STC Short-Circuit Current Prediction and *I–V* Simulation of Colored BIPV Modules With Machine Learning and One-Diode Equivalent Circuit Models

Min Hsian Saw, Mauro Pravettoni, and Erik Birgersson

***Abstract*—Colored photovoltaic (PV) modules offer improved aesthetics at the cost of electrical performance loss. Here, we demonstrate a hybrid approach combining experiments, machine learning, and equivalent-circuit model to predict and simulate *I–V* parameters of colored crystalline-silicon PV modules. First, to predict the short-circuit current for different colors and opacity levels, three models—multiple linear regression (MLR), optimized support vector regression (SVR), and optimized Gaussian process regression (GPR)—are trained and evaluated with a ten-fold cross validation. The MLR model shows an MAE = 3.58 and an RMSE = 5.17. The accuracy could be further improved with the more advanced models, i.e., an optimized SVR (MAE = 0.22, RMSE = 0.24) or an optimized GPR (MAE = 0.13, RMSE = 0.17). Following, by taking the predicted short-circuit current values as inputs into a one-diode equivalent circuit model, *I–V* curves of two multicolored modules are simulated; and information such as module power P_{mpp} and current mismatch loss are extracted. There is a twofold advantage of implementing our approach: first, it serves as an efficient design space exploration methodology for a wide parameter space (e.g., colors), i.e., evaluating the PV performance for new colors without samples fabrication, thus saving time and resources; and second, it guides architects, designs, and engineers in color/design selection to achieve a balance between aesthetics and engineering considerations.**

***Index Terms*—Building-integrated photovoltaics (BIPV), colored photovoltaics (PV), equivalent circuit model, *I–V* performance, machine learning (ML), RGB color value.**

I. Introduction

ONE measure to improve building energy efficiency is by adopting renewable energy technologies, e.g., solar photovoltaics (PV) [1]–[5]. By retrofitting or integrating PV modules onto/into building envelopes, known as building integrated photovoltaics (BIPV), solar energy can be harvested and converted into electricity for buildings' self-consumption, which in turn reduces grid consumption [6]–[9]. There are a number of building components that provide opportunities for PV integration, such as roof, curtain walls, glazing, façades, canopy, shading device, etc.

From an architectural perspective, BIPV modules should have high variability in size, appearance (e.g., color, texture), and shape to satisfy the aesthetical need [10]. The aesthetic qualities are particularly crucial for façade elements. A broad range of research and development to create colored BIPV solutions has been conducted [11], [12]. One way is to modify the front glass cover of PV modules by creating colors with dyes and pigments. An example is digital ceramic printing [13].

The coloration approach circumvents the aesthetic concerns, making BIPV more attractive to architects. Nonetheless, it is accompanied by loss in electrical performance when compared to standard PV modules, due to light absorption by the glass fritz and pigment, and light reflection. Our previous study [14] and several other publications [15]–[17] have investigated the effects of digital ceramic printing on the light transmittance and electrical performance of PV modules, mainly based on measurements. Our previous study presented an experimental approach combined with single-variable statistical regression analysis to investigate the effects of digital ceramic printing of six natural color system (NCS) colors with varying opacity level on the module *I–V* parameters. For each of the six NCS colors, an individual regression fitting was performed, resulting in six color-specific statistical regression models. The approach suffers from the limitation that each regression equation is only applicable for a single color. In architectural designs, the choices of colors are vast. While it would be ideal if we could quantify the PV performance loss for varieties of colors and varying printing parameters solely via experimental approach, this often takes a considerable amount of time (months, starting with planning of experiments, obtaining samples to conducting measurements) and resources (e.g., personnel, instrument usage, monetary budget, etc.).

In view of the abovementioned points, our motivation for this work is twofold: 1) propose an efficient design space exploration methodology for a wide parameter space (large amount of color choices in architectural designs) that is more resource-effective compared to a pure experimental approach; and 2) bridge the gap

Manuscript received 12 May 2022; revised 27 June 2022; accepted 25 July 2022. Date of publication 10 August 2022; date of current version 28 November 2022. This work was supported by the Singapore Economic Development Board, under its Solar Competitive Research Program Grant S18-1175-SCRP. SERIS is a research institute at the National University of Singapore (NUS). SERIS is supported by NUS, the National Research Foundation Singapore, the Energy Market Authority of Singapore, and the Singapore Economic Development Board. *(Corresponding author: Erik Birgersson.)*

Min Hsian Saw and Erik Birgersson are with the Solar Energy Research Institute of Singapore, National University of Singapore, Singapore 117574, and also with the Department of Mechanical Engineering, National University of Singapore, Singapore 117575 (e-mail: sersmh@nus.edu.sg; mpebke@nus.edu.sg).

Mauro Pravettoni is with the Solar Energy Research Institute of Singapore, National University of Singapore, Singapore 117574 (e-mail: mauro.pravettoni@nus.edu.sg).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JPHOTOV.2022.3195693.

Digital Object Identifier 10.1109/JPHOTOV.2022.3195693

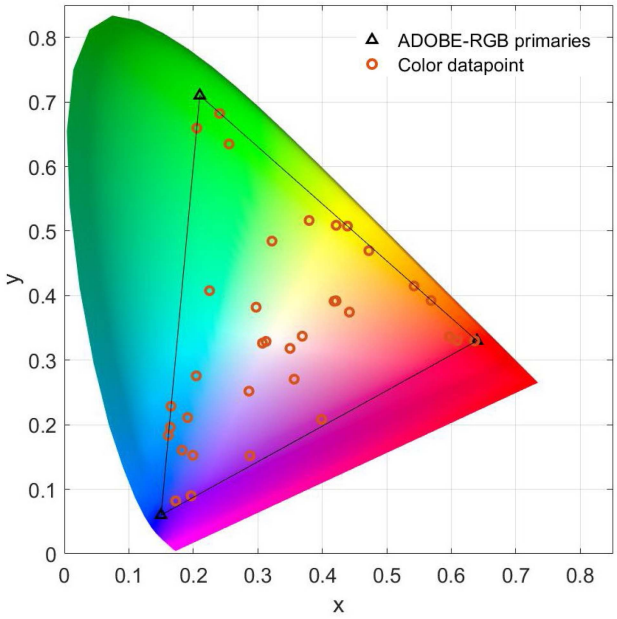

Fig. 1. Visualization of six base ink colors and 30 mixed colors in the CIE *xy* color space, leaving out the fourth variable A^*.

between architectural designs and engineering: paving a path to the development of a tool that can be employed by architects, designers, and engineers in color/design selection, balancing between aesthetics and PV module performance.

This article aims to develop data-driven machine learning (ML) model, coupled with an equivalent circuit model, to estimate the electrical performance of digital-ceramic printed colored BIPV modules. Also, the applied values of the ML model are demonstrated. Primarily, with the ML model, the estimation of the overall power for new multicolored PV modules, their $I_{\rm sc}$-mismatch loss, etc., is simplified, without the need of prototype fabrication and characterization. Secondarily, it contributes to better accuracy in energy yield simulation for colorful BIPV façades during design and conceptual phase. Module *I–V* parameters are input variables required for energy yield simulation [18]. At present, these parameters are often extracted from manufacturer data sheets for commercial products or in-house measurements for new prototypes. However, the amount of commercial colored BIPV modules are limited, and most of them are monochromatic. When new color designs are proposed, architects or engineers can only approximate the new module *I–V* parameters by either using the datasheet values of commercial colored PV modules or literature data that are believed to have the closest color to their designs, thereby introducing inaccuracy to the energy simulation for colorful BIPV façades. In contrast, our hybrid approach provides an estimate for the actual color designs that the architects or engineers are looking at, which in turn would result in more accurate energy simulation.

The first part of this study addresses the training and assessment of three ML algorithms—multiple linear regression (MLR), support vector regression (SVR), and Gaussian process regression (GPR), to predict the short-circuit current $I_{\rm sc}$ for different colors and opacity levels. Upon identifying the ML algorithm that is best-suited for our problem at hand, we demonstrate the primary applied value of the chosen ML model by the following:

1) predicting the $I_{\rm sc}$ for new colors designs;
2) constructing a one-diode equivalent circuit model for two large-size modules with the given color designs;
3) adopting the predicted $I_{\rm sc}$ values as inputs for the diode model;
4) simulating other *I–V* parameters such as $V_{\rm oc}$ and $P_{\rm mpp}$ of the colorful modules.

This article is organized as follows: Section II describes the ML protocol for model training and evaluation; Section III discusses the performance of the trained models in terms of predicting short-circuit current; Section IV shows the application of the ML model in simulating *I–V* parameters for new color designs; and Section V presents the conclusion and future work.

II. Development of Machine Learning Model

A. Datasets

In digital ceramic printing, the two main controlling factors are the color and opacity: the former is expressed as a combination of the three additive primary color channels—red (*R*), green (*G*), and blue (*B*), with each color value ranging from 0 to 255; and the latter is given by the degree of opacity that can be controlled with alpha compositing (*A*), ranging from 0% to 100%. Printing was done with a DipTech GlassJet Jumbo AR6000 digital printer, which has six ink channels corresponding to the six elementary colors defined in the NCS, allowing for a wide color gamut by mixing the inks. The degree of opacity was controlled by varying the translucency level of a raw image file in graphical software such as Adobe Photoshop/Illustrator. The four variables—*R, G, B,* and *A*—were normalized and denoted with $*$, to avoid bias toward features with higher magnitudes: the three-color variables were divided by 255, while the degree of opacity was divided by 100, to normalize them between 0 and 1. A total of 110 experimental observations were used for model training and evaluation. Out of the 110 observations, 60 are from our previous study for the six base ink colors with varying degrees of opacity [14], and the additional 50 are newly added observations for mixed-color samples. By combining both the base- and mixed-colored samples, we covered more of the color gamut, as we considered not only the six base colors, but also colors that are generated by mixing two or more of the base inks. By adding in the mixed-colored samples, the ML models were trained on a more versatile dataset that better represents the color range that can be printed with the technology. Fig. 1 visualizes the colors covered in this study—6 base ink colors that were considered in our previous study and additional 30 mixed colors which were newly added, in the CIE *xy* color space, leaving out the fourth variable (A^*). Adding in the fourth variable, there are 60 *RGBA* pairs for the base color dataset (6 base ink colors, each with 10 opacity levels), and 50 *RGBA* pairs for the mixed-color samples. Among the 50 mixed-color observations, 30 data points are made of 10 mixed-colors, each

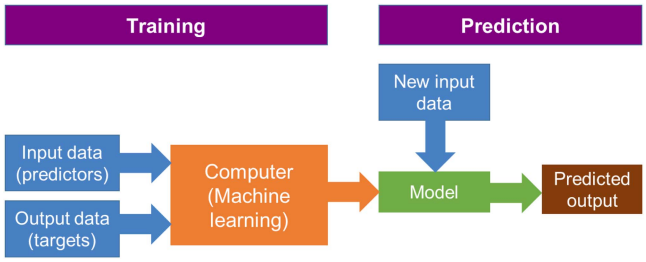

Fig. 2. Overview of machine learning pipeline.

with 3 opacity levels; while the remaining 20 data points are made of 20 different mixed-colors, each with 1 opacity level.

The data collection process was described in our previous paper [14]. To eliminate the inherited variation effects between the solar cells used to fabricate the mini-modules samples (both colored and reference), the measured module I_{sc} was normalized with respect to its corresponding cell I_{sc} via the following:

$$I_{sc}^{*} = \frac{I_{sc,mod}}{I_{sc,cell}} \tag{1}$$

where I_{sc}^{*} is the normalized short-circuit current of solar modules; $I_{sc,mod}$ is the measured short-circuit current of solar modules; and $I_{sc,cell}$ is the measured short-circuit current of the solar cells.

Next, the normalized short-circuit current of colored modules were evaluated against the normalized short-circuit current of the reference module with (2), giving the percentage short-circuit current ratio:

$$\% - I_{sc} = \frac{I_{sc,col}^{*}}{I_{sc,ref}^{*}} \tag{2}$$

where $I_{sc,col}^{*}$ is the normalized short-circuit current of the colored solar modules; $I_{sc,ref}^{*}$ is the normalized short-circuit current of the reference unprinted solar module ($I_{sc,ref}^{*} = 0.917$); and $\% - I_{sc}$ is the percentage short-circuit current ratio, which is also the output response for our ML models.

B. Machine Learning Protocol

ML is a data analytics technique which adopts computational methods to "learn" from data. It identifies patterns or trends in large pools of data and relies on inferences to solve different tasks such as classification, regression, clustering, and association [19]–[21]. The performance of ML algorithms adaptively improves as the number of samples available for learning increases. The first part of this study is focused on developing a multivariable predictive model with ML to estimate short-circuit current I_{sc} of colored BIPV modules. Supervised machine learning was conducted in the form of regression in MATLAB (see Fig. 2), as an extension to our previous study. The herein presented study addresses the following objectives.

1) Compare different algorithms and identify the ML algorithm that is best-suited for the problem at hand.
2) Optimize the hyperparameters of the learning algorithm and select the best-performing model.
3) Estimate the generalization performance (generalization accuracy or error) of the trained (and optimized) models on unseen data.

Due to the relatively small sample size (110 observations), a *k-fold* cross validation (CV) was implemented. This resampling method applies a randomized partitioning of the observations into k folds/subsets, of approximately equal size. The first fold is held out from the training and kept as a testing subset, while the ML models are trained with the remaining $k-1$ folds. This process is repeated k times, with each subset/fold left out exactly once in the testing. For each repetition, the prediction accuracies, e.g., root mean square error (RMSE), are calculated based on the hold-out subset, resulting in k estimates of the generalization accuracies.

It was unknown to us how the number of folds affects the testing accuracy of the models; hence, a sensitivity analysis was performed by varying the number of folds. Also, to check the sensitivity to the randomized splitting of the dataset across different runs, we ran the CV procedure ten times. From the empirical analysis performed on an MLR model, the variation in RMSE is shown to be marginal for different runs. Also, the effect of the number of folds is found to be negligible. Considering the bias-variance tradeoff [29], [30] and computational cost, a single-run (nonrepeated) ten-fold CV procedure was adopted for the ML model training and assessment.

In addition, an optimization algorithm was adopted to look for the optimized hyperparameters of nonparametric models. The models were then trained with the hyperparameter configurations that provide the best results.

C. Machine Learning Models

In this study, three ML models were investigated and compared—MLR, SVR, and GPR. MLR calculates the response variable from a linear combination of the predictor variables. As linear models are simple, and easy to train and interpret, they are often the first model to be fitted to a dataset [22], [23]. This model assumes a linear correlation between the response and predictors. As we have four predictor variables, i.e., $R*$, $G*$, $B*$, and A^{*}, a multiple linear regression model was applied to determine the association between short-circuit current loss and the color variables as well as the print opacity. The interaction effects between the predictor variables were included in the model too. The model is expressed as follows:

$$Y = \beta_0 + \beta_1 R^{*} + \beta_2 G^{*} + \beta_3 B^{*} + \beta_4 A^{*} + \beta_{12} R^{*} G^{*} + \cdots + \beta_{34} B^{*} A^{*} + \varepsilon \tag{3}$$

where Y is the response variable, β-parameters are coefficients of regression, and ε is the error term.

SVR and GPR models are "black-box" learning algorithms, which are created directly from the observations and hardly interpretable in making predictions. SVR is often adopted for problems with large number of predictor variables [24]–[26]. In SVR, kernel function, such as linear, polynomial, or Gaussian, is applied to map the training data into a higher dimensional space, known as a hyperplane. GPR is a kernel-based probabilistic model. In supervised learning, it is expected that the points with

TABLE I
COEFFICIENTS OF MULTIPLE LINEAR REGRESSION MODEL (MLR) FOR SHORT-CIRCUIT CURRENT PREDICTION

Predictor variables	Coefficient parameter	Coefficient values	Standard error	*t*-statistics	*p*-Value
Intercept	β_0	83.71	2.649	31.60	1.585e-53
R^*	β_1	7.666	4.893	1.567	0.1204
G^*	β_2	29.95	7.427	4.032	1.086e-04
B^*	β_3	20.20	5.629	3.589	5.181e-04
A^*	β_4	–55.30	4.109	–13.57	2.344e-22
R^*G^*	β_{12}	–19.56	6.242	–3.133	0.0023
R^*B^*	β_{13}	–4.094	6.073	–0.674	0.5018
R^*A^*	β_{14}	15.58	6.219	2.505	0.0139
G^*B^*	β_{23}	–18.28	6.513	–2.807	0.0060
G^*A^*	β_{24}	7.172	8.382	0.855	0.3942
B^*A^*	β_{34}	13.18	6.235	2.1143	0.0370

similar predictor values x naturally have close response values y. The similarity can be expressed by kernel-based covariance functions, e.g., exponential, Matern, and rational quadratic. The kernel function determines how the response at one point, e.g., x_2, is affected by responses at other points, e.g., x_1 [27]. The performance of these two models could be strongly influenced by the choices of the hyperparameter values; hence, hyperparameter optimization is crucial. Here, both SVR and GPR models were trained and optimized with a Bayesian optimization algorithm, which is more efficient as it chooses hyperparameters in an informed manner. By prioritizing hyperparameters that appear more promising from previous results, Bayesian algorithm finds the best hyperparameters in fewer iterations compared to other optimization methods such as grid search and random search. Different combinations of hyperparameters were tested, and the performance of each set of hyperparameters was evaluated by mean square error (MSE) from a *k*-fold CV. The hyperparameters were optimized by achieving the minimum MSE.

D. Assessment of Model Prediction Accuracy

To assess the model performance, the following accuracy statistics are evaluated: RMSE, mean absolute error (MAE), and coefficient of determination R^2. The accuracy metrics are expressed in the following equations [28]:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|\widehat{y_i} - y_i\right|, \tag{4}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{y_i} - y_i\right)^2}, \tag{5}$$

$$R^2 = 1 - \frac{\sum\left(y_i - \widehat{y_i}\right)^2}{\sum\left(y_i - \bar{y}\right)^2} \tag{6}$$

where y_i is the measured value, $\widehat{y_i}$ is the predicted value, $\bar{y}$ is the mean of the observations, and N is the number of observations. For the ten-fold CV, the three metrics are calculated for each fold, and the final presented results are the average estimates from the ten-fold data split.

III. PERFORMANCE EVALUATION OF THE MACHINE LEARNING (ML) MODELS

A. Multiple Linear Regression (MLR) Model

The MLR model included four variables $R*$, $G*$, $B*$, and A^* as predictors, as well as the interaction terms between these four variables. Following, we look into a more detailed analysis of this model to analyze the causal relationship between the four predictor variables and the response variable $\%\text{-}I_{\mathrm{sc}}$. Table I summarizes the estimates of the regression coefficients for the model, which are obtained with least squares criterion. We first look at the *p*-value for each predictor variable and the interaction terms to test the null hypothesis that this variable has no correlation with the response variable. Taking a significance level of 0.10, if the *p*-value for a variable is <0.10, we conclude that our samples provide sufficient evidence to reject the null hypothesis, indicating that there is a correlation between the predictor variable/interaction term with the response variable. Except the color channel variable $R*$, other individual variables, i.e., $G*$, $B*$, and A^*, are found to be statistically significant.

Looking at the interaction terms, the interaction effects between $R*$ and $G*$, $R*$ and A^*, $G*$ and $B*$, as well as $B*$ and A^*, appear to be statistically significant. In contrast, interaction effects between $R*$ and $B*$, as well as $G*$ and A^*, are not statistically significant, given their *p*-values >0.10. From the coefficient values, it is identified that for variables $R*$, $G*$, and $B*$, when one of them increases, while holding the other two constant, percentage short-circuit current $\%\text{-}I_{\mathrm{sc}}$ of the colored PV (with respect to the reference module) increases. This might correlate with the color addictive model whereby having zero intensity (value $= 0$) for each *RGB* component results in the darkest color (black). Black ink absorbs a significant portion of the solar spectrum in visible wavelengths (wavelength range in which the solar cell spectral responsivity is the highest), thus

TABLE II
OPTIMIZED HYPERPARAMETERS OF SVR AND GPR MODELS

Model	Hyper-parameters	Hyper-parameter search range	Optimized value
SVR	Kernel function	Linear; Quadratic; Cubic; Gaussian	Gaussian
	Kernel scale	0.001 – 1000	0.4076
	Epsilon	0.017535 – 1753.5	0.2008
	Box constraint	0.001 – 1000	744.17
GPR	Basis function	Zero; Constant; Linear	Constant
	Kernel function	Exponential; Matern 3/2; Matern 5/2; Rational quadratic; Squared exponential;	Matern 3/2, with a different length scale for each predictor
	Sigma	0.0001–173.41	0.0202

TABLE III
MODEL GENERALIZATION PERFORMANCE BASED ON A TEN-FOLD CROSS VALIDATION

Model	MAE	RMSE	R^2
MLR	3.58	5.17	0.89
Optimized SVR	0.22	0.24	0.99
Optimized GPR	0.13	0.17	0.99

reducing the light transmission to the solar cell remarkably. For variable A^*, the percentage short-circuit current of the colored PV (with respect to the reference module) shows an inverse relationship, i.e., the percentage short-circuit current decreases with increasing A^* value.

B. Optimized SVR and GPR Models

Turning our attention to the SVR and GPR models, their hyperparameters were optimized with a Bayesian optimization algorithm, in which their optimized hyperparameters were identified by looking for the minimum MSE during a ten-fold CV. The hyperparameter search range and optimized values are summarized in Table II. The Gaussian kernel function for SVR is considered a universal kernel and can be easily integrated against many functions. The low Epsilon value and high box constraint value suggest that a model with low bias is preferred to fit closely with the data. For the GPR model, the optimized hyperparameter combinations that produce the lowest RMSE are: constant basis function, Matern 3/2 kernel function with a different length scale for each predictor, and low sigma value.

C. Model Comparison Using Cross-Validation (CV) Performance

Now, we turn our attention to the ten-fold CV results summarized in Table III, which are used to assess the generalization accuracies of the models over unseen data.

From the ten-fold CV, the MLR model shows an RMSE = 5.17, MAE = 3.58, and $R^2 = 0.89$. Compared to the MLR model, both optimized SVR and GPR models show notably improved prediction accuracy from the 10-fold CV: RMSE = 0.24, MAE = 0.22, and $R^2 = 0.99$ for the optimized SVR model; and RMSE = 0.17, MAE = 0.13, and $R^2 = 0.99$ for the optimized GPR model. Fig. 3(a) shows the scatter plot of the ten-fold CV predictions as a function of the measured values for the MLR model: majority of the data points scatter along the diagonal line, indicating a reasonably strong correlation between the model's ten-fold CV predictions and its experimental results. From Fig. 3(b) and (c), the scatter plots show excellent convergence of the data points to the black diagonal line, indicating exceptionally good fitting accuracy of the optimized SVR and GPR models to the CV test set. Looking at the histograms of the differences between predicted and measured values in Fig. 3(d)–(f), we see that for the MLR model, about 80% of the predicted values deviate from the experimental values in the range of [–5%, 5%]; while for the optimized SVR and GPR models, all predicted values show a deviation within the range of [–1%, 1%]. A significantly better performance (although could possibly overfit) of optimized SVR and GPR than MLR could be due to the existence of a nonlinear relationship between %-I_{sc} and the predictor variables, which MLR has limitations in acknowledging such nonlinear relationship. Choosing between the MLR model and the advanced models for a final deployment, a tradeoff between accuracy and model interpretability is considered. The more simplistic MLR model is easier to interpret and allows us to better understand the correlation between the predictor variables and the response with a closed form equation. Whereas in the case where higher accuracy is desired, optimized SVR or GPR model could be adopted, but such model is complex and hard to interpret.

IV. DEMONSTRATING THE APPLIED VALUES OF THE PREDICTIVE MODEL

Based on our previous study, we identified that the cost to determine the electrical performance of colored PV modules solely through experiments is high due to the need of prototyping and characterization. Our developed predictive model, coupled with a solar cell equivalent circuit model, helps to overcome the resources constraints and can be adopted by the PV community and architects to estimate the electrical performance of colored PV for any color and opacity level of interest, which aids the architects and engineers in making an optimal decision between aesthetics and PV performance. Here, by adopting the optimized SVR model, and coupling it with an equivalent circuit model (based on one-diode model), we demonstrate how

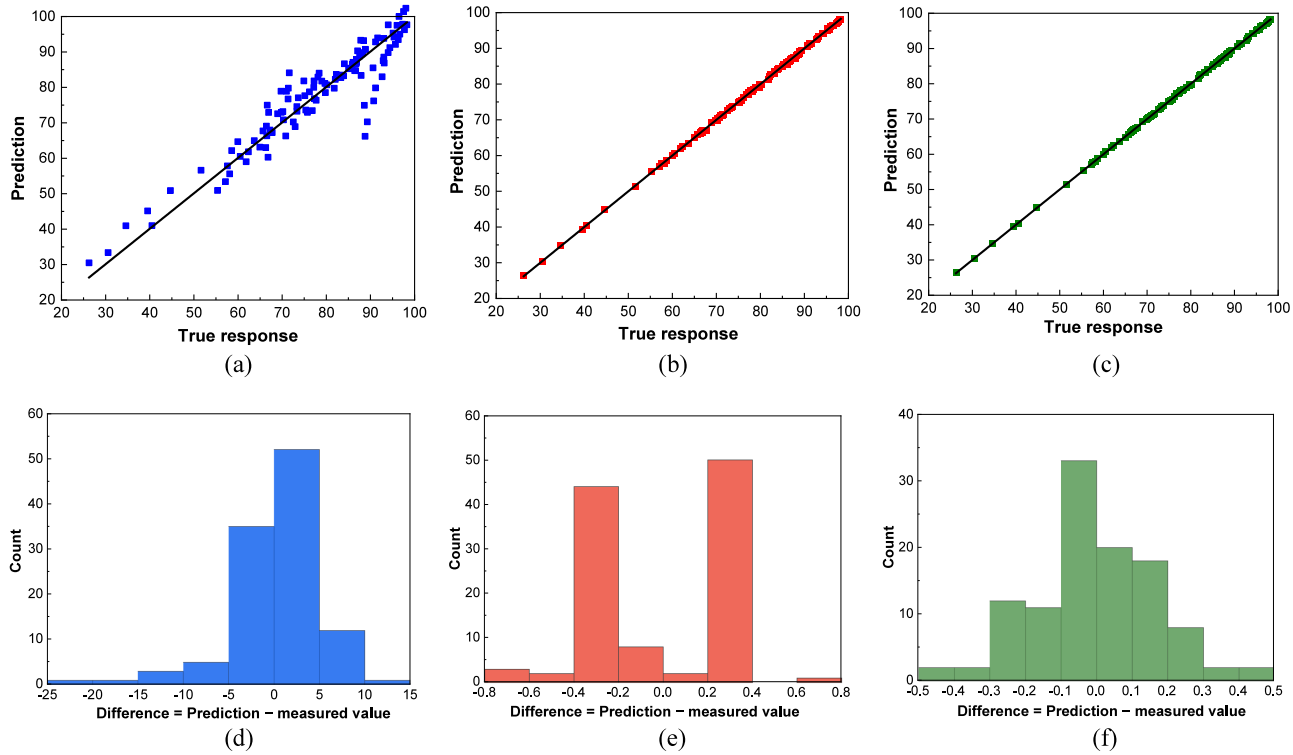

Fig. 3. (a)–(c) Scatter plot of the ten-fold cross-validation test results for the three models, respectively. (d)–(f) Error histogram of the ten-fold cross-validation test results for the three models, respectively. (a) Scatter plot - MLR. (b) Scatter plot - optimized SVR. (c) Scatter plot - optimized GPR. (d) Error histogram - MLR. (e) Error histogram - optimized SVR. (f) Error histogram - optimized GPR.

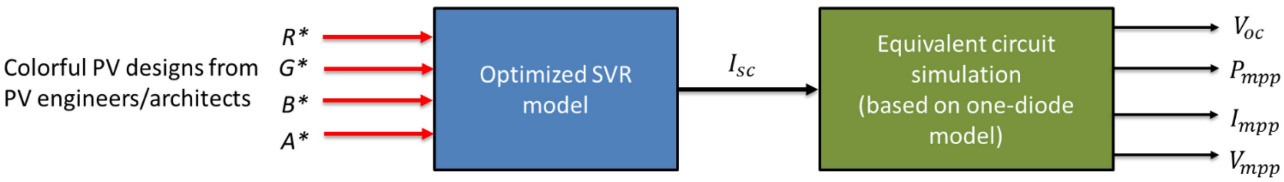

Fig. 4. Estimating the *I*–*V* parameters of multicolored PV modules with the optimized SVR model and equivalent circuit simulation.

TABLE IV
PARAMETERS OF THE ONE-DIODE MODEL

Model parameter	Value
R_s	0.004 Ω
R_{sh}	106 Ω
a	1.15
I_o	1.6E-19
I_{ph}	I_{sc}

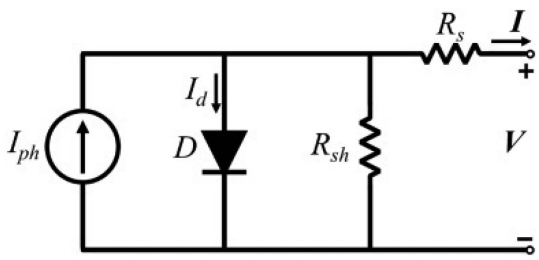

Fig. 5. Schematic of one-diode model.

the voltage–current (*I*–*V*) curves of large-size colorful modules can be simulated to estimate their power, current mismatch, etc. A one-diode model is adopted for the demonstration as it has sufficient degree of accuracy with a lower complexity and computational cost. The optimized SVR model is equally utilizable with other more complex equivalent circuit models such as the two-diode or three-diode models. Fig. 4 presents the workflow overview.

A. Equivalent Circuit Model Based on One-Diode Model

Here, we constructed an equivalent circuit model for a large-size PV module with 60 solar cells in LTSpice, based on the one-diode five-parameters model. In its simplest form, as shown in Fig. 5, the solar cell was modelled as a single diode with a series resistance and a shunt resistance [31]–[33]. Cell (and module) capacitance which is in parallel to the current source shows important effects in high efficiency c-Si modules:

TABLE V
SIMULATED *I–V* PARAMETERS OF THE VALIDATION AND DEMONSTRATION MODULES

	I_{sc} (A)	V_{oc} (V)	P_{mpp} (W)	I_{mpp} (A)	V_{mpp} (V)
Baseline module (simulation)	8.94	40.05	277.3	8.45	32.84
Baseline module (experiment)	8.94	40.05	277.5	8.46	32.80
Validation color module (simulation)	6.68	39.53	207.3	6.32	32.80
Validation color module (experiment)	6.68	39.65	209.0	6.32	33.03
Tri-color demo module (simulation)	4.72	39.18	156.86	4.62	33.97
Penta-color demo module (simulation)	5.29	39.20	152.23	4.40	34.60
Best-performing color demo module (simulation)	7.34	39.70	227.87	7.49	32.83
Worst-performing color demo module (simulation)	4.45	38.81	137.06	4.21	32.57

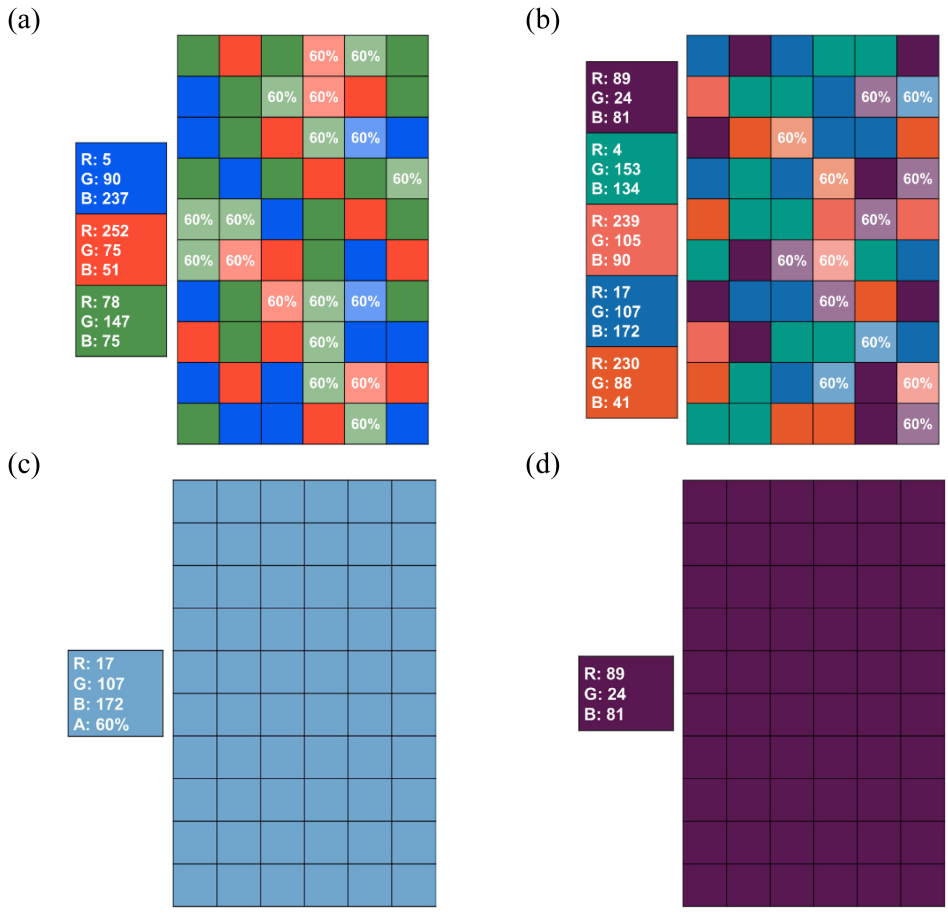

Fig. 6. Four new color designs used as demonstration modules: (a) tri-color, (b) penta-color, (c) best-performing color, and (d) worst-performing color.

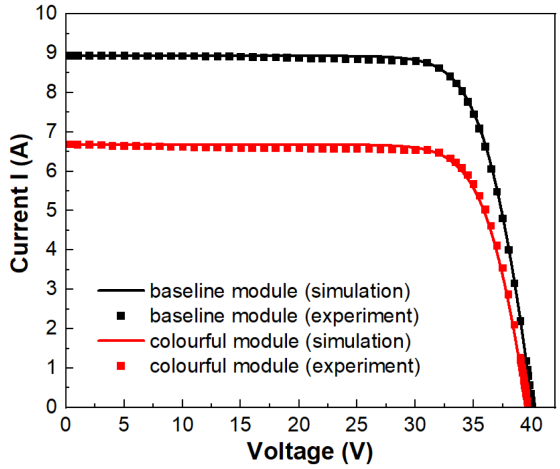

Fig. 7. Simulated (solid line) and experimental *I–V* curves (dotted) of the two validation modules.

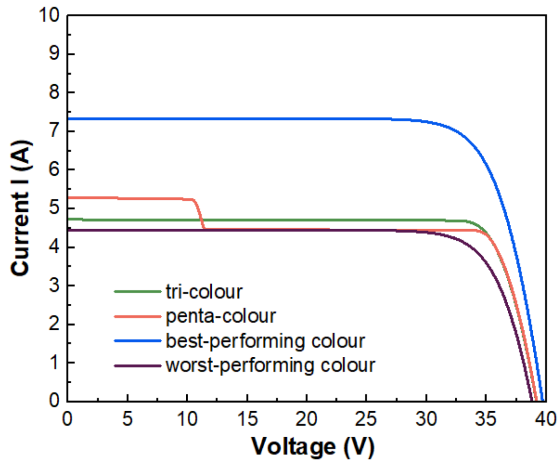

Fig. 8. Simulated *I–V* curves of the four demonstration modules.

however, revising the effect of capacitance in the one-diode model approximation goes beyond the goals of this work and therefore is not considered in our model; interested readers can find details in [34].

The well-known equation that describes the *I–V* relationship is given as follows:

$$I = I_{\text{ph}} - I_o \left[\exp\left(\frac{V + IR_s}{n_s V_T}\right) - 1\right] - \left(\frac{V + IR_s}{R_{\text{sh}}}\right) \quad (7)$$

where V_T in the above equation is expressed as follows:

$$V_T = \frac{akT_{\text{STC}}}{q}. \quad (8)$$

The five model parameters are as follows: I_{ph} which is the photo-generated current at STC; I_o which is dark saturation current at STC; R_s which is the series resistance; R_{sh} which is the shunt (parallel) resistance; and a which is the diode ideality factor. In addition, n_s is the number of cells connected in series; k is Boltzmann's constant; q is the charge of an electron; and T_{STC} is the temperature at STC.

B. Determination of the Model Parameters

Before the modeling exercise, we first measured the *I–V* curve of a baseline reference PV module (without color printing) with a class A+A+A+ solar simulator. From the *I–V* curve, we determined the series resistance R_s by calculating the inverse of the slope of the *I–V* curve at the open circuit voltage ($I = 0$); and the shunt resistance R_{sh} from the inverse of the slope of the *I–V* curve at the short-circuit condition ($V = 0$). The third parameter, photogenerated current I_{ph}, is evaluated as follows:

$$I_{\text{ph}} = \frac{G}{G_{\text{STC}}} I_{\text{sc}} + k_i \ (T_c - T_{\text{STC}}) \quad (9)$$

where G is the irradiance; G_{STC} is the irradiance at STC (1000 W/m^2); T_c is the cell temperature; and T_{STC} is the temperature at STC. Given that the module was measured at STC, it gives $I_{\text{ph}} = I_{\text{sc}}$.

Following, by calibrating the model using the module measured *I–V* curve, the diode ideality factor a, and the dark saturation current I_o were determined. Table IV summarizes the model parameters.

C. Simulation Results

Moving on to the simulation demonstrations, the equivalent circuit model was first validated with two modules: 1) baseline reference module used to calibrate the model in Section IV-B; and 2) a colorful module with known experimental *I–V*. Simulations were then performed for new color designs curated by a resident architect.

1) 60-cells tri-color module, with each solar cell covered by a color, shown in Fig. 6(a).
2) 60-cells penta-color module, with each solar cell covered by a color, shown in Fig. 6(b).
3) Two limit cases: one module with only the best-performing color (amongst the eight color designs) for all cells [see Fig. 6(c))]; and one module with the worst-performing color for all cells [see Fig. 6(d)].

Each color was represented with a set of $R{*}G{*}B{*}A{*}$ values, which were fed as inputs into our chosen optimized SVR model presented in Section III to estimate the I_{sc} values. Subsequently, the predicted I_{sc} values were given as inputs to the I_{ph} of the corresponding solar cells to obtain the overall module *I–V*.

The simulated and experimental *I–V* curves of the two validation modules are presented in Fig. 7; while the simulated *I–V* curves of the four demonstration modules are illustrated in Fig. 8. The *I–V* parameters, i.e., I_{sc}, V_{oc}, I_{mpp}, V_{mpp}, and P_{mpp}, of the validation and demonstration modules are summarized in Table V. Looking at Fig. 7, a good agreement between the simulated and measured *I–V* curves is observed, proving the validity of the equivalent circuit model. From Fig. 8, we see that the penta-color module shows current mismatch loss (step-curve), indicating nonuniform shading casted by the different colors on the solar cells. Contrarily, the tricolor module does not show any current mismatch effect, which makes it a more favorable option for BIPV installations from the technical perspectives.

V. Conclusion and Outlook

Determining the electrical performance of colored PV modules solely through experiments is costly considering the varieties of colors available in architectural design. To know more with less resources, this work has developed and evaluated three ML-based models to estimate the short-circuit current for any color and opacity level of interest. Based on an evaluation with a ten-fold CV, the two advanced nonparametric models: optimized SVR, and optimized GPR, show significantly better prediction accuracies than the MLR model. Comparing these three models, it can be concluded that when model understanding and interpretability is preferred over the accuracy, the multiple linear regression model can be adopted with a reasonable accuracy (RMSE = 5.17, MAE = 3.58). Whereas in the case of higher accuracy requirements (although could possibly overfit), models of higher complexity, i.e., optimized SVR (RMSE = 0.24, MAE = 0.22) or optimized GPR (RMSE = 0.17, MAE = 0.13), could be adopted. Following, the applied value of the predictive ML model is demonstrated: complemented with an equivalent circuit model, the voltage–current (*I–V*) curves and parameters for four new color designs are estimated.

While this study shows a number of important advantages, several limitations should be considered when interpreting the results. First, the prediction accuracy of the ML model could be further improved with the addition of more color points, subjected to resource availability. Second, our ML dataset is based on the colored PV samples created with a pigment/dye coloration technique, i.e., digital ceramic printing technology, and hence our results might be more applicable to PV modules that are created with similar technology.

Moving forward, we would extend our current study to include additional printing parameters such as ink layer thickness in the ML model to improves its versatility. Also, we would adopt the complementing ML and equivalent circuit models to estimate the *I–V* of multicolored BIPV façades and perform energy yield simulation for these case studies. At present, module *I–V* parameters required by the energy yield simulation are often extracted from the data sheet of commercial products; however, existing colored BIPV solutions are mostly limited to monochromatic modules. Using our models to estimate the module parameters for the multicolored façades, better accuracies in the energy yield simulation could be achieved.

Acknowledgment

The authors would like to gratefully acknowledge S. Nalluri, A. R. Alcain, and G. P. Raharjo for their technical contributions. The dataset that supports the findings of this study are available upon request.

References

[1] "International energy outlook 2019 with projections to 2050," U.S. Energy Information Administration (EIA), 2019.

[2] A. Chel and G. Kaushik, "Renewable energy technologies for sustainable development of energy efficient building," *Alexandria Eng. J.*, vol. 57, no. 2, pp. 655–669, Jun. 2018, doi: 10.1016/j.aej.2017.02.027.

[3] S. A. Kalogirou, "Building integration of solar renewable energy systems towards zero or nearly zero energy buildings," *Int. J. Low-Carbon Technol.*, vol. 10, no. 4, pp. 379–385, Dec. 2015, doi: 10.1093/ijlct/ctt071.

[4] A. Allouhi et al., "Energy consumption and efficiency in buildings: Current status and future trends," *J. Cleaner Prod.*, vol. 109, pp. 118–130, Dec. 2015, doi: 10.1016/j.jclepro.2015.05.139.

[5] L. Pérez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy Buildings*, vol. 40, no. 3, pp. 394–398, Jan. 2008, doi: 10.1016/j.enbuild.2007.03.007.

[6] B. P. Jelle, "Building integrated photovoltaics: A concise description of the current state of the art and possible research pathways," *Energies*, vol. 9, no. 1, 2016, Art. no. 21.

[7] C. Ferrara, H. R. Wilson, and W. Sprenger, "8 - Building-integrated photovoltaics (BIPV)," in *The Performance of Photovoltaic (PV) Systems*, N. Pearsall, Ed., Sawston, U.K.: Woodhead Publishing, 2017, pp. 235–250.

[8] H.-J. Kuo, S.-H. Hsieh, R.-C. Guo, and C.-C. Chan, "A verification study for energy analysis of BIPV buildings with BIM," *Energy Buildings*, vol. 130, pp. 676–691, Oct. 2016, doi: 10.1016/j.enbuild.2016.08.048.

[9] S. Domjan, C. Arkar, Ž. Begelj, and S. Medved, "Evolution of all-glass nearly zero energy buildings with respect to the local climate and free-cooling techniques," *Building Environ.*, vol. 160, Aug. 2019, Art. no. 106183, doi: 10.1016/j.buildenv.2019.106183.

[10] N. Sánchez-Pantoja, R. Vidal, and M. C. Pastor, "Aesthetic impact of solar energy systems," *Renewable Sustain. Energy Rev.*, vol. 98, pp. 227–238, Dec. 2018, doi: 10.1016/j.rser.2018.09.021.

[11] A. K. Shukla, K. Sudhakar, and P. Baredar, "Recent advancement in BIPV product technologies: A review," *Energy Buildings*, vol. 140, pp. 188–195, Apr. 2017, doi: 10.1016/j.enbuild.2017.02.015.

[12] G. Eder, P. Bonomo, and H. R. Wilson, "Coloured BIPV market, research and development," Int. Energy Agency, London, U.K., Rep. IEA-PVPS T15-07, 2019.

[13] B. Hoffman, "The impact of digital printing with ceramic inks on decorative and functional glass," *Glass Int.*, vol. 36, pp. 38–40, 2013.

[14] M. H. Saw, J. P. Singh, Y. Wang, K. E. Birgersson, and Y. S. Khoo, "Electrical performance study of colored c-Si building-integrated PV modules," *IEEE J. Photovolt.*, vol. 10, no. 4, pp. 1027–1034, Jul. 2020.

[15] E. Saretta, F. Frontini, P. Bonomo, T. Weber, and J. Berghold, "Indoor and outdoor characterization of innovative colored BIPV modules for facade application," in *Proc. 32nd Eur. Photovolt. Solar Energy Conf. Exhib.*, 2016, pp. 2498–2502.

[16] G. Eder et al., "Designed BIPV-elements with printed front-glass: Simulation and experimental evaluation of the effect of printing on the electrical performance," in *Proc. 12th Conf. Adv. Building Skins*, 2017, pp. 1255–1262.

[17] C. Kutter et al., "Decorated building integrated photovoltaic modules: Power loss, color appearance and cost analysis," in *Proc. 35th Eur. Photovolt. Sol. Energy Conf. Exhib.*, 2018, pp. 1488–1492.

[18] IEA, London, U.K., "BIPV design and performance modelling: Tools and methods," Tech. Rep. IEA-PVPS T15-09, 2019.

[19] R. Fernandes de Mello and M. A. Ponti, "A brief review on machine learning," in *Machine Learning: A Practical Approach on the Statistical Learning Theory*. Cham, Switzerland: Springer, 2018, pp. 1–74, doi: 10.1007/978-3-319-94989-5_1.

[20] G. James, D. Witten, T. Hastie, and R. Tibshirani, "Classification, an introduction to statistical learning: With applications in R," in *Springer Texts in Statistics*, New York, NY, USA: Springer, 2013, pp. 127–173, doi: 10.1007/978-1-4614-7138-7_4.

[21] G. Rebala, A. Ravi, and S. Churiwala, *Machine Learning Definition and Basics, An Introduction to Machine Learning*. Cham, Switzerland: Springer, 2019, pp. 1–17, doi: 10.1007/978-3-030-15729-6_1.

[22] A. Géron, *Hands-on Machine Learning With Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Sebastopol, CA, USA: O'Reilly Media, Inc, 2017.

[23] G. James, D. Witten, T. Hastie, and R. Tibshirani, *Linear Regression, An Introduction to Statistical Learning: With Applications in R, Springer Texts in Statistics*. New York, NY, USA: Springer, 2013, pp. 59–126, doi: 10.1007/978-1-4614-7138-7_3.

[24] R. F. de Mello and M. A. Ponti, "Introduction to support vector machines," in *Machine Learning: A Practical Approach on the Statistical Learning Theory*. Cham, Switzerland: Springer, 2018, pp. 163–226, doi: 10.1007/978-3-319-94989-5_4.

[25] G. James, D. Witten, T. Hastie, and R. Tibshirani, *Support Vector Machines, An Introduction to Statistical Learning: With Applications in R*. New York, NY, USA: Springer, 2013, pp. 337–372, doi: 10.1007/978-1-4614-7138-7_9.

[26] H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, "Support vector regression machines," in *Proc. Adv. Neural Inf. Process. Syst.*, 1997, pp. 155–161.

[27] C. E. Rasmussen, "Gaussian processes in machine learning," in *Advanced Lectures on Machine Learning*, O. Bousquet, U. von Luxburg, and G. Rätsch, Eds., Berlin, Germany: Springer, 2004, pp. 63–71, doi: 10.1007/978-3-540-28650-9_4.

[28] D. C. Montgomery, *Design and Analysis of Experiments*. Hoboken, NJ, USA: Wiley, 2017.

[29] G. James, D. Witten, T. Hastie, and R. Tibshirani, *Resampling Methods, An Introduction to Statistical Learning: With Applications in R, Springer Texts in Statistics*. New York, NY, USA: Springer, 2013, pp. 175–201, doi: 10.1007/978-1-4614-7138-7_5.

[30] J. D. Rodriguez, A. Perez, and J. A. Lozano, "Sensitivity analysis of k-Fold cross validation in prediction error estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 569–575, Mar. 2010.

[31] D. Sera, R. Teodorescu, and P. Rodriguez, "PV panel model based on datasheet values," in *Proc. IEEE Int. Symp. Ind. Electron.*, 2007, pp. 2392–2396.

[32] D. S. H. Chan and J. C. H. Phang, "Analytical methods for the extraction of solar-cell single- and double-diode model parameters from I-V characteristics," *IEEE Trans. Electron Devices*, vol. 34, no. 2, pp. 286–293, Feb. 1987.

[33] A. Laudani, F. R. Fulginei, and A. Salvini, "Identification of the one-diode model for photovoltaic modules from datasheet values," *Sol. Energy*, vol. 108, pp. 432–446, 2014.

[34] M. Pravettoni, D. H. G. Poh, J. P. Singh, J. W. Ho, and K. Nakayashiki, "The effect of capacitance on high-efficiency photovoltaic modules: A review of testing methods and related uncertainties," *J. Phys. D, Appl. Phys.*, vol. 54, 2021, Art. no. 193001.