{MoO}_3$  with PEDOT:PSS suppresses this additional  $\text{PbI}_2$  peak.

The photovoltaic performance when using each HTL material and the J-V curves are shown in Figure C.5. Table C.1 displays each of the performance parameters ( $J_{sc}$ ,  $V_{oc}$ , FF and PCE) for the champion devices using each HTL material. The average value between the reverse and forward scans are shown. Comparison of the performance parameters for each device shows that the device utilising  $\text{NiO}_x$  as the HTL leads to the best PCE, followed by PEDOT:PSS. The worst PCE performance is seen for the device utilising  $\text{V}_2\text{O}_5$  as the HTL.

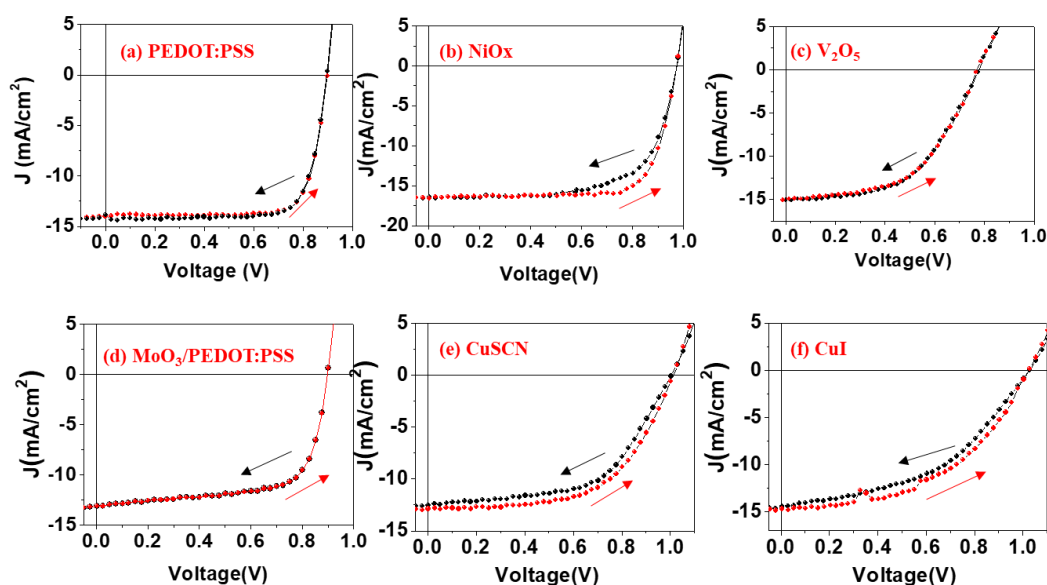


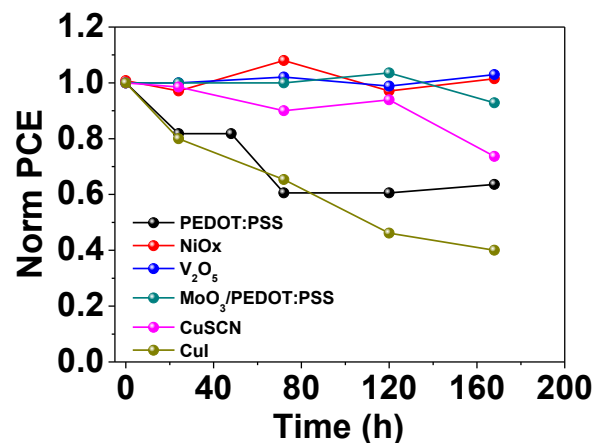
Figure C.5: Photovoltaic performances for each device architecture using (a) PEDOT:PSS, (b)  $\text{NiO}_x$ , (c)  $\text{V}_2\text{O}_5$ , (d)  $\text{MoO}_3/\text{PEDOT:PSS}$ , (e) CuSCN and (f) CuI as HTLs. Reverse scan is shown in black and forward scan is shown in red.



	$J_{sc}$ (mA/cm <sup>2</sup> )	$V_{oc}$ (V)	FF	PCE
PEDOT:PSS	14.2	0.89	79.3 %	10.07 %
NiO <sub>x</sub>	16.5	0.97	74 %	12.0 %
V <sub>2</sub> O <sub>5</sub>	15	0.78	53%	6.3%
MoO <sub>3</sub> /PEDOT:PSS	13	0.90	68%	8.1%
CuSCN	12.4	1.0	60.5 %	7.6 %
CuI	14.8	1.0	49.2 %	7.2 %

*Table C.1: Photovoltaic performance parameters ( $J_{sc}$ ,  $V_{oc}$ , FF and PCE) for each device employing the stated HTL.*

Finally, the dark storage stability of the devices is tested. Figure C.6 illustrates the dark storage stability of each device type, where the normalised PCE is plotted as a function of time. The dark storage test is performed over the course of 170 hours and allows the discrepancy between the 5 different device types to be identified. It is seen that devices using metal oxide HTL display significantly improved stabilities with NiO<sub>x</sub>, V<sub>2</sub>O<sub>5</sub> and MoO<sub>3</sub>/PEDOT:PSS devices maintaining their initial efficiencies for the full test duration. CuSCN, PEDOT:PSS and CuI devices display correspondingly lower lifetimes and performance retention.



*Figure C.6: Dark storage stability of each device type, where normalised PCE is plotted as a function of time.*

Comparison of the stability data, illustrated in Figure C.6 and the performance metrics shown in Table C.1, shows that both the best and worst performing devices lead to the longest device lifetimes. In addition, PEDOT:PSS devices, displaying the second best performance, along with CuI devices, display significantly poorer stabilities. In addition, the inclusion of MoO<sub>3</sub>, combined with PEDOT:PSS, significantly improves the stability, as compared to PEDOT:PSS used individually. The stability results shown in Figure C.6 identifies metal oxides as significant factors for improving the lifetime and stability of PSCs, whilst also improving the performance as is shown in the case of devices employing NiO<sub>x</sub> HTLs.

## C.4 Results and Discussion – Intercalated Carbon-Based Perovskites

### C.4.1 Device Characterisation

In contrast to investigating the role of HTLs in PSC performance and stability, intercalated carbon-based PSCs performance and stability was also conducted. Figure C.7 shows an image of a device with contacts soldered on.

The devices were initially characterised using J-V measurements which do not utilise an HTL material. The devices demonstrated an interesting phenomenon, whereby the efficiency of the devices increased with light soaking, thus demonstrating the need for preconditioning. Figure C.8 illustrates the J-V characteristics at the initial instance of illumination and after 10 minutes of light soaking. After exactly 10 minutes of light soaking at 1 Sun, the PCE is observed to increase from 4.10% to a maximum value of 7.05% with the FF increasing from 31.8% to 44.0%.

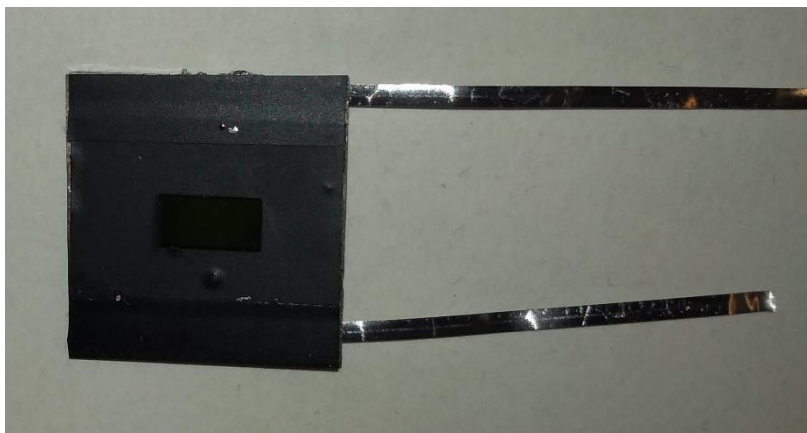


Figure C.7: Image of intercalated carbon – based PSC with contacts soldered on.

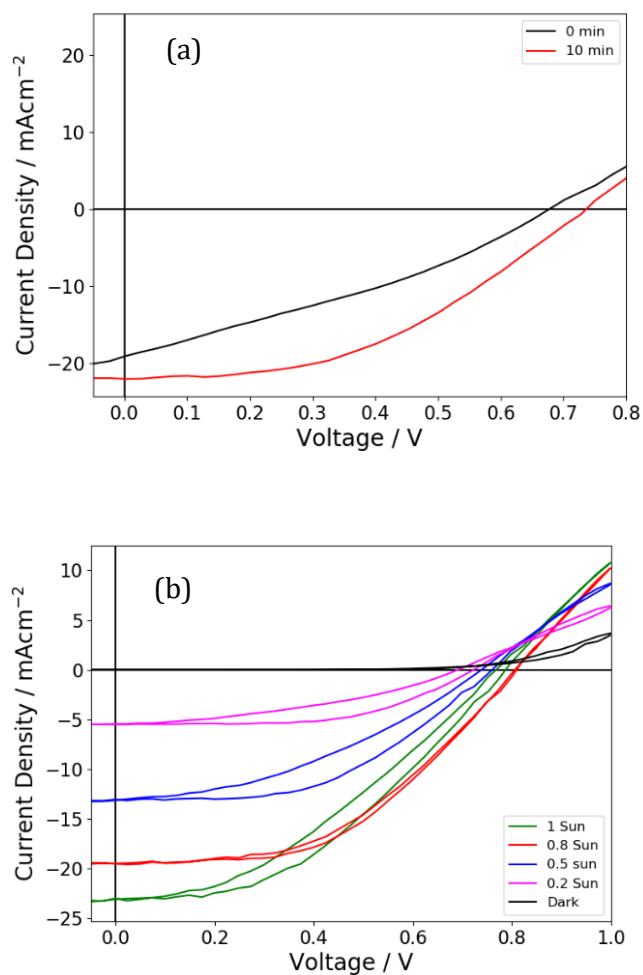


Figure C.8: (a) J-V characteristics of intercalated carbon based at the initial instance of illumination and after 10 minutes of light soaking. (b) Current – voltage characteristics of intercalated carbon-based PSC under varying light intensity.

The origin of the slow response under light soaking in Carbon based perovskites was investigated by Pockett *et al.* and attributed to the presence of the AVA additive, which could restrict ion migration *via* a 2D perovskite region [196]. The cause of the poor initial FF displayed for these types of devices could result from their structural characteristics; the intercalation could result in poor interfacial layers forming between respective components of the cell, but this feature could also result in the relatively large  $J_{sc}$  observed at 0 minutes, since it could enable more efficient charge collection at the interface. The effect of infiltration into the stack is discussed in greater detail by Lakhiani *et al.* [197]. The increase in performance with initial light soaking could signify the increased participation of charge carries due to the enhanced  $J_{sc}$  as well as the ionic reorganisation due to the increased  $V_{oc}$ . The performance parameters at 0 minutes and after 10 minutes of light soaking are shown in Table C.2.

	$J_{sc}$ / $\text{mAcm}^{-2}$	$V_{oc}$ / mV	FF / %	PCE / %
0 min	18.9	670	31.8	4.0
10 min	21.7	730	44.9	7.0

*Table C.2: Performance parameters at the initial instance of illumination and after 10 minutes of light soaking.*

In order to characterise the device after light soaking, J-V characteristics are measured at varying illumination intensities. Figure C.8 (b) shows the J-V characteristics at  $100 \text{ mWcm}^{-2}$ ,  $80 \text{ mWcm}^{-2}$ ,  $50 \text{ mWcm}^{-2}$ ,  $20 \text{ mWcm}^{-2}$  and under dark conditions. The incident light intensity is varied using neutral density filters. As the light intensity is increased, there is a corresponding increase in the short-circuit current density. The photocurrent density is plotted as a function of irradiance and shown in Figure C.9 (a). Generally, this follows a power law governed by  $J_{ph} \propto P^\alpha$ . When  $\alpha$  has a value of 0.75, the current is space charge limited and when  $\alpha$  is 1, the current is trap charge limited.

From analysis of Figure C.9 (a), a value of 0.96 is derived for  $\alpha$  which signifies that the devices demonstrate partial space charge limited current [198]. Figure C.9 (b) illustrates the variation of  $V_{oc}$  as a function of illumination irradiance and demonstrates an exponent of 1.94. The deviation of this exponent from 1 suggests that there is interfacial recombination or other loss mechanism at the perovskite-electrode interface. Figure C.9 (c) illustrates the variation of  $J_{sc}$  as a function of illumination intensity.

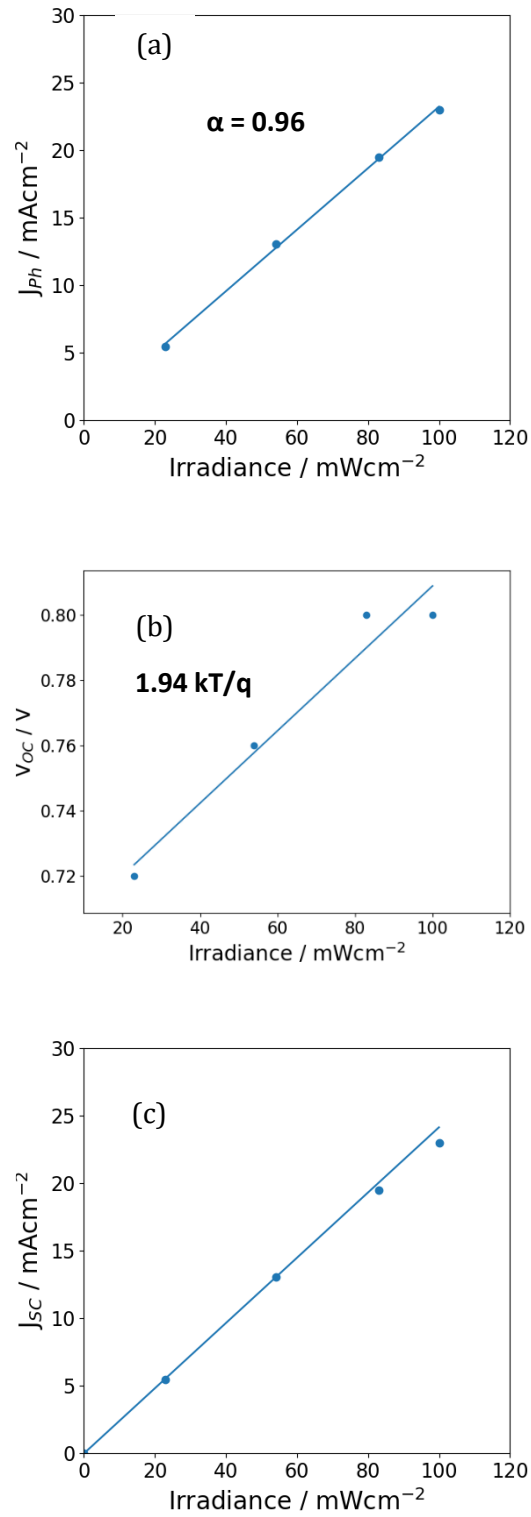
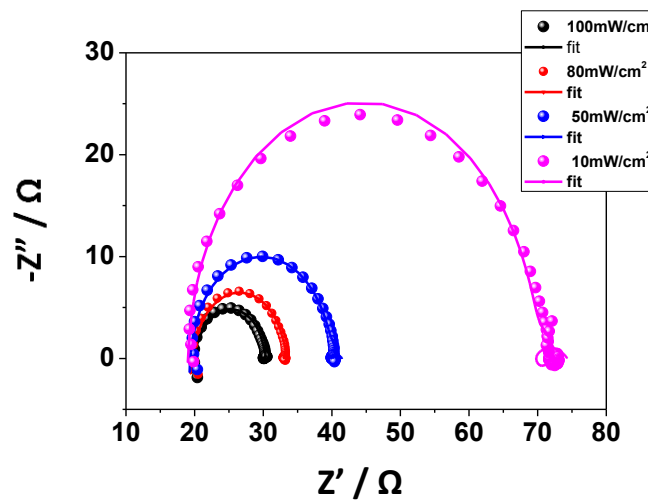


Figure C.9 (a) Variation of photocurrent density with irradiance (b) Variation of  $V_{oc}$  with incident illumination irradiance (c) Variation of  $J_{sc}$  with incident light intensity.

The increase in  $J_{sc}$  with incident light intensity displays a linear trend up to an intensity of 0.5 sun. As the  $J_{sc}$  values for high irradiance levels show a deviation, below the expected trend, this suggests some saturation of the  $J_{sc}$  value at higher irradiance. The trend line shown has a gradient of  $0.26 \text{ V}^{-1}$  and an intercept of  $0.1 \text{ mAcm}^{-2}$ .

A linear relationship of the short-circuit current density with varying light intensity suggests that the recombination of electrons and holes within the active layer is governed by first-order mechanisms. This means that in a range between the short-circuit current density and the maximum power point current density, the recombination is due to mono-molecular processes whilst bimolecular processes are less significant [199]. Mono-molecular recombination refers, primarily, to a first-order process such as geminate or Shockley-Read-Hall (SRH) recombination. A geminate process occurs when an electron-hole pair recombine before free charge carriers are created and an SRH process occurs when an electron-hole pair recombine via a trap state due to interfacial defects or impurities [200].



*Figure C.10: Variation in EIS as a function of different illumination intensities.*

The variation in the EIS as a function of the light intensity is shown in Figure C.10. The EIS demonstrates how the charge – transfer resistance dramatically increases from  $10.8 \Omega$  to  $50.5 \Omega$  as the irradiance is varied from  $100 \text{ mWcm}^{-2}$  to  $10 \text{ mWcm}^{-2}$ , whilst the contact resistance remains relatively constant with less than  $0.5 \Omega$  variation between the

different measurements. This increase in charge – transfer resistance at low illumination intensities supports the observation of a reduced photocurrent at low irradiances. made from the J-V analysis for varying illumination intensities.

#### C.4.2 Device Stability

The stability of the intercalated carbon devices is monitored under ISOS-D-2 (65 °C and ambient humidity), a modified ISOS-D-3 (65 °C and 35% relative humidity (RH)) and ISOS-L-2 (65 °C and 1 sun illumination). The ISOS-D-3 protocol is modified since testing at 65 °C/85% RH (as the standard dictates) results in almost instantaneous degradation of the device; by conducting the test at less harsh conditions, the variation in device characteristics can be understood in greater detail since it allows for a greater number of measurements to be performed over time.

During all stability tests, the EIS spectra were simultaneously taken and modelled by the equivalent circuits shown in Figure C.11. The variation in the normalised PCE and EIS over the course of the ISOS-D-2 test are shown in Figure C.12 (a) and (b) respectively. The variation in normalised PCE under ISOS-D-2 demonstrates the high intrinsic stability of the devices, with performance maintained above 90% of  $E_0$  for over 1000 hours. A similar stability trend was observed in [201]. The improved stability trend observed in carbon-based PSCs has also been investigated by Péan *et al.* and has been associated with the inclusion of the 5-AVAI additive in the perovskite (AVA-MAPI) when infiltrated in mesoporous layers; in this case the superoxide formation is significantly reduced as compared to MAPI deposited on glass [202].



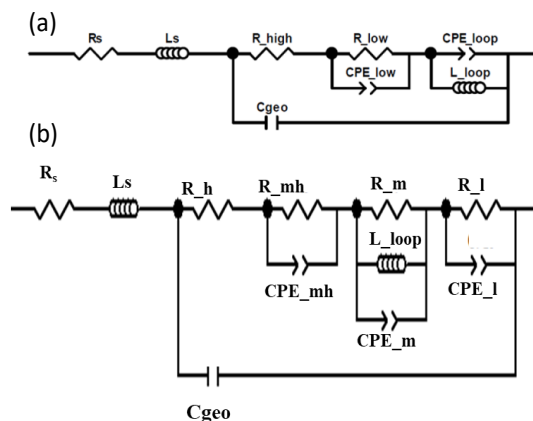


Figure C.11: Equivalent circuits used for modelling variation in EIS spectra for intercalated carbon devices under ISOS-D-2, ISOS-D-3 and ISOS-L-2.

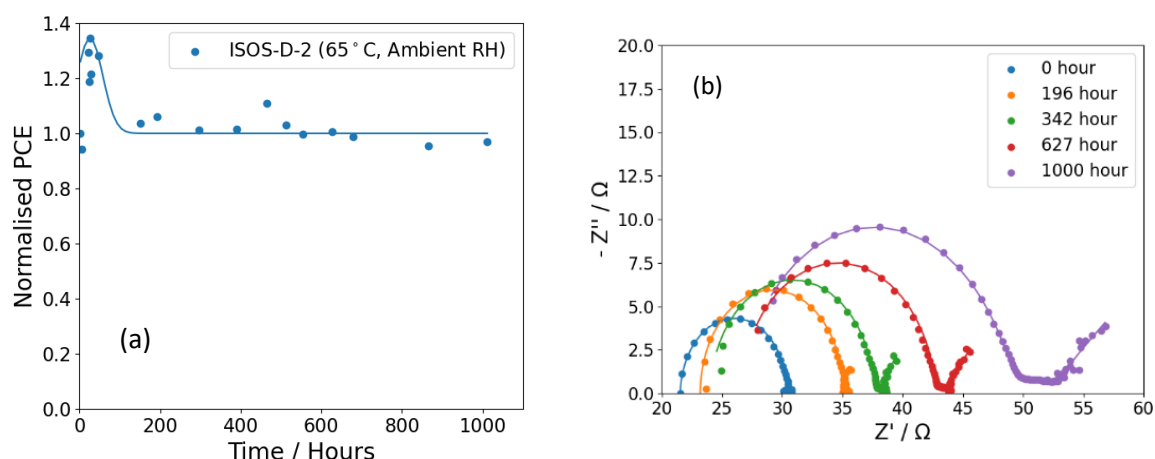
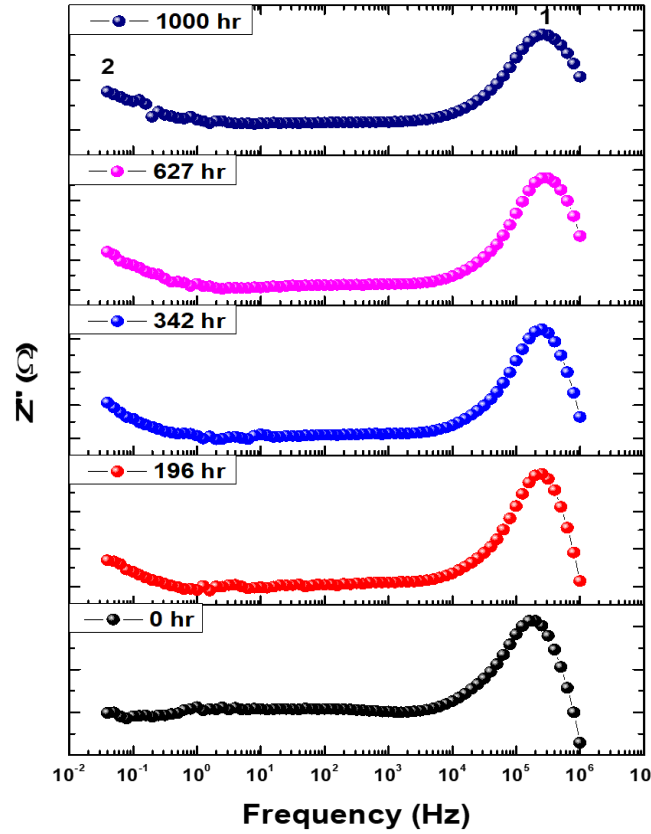


Figure C.12: Variation in (a) normalised PCE and (b) EIS for intercalated carbon devices under ISOS-D-2.

The variation in the EIS spectra for ISOS-D-2 is shown in Figure C.12 (b). The spectra are fitted using equivalent circuit analysis and allows the device parameters to be determined. The spectra demonstrate that both the contact resistance and the charge transfer resistance increase significantly during the test. This could be due to ion accumulation at the perovskite/electrode interface. Two equivalent circuits are required to model the evolution of the EIS since, after 196 hours, a low frequency feature develops, thus requiring an additional R-CPE element in the equivalent circuit and the circuit shown in Figure C.11 (b) was used. A similar observation can be made from the imaginary impedance ( $Z$ ) versus frequency curve where a clear difference can be observed in the low frequency region, shown in Figure C.13. Region 1 represents the recombination

process and region 2 represents the slow ionic motion or surface accumulation. With aging, the low frequency, slow processes due to ionic motion becomes more prominent.

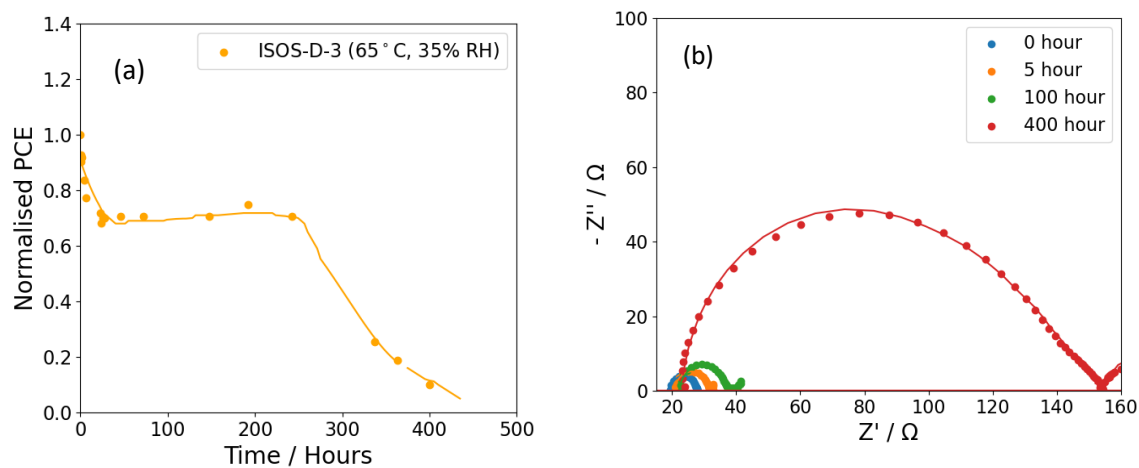


*Figure C.13: Frequency plotted as a function of  $Z''$  for different times under ISOS – D – 2 testing.*

The EIS analysis of the ISOS-D-2 stability test demonstrates how the physical properties of the PSC changes due to extended periods of heating. The two most prominent evolutionary features are the increase in the semi – circular width of the EIS and the shift of the high frequency intercept towards higher real impedance values. The shift of the intercept corresponds to an increase in the contact resistance ( $R_s$ ) of the PSC whilst the increase in width corresponds to the increase in charge – transfer resistance within the PSC. Specifically, the contact resistance of the device is seen to increase by  $6.5 \Omega$  over the course of 1000 hours and the charge – transfer resistance increases by  $14.0 \Omega$ .

Two other notable features develop due to the heating process; the high frequency offset of the EIS due to ageing and the introduction of an inductive loop after 196 hours. The high frequency offset is the result of the reduction of the transient constant phase element (CPE-T), which can be considered as the real capacitance of the PSC. The analysis shows that this element decreases by 2.5 pF over the course of 1000 hours. The geometrical capacitance is found to decrease 62 nF.

Subsequently, the variation in the PCE and EIS over the course of the modified ISOS-D-3 test are shown in Figure C.14 (a) and (b) respectively. The damp-heat test for the modified ISOS-D-3 demonstrates much reduced stability with  $T_{50\%}$  being reached in approximately 300 hours. However, the devices do maintain a 70% efficiency up to this time. This could signify the catastrophic failure of the electrode or breakdown in the ETL stack due to water infiltration.



*Figure C.14: Variation in (a) normalised PCE and (b) EIS for intercalated carbon devices under ISOS-D-3.*

Figure C.14 (b) illustrates the variation in EIS for the modified ISOS-D-3 test for 0 hours, 5 hours, 100 hours and 400 hours. The spectra demonstrate a significantly accelerated variation, in comparison with the ISOS-D-2 evolution, specifically demonstrating how moisture infiltration can significantly increase the charge transfer

resistance and contact resistance. For the equivalent circuit fitting, the same model is employed for ISOS-D-3. Similar to the ISOS-D-2 fitting, when using the ISOS-D-3 protocol an additional R-CPE element was used in the equivalent circuit which may also be attributed to the induced trap states or interfacial reactions due to the elevated moisture levels. The EIS analysis of the ISOS-D-3 stability test demonstrates how the physical properties of the PSC changes due to extended periods of heating and humidity exposure. The most prominent evolutionary feature is the significant increase in the semi – circular width of the EIS. The contact resistance of the device is seen to increase by  $4.5 \Omega$  over the course of 400 hours and the charge – transfer resistance increases by  $122 \Omega$ . This dramatic increase in charge transfer resistance correlates with the rapid degradation of PCE observed in Figure C.14 (a) for ISOS-D-3. This result demonstrates how the presence of water can rapidly deteriorate the PSC by specifically deteriorating the charge – transfer abilities of the active material, which is far more significant than the deterioration of the contact. In contrast to the ISOS-D-2 test, the capacitance for the device increases significantly by  $1.76 \text{ nF}$ .

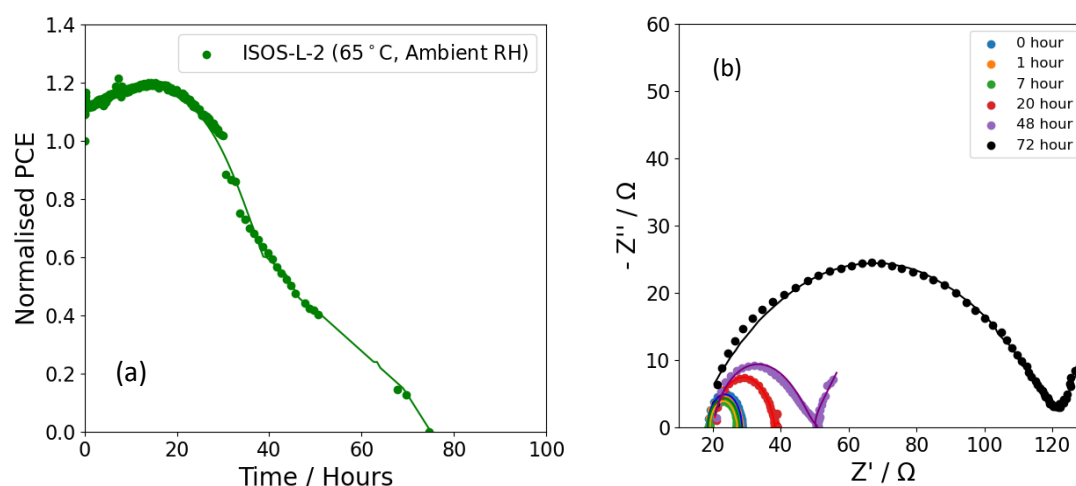


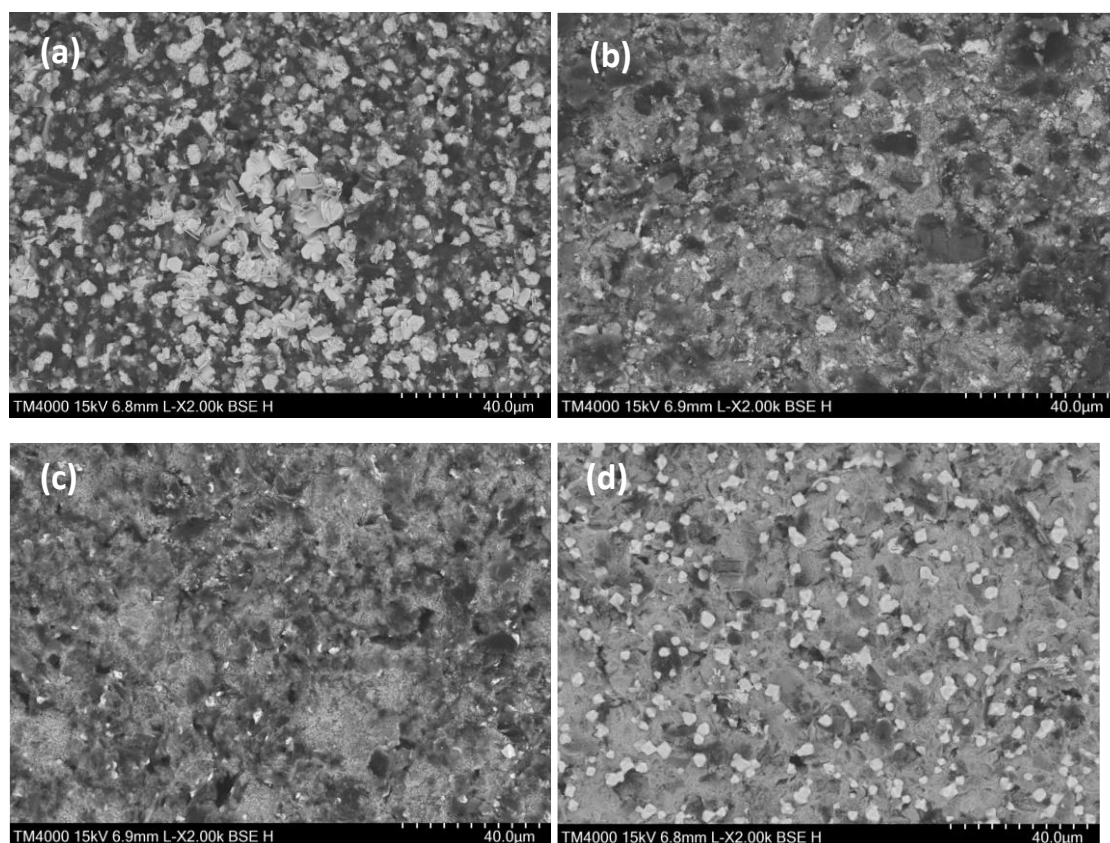
Figure C.15: Variation in (a) normalised PCE and (b) EIS for intercalated carbon devices under ISOS-L-2.

The variation in the PCE and EIS over the course of the ISOS-L-2 test are shown in Figure C.15 (a) and (b) respectively. The ISOS-L-2 test demonstrates very little stability, with the PCE falling to zero under 100 hours.

The variation in EIS under ISOS-L-2 is illustrated in Figure C.15 (b). In comparison to ISOS-D-2 and ISOS-D-3, the contact resistance proves to be maintained under ISOS-L-2. However, rapid increase in charge-transfer resistance is observed. In the case of ISOS-L-2, the equivalent circuit in Figure C.11 (b) is used after 7 hours of testing. For the aged devices we have also added an R-CPE element to model the increase in low frequency resistance. A slight decrease in perovskite capacitance has also been observed which was evident at the time of measurement when a change in perovskite layer colour was observed.

The EIS analysis of the ISOS-L-2 stability test demonstrates how the physical properties of the PSC changes due to extended periods of illumination. The most prominent evolutionary feature is the significant increase in the semi – circular width of the. The contact resistance of the device is seen to increase by  $2.6\ \Omega$  over the course of 72 hours and the charge – transfer resistance increases by  $88.7\ \Omega$ . This dramatic increase in charge transfer resistance correlates with the rapid degradation of PCE observed in Figure 6.14 for ISOS – L – 2. This result demonstrates how illumination can rapidly deteriorate the PSC by specifically deteriorating the charge – transfer abilities of the active material. Similarly to the ISOS – D – 3 test, the capacitance also increases during the ISOS-L-2 test, all be it far less significantly, by  $13.3\ \text{pF}$ .

The SEM images through the carbon electrode are shown in Figure C.16 where the device before and after aging for (a) a fresh device, (b) ISOS – D – 2 aging, (c) ISOS – D – 3 aging and (d) ISOS – L – 2 aging. The images shown in Figure C.16 illustrate how the carbon electrode deteriorates during each test. For each degraded device, the SEM images display a more homogenous structure, as compared with the fresh device, in particular for ISOS – D – 2 and ISOS – D – 3. For ISOS – L – 2 testing, crystal structure can still be observed, although with a reduction in size.



*Figure C.16: SEM image through carbon electrode for (a) fresh device, (b) device aged by ISOS-D-2, (c) device aged by ISOS-D-3 and (d) device aged by ISOS-L-2.*

In order to understand the operational performance of intercalated carbon devices, light cycling was performed, which simulates real-world outdoor conditions. However, it should be noted that the irradiance level is not varied during the illumination period. Therefore, the carbon devices are subjected to alternating periods of illumination and dark storage; the devices are light soaked for 8 hours followed by 16 hours dark storage at ambient conditions in the dark. Figure C.17 illustrates the variation in PCE during the light cycling test.

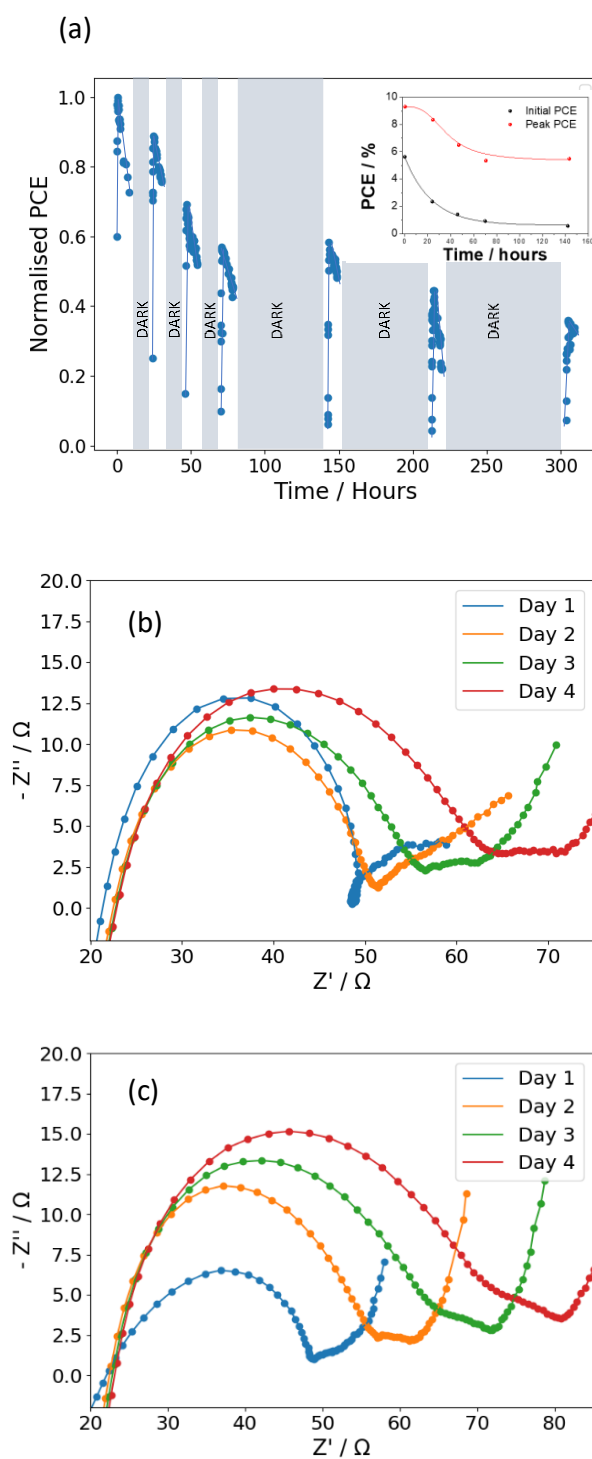


Figure C.17: (a) Variation in PCE during alternating light soaking and dark storage. Inset shows the exponential decay of the initial PCE measured at the start of each light soaking period and the decay of the peak PCE after initial light soaking. (b) Variation in EIS at start of light soaking and (c) variation in EIS at end of light soaking period.

During the initial 30 minutes of light soaking, the PCE increases in magnitude from 5.63% to 9.31%. This is attributed to the increased participation of trapped charge carriers in conduction. However, over the next 8 hours of light soaking, the PCE falls to approximately 70% of its peak value. After storage in dark conditions following the first light soaking period, the PCE recovers to 8.35%. The same level of recovery is not observed after subsequent dark storage periods, but in each instance, the PCE does recover to a value greater than the final measurement for the previous light soaking duration.

The inset in Figure C.17 (a) illustrates the variation in the PCE measured at the start of each light cycling period. The data follows an exponential decay curve very closely, despite the fact that further light soaking results in a very prominent recovery of the PCE. In addition, it can be seen that when the device is stored in dark conditions for a period of 64 hours, the maximum PCE observed after light soaking for approximately 30 minutes is greater than the maximum PCE observed during the previous light soaking period. This suggests that there is a correlation between the duration of dark storage and the level of recovery observed for the device. In addition, EIS analysis is performed during the light cycling test whereby the spectra are taken at the start of the illumination period and at the end of the illumination period.

Figure C.17 (b) and (c) show the variation in EIS spectra evolution during light cycling. The EIS at the start of each light soaking period is shown in (b) and the EIS at the end of each light soaking period is shown in (c). The obtained EIS display a contact resistance of  $(22 \pm 2) \Omega$  for all the measurements taken. The charge transfer resistance for the initial measurement is  $27 \Omega$ , and this maintained after 480 minutes of light soaking; the phase of the CPE component, decreases, as demonstrated by the depression of the semi-circle; this could signify the ionic reorganisation during the light soaking period, leading to reduced capacitance. The depression of the EIS spectrum is only observed after the initial 8 hour period of light soaking, possibly suggesting that the ionic reorganisation only occurs after the first instance of light exposure. The EIS taken at the start of the second day of light soaking demonstrates a partial recovery of the CPE-P term and could explain the recovery in the PCE recorded after storage in dark conditions. The second 8 hour period of light soaking results in the charge transfer resistance increasing significantly from  $51.4 \Omega$  to  $57.5 \Omega$ , with an associated, small increase in the capacitance.



Figure C.17 shows that the charge transfer resistance increases by  $11\Omega$  after each dark storage period and subsequent light soaking time. It is noteworthy that the charge – transfer resistance has decreased by approximately  $3\Omega$  between the end of day 2 and the start of day 3, suggesting a recovery of the active material; a decrease of approximately  $2\Omega$  is observed between the end of day 3 and the start of day 4.

## Appendix D Additional Studies – Large area OPVs

### D.1 Results and Discussion

Large area monolithically connected OPV modules were bonded onto rigid polycarbonate substrates and mounted at an inclination angle of  $35^{\circ}$  and orientated southwards. The edges of the flexible modules were sealed with pressure sensitive adhesive tape in order to minimise the effect of water infiltration into the modules. The modules were elevated in order to avoid shadowing effects. The active areas of the modules are  $360\text{cm}^2$ ,  $720\text{cm}^2$ ,  $2928\text{cm}^2$  and  $5904\text{cm}^2$ . Modules were made by InfinityPV in Denmark. Low resistance cables were soldered onto the contact points and were fed to a high voltage source measure unit (SMU). The performance of the modules was monitored with current-voltage (I-V) sweeps recorded at 10-minute intervals.

#### D.1.1 Diurnal Performance and Stability

Analysis of the diurnal performance of the modules allowed the response of the modules over a wide range of irradiance conditions to be tested. In addition, analysis of the performance under differing climatic conditions, such as low irradiance and cloudiness, allows the effect of diffuse light and cloud lensing to be observed and studied. Figure C.1 illustrates the variation in power conversion efficiency (PCE), short-circuit current density ( $J_{sc}$ ), open circuit voltage ( $V_{oc}$ ) and fill factor (FF) over the course of a sunny day with little or no cloud obstruction (5/8/2018). The irradiance curve for this day can be seen in Figure C.2 with a peak irradiance of approximately  $1000\text{ W/m}^2$ .

The PCE,  $J_{sc}$  and FF are relatively similar across all module sizes, with the only major variations being in  $V_{oc}$  due to the different number of cells within each module. Slight fluctuations can be observed in the  $J_{sc}$  and these can be attributed to minor cloud shading, as can be seen from the irradiance curve (See Figure C.2). The PCE displays

significant enhancement in the early morning and late in the evening, correlating strongly with the FF. This suggests that the low irradiance levels experienced at these times result in an increase in the FF for all modules. This effect is discussed further when partly cloudy conditions are investigated. Overall, the three largest modules show relatively similar PCE and characteristics irrespective of module sizes. However, a noticeable drop in performance is witnessed for the smaller modules.

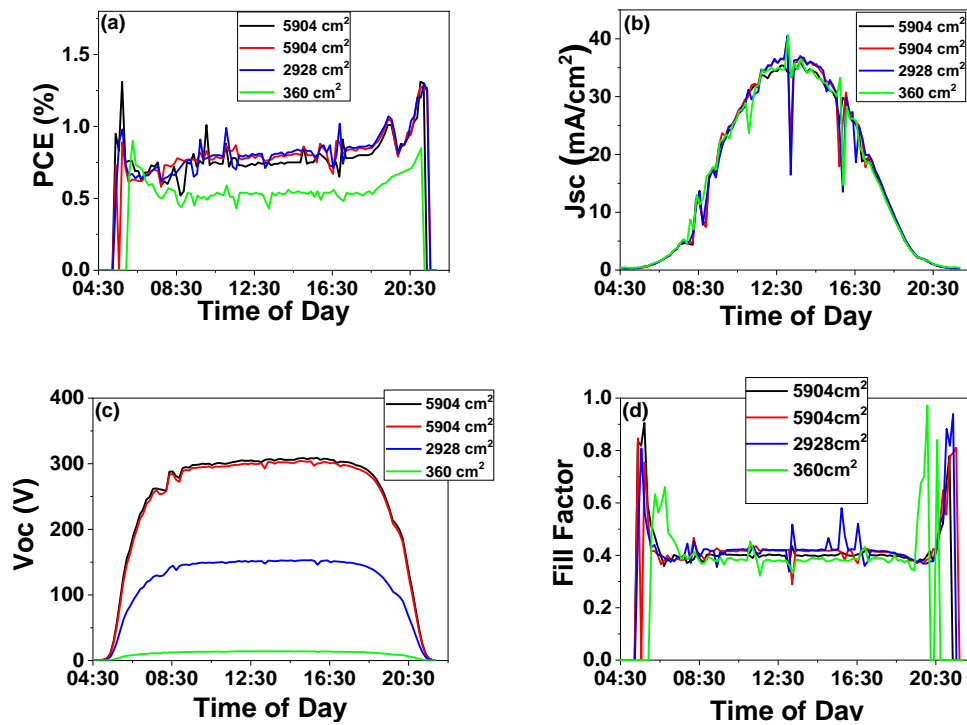


Figure C.1: Diurnal variation in solar module parameters (a) PCE, (b)  $J_{sc}$ , (c)  $V_{oc}$  and (d) FF over the course of a sunny day with peak irradiance of approximately 1 Sun for OPVS with four differing module sizes on the 5/8/2018

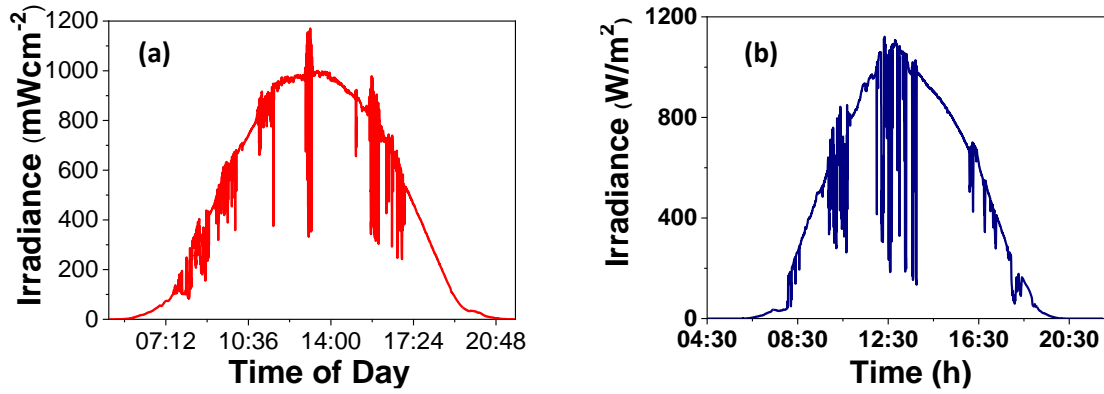


Figure C.2 (a): Irradiance variation over the course of the sunny day on the 5/8/2018 and (b) Irradiance variation on partly cloudy day 5/9/2018.

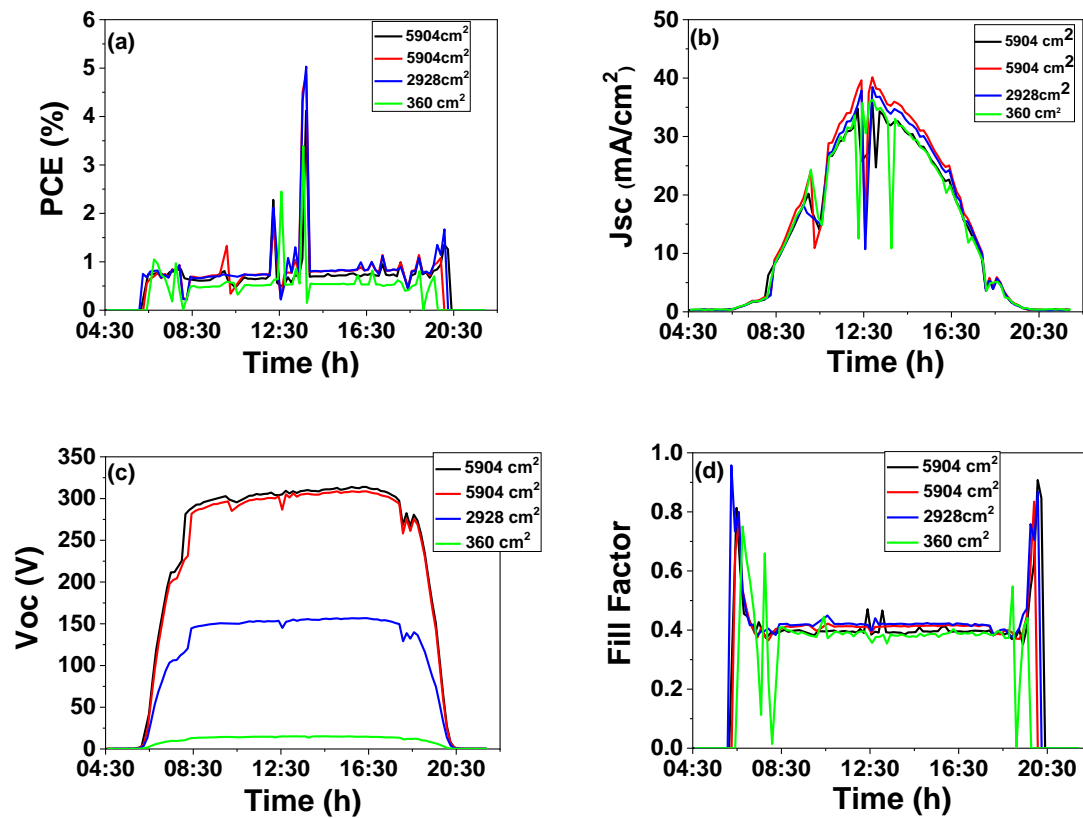


Figure C.3: Diurnal variation in solar module parameters (a) PCE, (b)  $J_{sc}$ , (c)  $V_{oc}$  and (d) FF over the course of a partly cloudy day.

To further highlight the effect of cloud shading and cloud lensing on the variations in the observed PCE, the diurnal performances on a partly cloudy day is investigated. On these days, a nearly smooth, bell-shaped irradiance curve is observed, except for significant cloud obstruction near to the period of maximum irradiance, as can be seen in Figure C.2. The diurnal variations of the module parameters are shown in Figure C.3. A spike in PCE is observed close to midday for all modules, with an associated depression in  $J_{sc}$ . The most likely cause of spikes in PCE is the slight mismatch between the time when irradiance was measured and when I-V scan was measured.

In addition to the diurnal performances of the modules, the degradation over the course of 4 months is monitored by tracking the PCE at a specific irradiance of  $300 \pm 10 \text{ Wm}^{-2}$ . This irradiance is selected in order to maximise the available data during both summer and winter months. Figure C.4 (a) illustrates the degradation curves for the modules between August 2018 and December 2018. The module with an area of  $720 \text{ cm}^2$  was observed to degrade rapidly. Figure C.4 (b) illustrates the variation of daily yield divided by daily dose. From Figure C.4, it is observed that the fastest rate of degradation is for smaller area modules, which begin degrading from the outset of the testing period. The module with an area of  $5904 \text{ cm}^2$  reaches  $T_{50}$  (the time taken to reach 50% of the initial efficiency) in approximately 5 months, whereas the module with an area of  $360 \text{ cm}^2$  reaches  $T_{50}$  in only 2 months. The degradation of the modules is most likely due to water infiltration between the encapsulating medium, resulting in degradation of the silver electrode grid as well as degradation of the solder bonds. Water and oxygen infiltration can also result in the degradation of the active material if water and oxygen infiltrate into the OPV stack. This will be enhanced during the summer months when there are higher levels of UV.

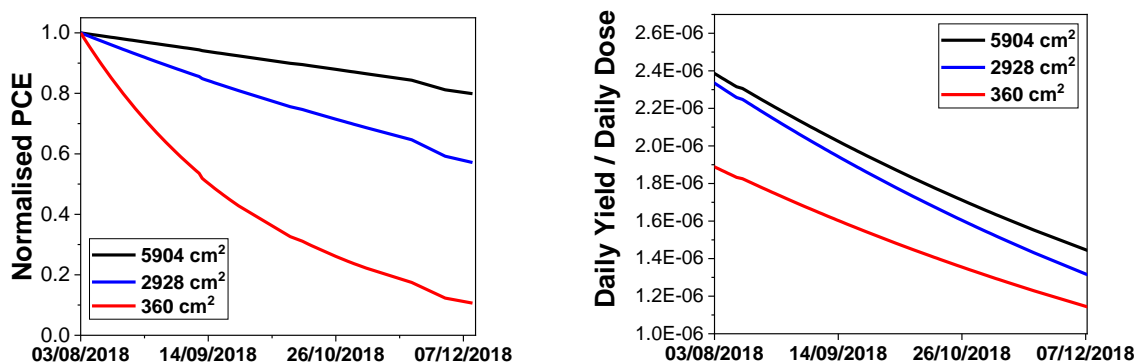
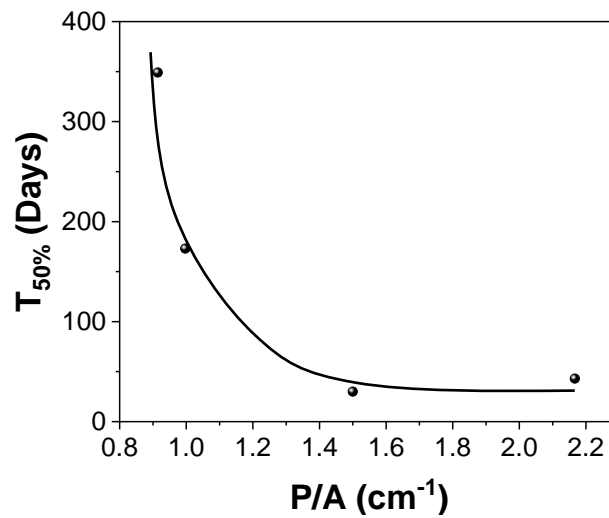


Figure C.4: (a) Degradation curves of modules over the course of 4 months at a fixed irradiance of  $300 \pm 10 \text{ Wm}^{-2}$ . (b) Variation of daily yield divided by daily dose over the course of 4 months.

The faster rate of degradation of the smaller modules can be attributed to the varying perimeter to active area ratio ( $P/A$ ) between modules. The  $P/A$  signifies how much of the active area can be reached by water infiltration through the perimeter of the module. Larger area modules possess a smaller  $P/A$  when compared with smaller area modules. For the largest area module, the  $P/A$  is found to be  $0.91 \text{ cm}^{-1}$  whilst for the smallest area module, the  $P/A$  is found to be  $2.2 \text{ cm}^{-1}$ . The majority of water infiltration into the modules will occur along the perimeter. This phenomenon will, therefore, lead to a faster rate of water infiltration for the smaller area devices. Figure C.5 illustrates the variation in  $P/A$  plotted as a function of  $T_{50}$  lifetime for each module size. From this plot it is evident that larger area modules display higher  $T_{50}$  lifetimes as a result of a smaller  $P/A$ , most likely due to less exposure to water infiltration and oxygen infiltration.



*Figure C.5: Variation in Time or reach 50% of the original efficiency ( $T_{50}$ ) lifetime as a function of Perimeter-to-area ratio ( $P/A$ ) ratio.*

Images of the modules before and after degradation are shown in Figure C.6. Figure C.6 (a) illustrates the colour of the active material at the beginning of the test period and Figure C.6 (b)-(e) illustrate the state of each module at the end of the test period (b=5904 cm<sup>2</sup>, c=2928 cm<sup>2</sup>, d=720 cm<sup>2</sup> and e=360 cm<sup>2</sup>). The colour change in the active material of the modules can clearly be seen when comparing Figure C.6 (a) and the Figure C.6 (b)-(e). In addition, the colour gradation between Figure C.6 (b) through to (e) can also be identified, corresponding to a decreasing module size. All images are taken at the centre of the module. This implies that there is greater degradation of the active material for the smaller area modules than for the larger area modules. One possible cause could be due to the faster rate of water infiltration as a consequence of the higher perimeter to area ratio for the smaller area modules. In addition, the edges of the smallest module can also be observed in Figure C.6 (e) and the effect of contact tarnishing can be seen (illustrated by the arrow), further demonstrating the infiltration of water into the module through solder joints and contacts. With improved encapsulation and edge sealants, this could be dramatically reduced.

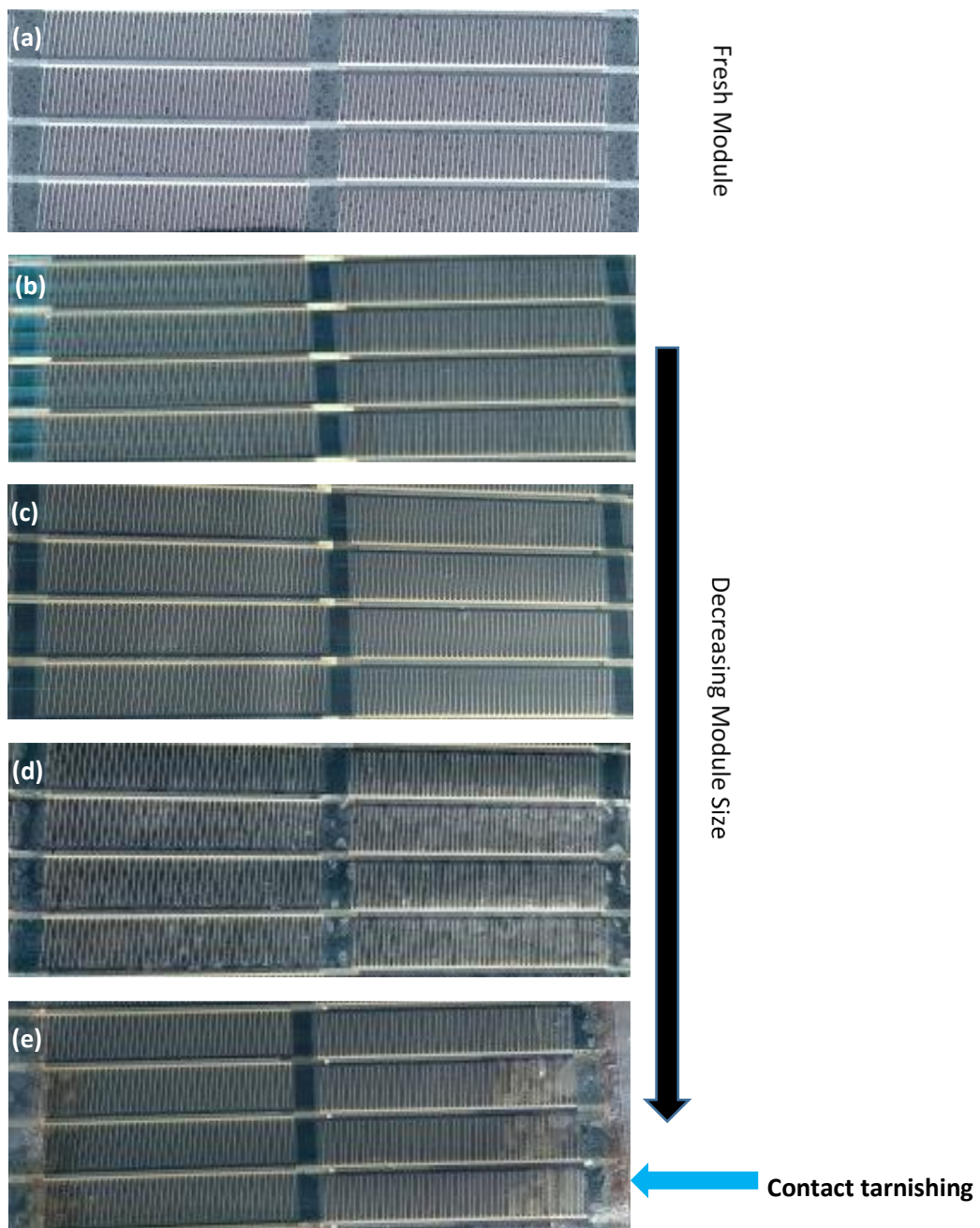


Figure C.6: (a) Active material at beginning of test period. (b) 5904 cm<sup>2</sup> module, (c) 2928 cm<sup>2</sup> module, (d) 720 cm<sup>2</sup> module and (e) 360 cm<sup>2</sup> modules after test period.