Data-Driven I–V Feature Extraction for Photovoltaic Modules

Xuan Ma , Wei-Heng Huang , Erdmut Schnabel , Michael Köhl , Jenný Brynjarsdóttir ,
Jennifer L. Braid , *Member, IEEE*, and Roger H. French , *Member, IEEE*

Abstract—In research on photovoltaic (PV) device degradation, current–voltage (I–V) datasets carry a large amount of information in addition to the maximum power point. Performance parameters such as short-circuit current, open-circuit voltage, shunt resistance, series resistance, and fill factor are essential for diagnosing the performance and degradation of solar cells and modules. To enable the scaling of I–V studies to millions of I–V curves, we have developed a data-driven method to extract I–V curve parameters and distributed this method as an open-source package in R. In contrast with the traditional practice of fitting the diode equation to I–V curves individually, which requires solving a transcendental equation, this data-driven method can be applied to large volumes of I–V data in a short time. Our data-driven feature extraction technique is tested on I–V curves generated with the single-diode model and applied to I–V curves with different data point densities collected from three different sources. This method has a high repeatability for extracting I–V features, without requiring knowledge of the device or expected parameters to be input by the researcher. We also demonstrate how this method can be applied to large datasets and accommodates nonstandard I–V curves including those showing artifacts of connection problems or shading where bypass diode activation produces multiple "steps." These features together make the data-driven I–V feature extraction method ideal for evaluating time-series I–V data and analyzing power degradation mechanisms in PV modules through cross comparisons of the extracted parameters.

Index Terms—Data driven, diode model, I–V curve, photovoltaic (PV) module, series resistance, shunt resistance.

X. Ma and J. Brynjarsdóttir are with the Department of Mathematics, Applied Mathematics and Statistics, Case Western Reserve University, Cleveland, OH 44106 USA (e-mail: xxm115@case.edu; jenny.brynjarsdottir@case.edu).

W.-H. Huang is with the Solar Durability and Lifetime Extension Research Center, Department of Materials Science and Engineering, Case Western Reserve University, Cleveland, OH 44106 USA, and also with the Department of Statistics, Feng Chia University, Taichung 40724, Taiwan (e-mail: wxh272@case.edu).

E. Schnabel and M. Köhl are with the Fraunhofer Institute for Solar Energy Systems, 79110 Freiburg, Germany (e-mail: erdmut.schnabel@ise.fraunhofer.de; michael.koehl@ise.fraunhofer.de).

J. L. Braid and R. H. French are with the Solar Durability and Lifetime Extension Research Center, Department of Materials Science and Engineering, Case Western Reserve University, Cleveland, OH 44106 USA (e-mail: jlb269@case.edu; roger.french@case.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JPHOTOV.2019.2928477

I. INTRODUCTION

CURRENT–voltage (I–V) curve parameters are the most commonly used measurements to evaluate the performance and degradation of photovoltaic (PV) cells and modules. These performance features include the maximum power point (P_{mp}), short-circuit current (I_{sc}), open-circuit voltage (V_{oc}), shunt resistance (R_{sh}), series resistance (R_s), and fill factor (FF). The reduction of P_{mp} represents power degradation of a PV module or cell [1]. Other I–V features, meanwhile, imply specific mechanisms of module or cell performance and degradation [2]–[4]. Fitting the diode model to a single I–V curve, based on the theory and physics of solar cell operation, is the traditional way to obtain these I–V features.

One method of fitting the I–V curve with the diode model is to use the Lambert W function to obtain an explicit analytical solution [5]–[9]. Iterative numerical methods are also time-consuming and require manual setting or prior knowledge of the approximate initial fitting parameters for each I–V curve [10]–[14]. When analyzing a large number of I–V curves, for example, millions of I–V curves acquired from commercial PV power plants utilizing time-series I–V scanning tools, fitting to the diode model becomes computationally and/or labor intensive.

There have been several studies in the literature considering time-series of I–V curves and their features [15]–[19], with studies of only inverter-obtained data such as I_{sc}, V_{oc}, FF, and P_{mp} being even more common. Our group has recently employed network structural equation modeling (netSEM) for PV module degradation studies [1], [20], [21]. netSEM evaluates datasets of stressors, mechanisms, and responses as time series to identify and quantify relevant mathematical models linking these variables. For outdoor studies of PV modules, environmental exposure stressors, such as irradiance, temperature, and humidity, are modeled with responses such as power and wet insulation resistance, and mechanistic predictors of degradation, such as I–V features, to reveal active degradation pathways. Here, we propose a data-driven I–V feature extraction method to increase the efficiency and repeatability of I–V time-series data stream analysis. This is based on linear regression methods applied to different regions of the I–V curve [14], [22]–[25]. In this paper, we scale the linear regression approach to I–V curve fitting to accurately and efficiently process millions of I–V curves from a diverse variety of sources.

Data-driven regression approaches, as being presented here, intrinsically have sensitivities arising from the specific nature of the data used and its noise [14], [26]. Yet, at the same time, the diode model is not always adequate for describing the operation of PV modules (involving multiple cells, bypass diodes, etc.) and degraded PV devices [27], [28]. For example, multiple "steps" observed in a PV module I–V curve serve as an indication of mismatch between cells and/or irradiance present in different areas of the PV array or module under test. This can arise from partial shading of the PV array or degradation and damage of PV cells in the string, thereby causing bypass diodes to activate [29], and the resulting I–V curve does not conform to the diode model. Therefore, many studies regard these curves as erroneous and throw them out. Furthermore, I–V data must exhibit low noise for accurate use of the diode model: 1% noise in an I–V curve leads to approximately 20% of relative error in the value of R_s extracted from the fitted diode model [8].

In this paper, we describe this data-driven I–V feature extraction method and algorithm for time-series I–V studies of PV modules. Statistical methods such as simple linear regression and smoothing spline are used [30]. A simulation study is conducted to evaluate the performance of the I–V feature algorithm on diode-model-generated I–V curves with various levels of noise. Datasets from different sources, as well as a time series of I–V curves, are used to demonstrate how the proposed method performs on real-world data and how it can be applied in practice.

II. EXPERIMENT AND METHOD

A. Data-Driven I–V Feature Extraction Method

The data-driven I–V feature extraction method uses linear regressions and basic computational practices on various regions of the I–V curve to obtain values for several I–V features as follows. I_{sc} is defined as the current at zero voltage (the y-intercept of the I–V curve), while V_{oc} is the voltage at zero current (the x-intercept). R_{sh} is calculated as the negative inverse slope of the I–V curve near $V = 0$, and R_s is the negative inverse slope of the I–V curve near V_{oc}. P_{mp} is the maximum product of current and voltage on the I–V curve. FF is defined as the ratio of the maximum power from the solar cell to the product of V_{oc} and I_{sc} and measures the "squareness" of the solar cell's I–V curve. I–V curve parameters, as defined in this method, are illustrated in Fig. 1.

In most PV module I–V curves, observation points are evenly spaced in voltage. However, when approaching to V_{oc}, the current decreases exponentially, resulting in few points in this pseudolinear region close to V_{oc}, which may introduce bias when estimating V_{oc} and R_s. Additionally, some I–V tracers acquire more data points near P_{mp} for more accurate determination of the ideal operating point and fewer points in the pseudolinear regions near I_{sc} and V_{oc}. Thus, we use a smoothing spline on each raw I–V curve to generate an equivalent I–V curve with 500 points with equal spacing in voltage, giving enough data points to estimate these features with low statistical uncertainty. The smoothing spline involves interpolation and nonparametric
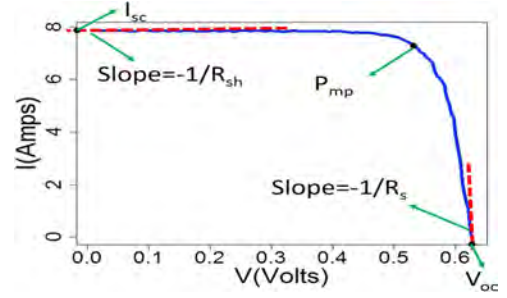

Fig. 1.   Standard one-step I–V curve and five I–V features: I_{sc}, R_{sh}, P_{mp}, V_{oc}, and R_s.
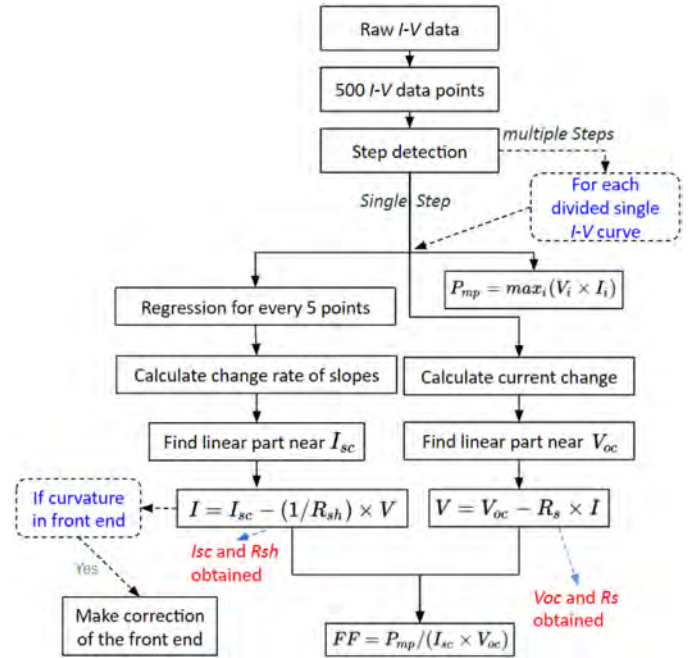
Fig. 2.   Detailed procedure of the data-driven method to calculate I–V features.

regression. Let $\{V_i, I_i : i = 1, 2, \ldots, n\}$ denote a set of n observations and $f(v)$ be a function that fits the observed data. The smoothing spline is the function f that minimizes

$$\sum_{i=1}^{n}\{I_i - f(V_i)\}^2 + \lambda \int \{f''(v)\}^2 dv \qquad (1)$$

where λ is a nonnegative tuning parameter that controls the roughness of the smoothing spline [30]. We use the stats::smoothspline function in R to perform the spline [31]. The data-driven I–V feature extraction method applies the above definitions of the six parameters (I_{sc}, R_{sh}, V_{oc}, R_s, P_{mp}, and FF) to automatically calculate their values, as illustrated in Fig. 2.

Because not all I–V curves have a single step, as shown in Fig. 1, we use segmented regression to find the number and locations of change points. Segmented regression can identify change points in a curve and is used here to figure out the voltage where a change point occurs [32], [33]. However, not all change points indicate the appearance of steps. The change

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

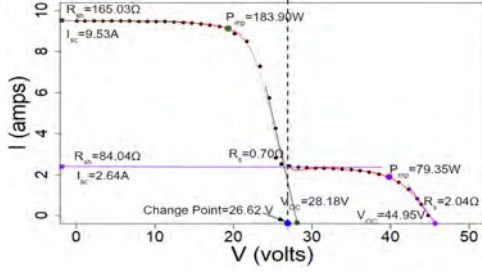MA *et al.*: DATA-DRIVEN *I–V* FEATURE EXTRACTION FOR PHOTOVOLTAIC MODULES

3

Fig. 3. *I–V* feature extracted results of an example of *I–V* curve with two steps.

Fig. 4. Change rate of the slope for regressions performed with a moving window of five points based on a standard *I–V* curve.

points between steps are those with the slope on the left "steeper" than the slope on the right, with the slope on the left being negative. In addition, the difference between the absolute value of slopes on the left and right sides of the point should be sufficiently large. Thus, we denote β_1 as the slope to the left of the change point, β_2 as the slope to the right of the change point, and a as a parameter decided by the noise of the *I–V* curve (for larger noise, set larger a). We build the following criterion to identify the change points that indicate multiple steps.

1) $|\beta_1| > |\beta_2|$.
2) $\beta_1 < 0$.
3) $|\beta_1| - |\beta_2| > a$.

We then extract *I–V* features on each step of the *I–V* curve.

Fig. 3 shows the *I–V* feature extracted result of an example *I–V* curve with two steps from the Fraunhofer-ISE dataset. In this example, we find that the voltage of the change point between steps is 26.62 V. Based on this point, the original *I–V* curve is divided into two single-step *I–V* curves, and we extract *I–V* features for each step.

To determine I_{sc} and R_{sh}, a linear regression is performed on the 500-point splined *I–V* curve. The linear regression model is shown follows:

$$Y = \alpha + \beta \times X + \epsilon \tag{2}$$

where X is the independent variable, Y is the dependent variable, α and β are coefficients, and ϵ is the error term.

The regression is performed on a moving window of five consecutive points, with current being the dependent variable and voltage being the independent variable, and the slope for each five-point window is stored. A five-point window (i.e., 1% of the splined data length) is used because only a very small number of observations approximate to a straight line; therefore, this length is most accurate to estimate slope as well as the change of slopes along the *I–V* curve.

We could expect that the slope coefficients for the low-voltage part of the *I–V* curve do not change much between windows. For the part of *I–V* curve that passes through the maximum power point, the slope coefficient changes sharply, and we use the rapid change in slope of the five-point moving box to identify some of the *I–V* features such as change points between steps [27]. A typical change of the slope pattern for a standard one-step *I–V* curve can be seen in Fig. 4. As shown in the figure, the change rate of the five-point line slope remains relatively stable from zero voltage to approaching the maximum power point,
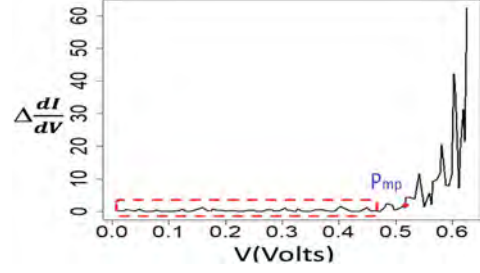
as this corresponds to the linear part in the *I–V* curve near I_{sc}. Thus, we set a critical value for the change rate of the five-point slope to find this linear region and determine the corresponding consecutive current and voltage points that have a change rate in slope smaller than this critical value. With the selected data points from the linear region for the *I–V* curve near I_{sc}, the linear regression model in (2) is used to find the slope and intercept. Based on our definition of I_{sc} and R_{sh} (see Fig. 1), I_{sc} is estimated with the intercept of the fitted line, and R_{sh} is estimated by the negative inverse of its slope. Note that the number of selected data points for the fit of I_{sc} and R_{sh} is typically 70–75 on the 500-point splined *I–V* curve.

Some *I–V* curves, especially from outdoor systems, exhibit a rapid change in slope approaching 0 V, which makes R_{sh} and I_{sc} determination challenging. As shown in Fig. 5(a), a nonlinear region near I_{sc}, due to the poor module connection or mis-recording by the *I–V* curve tracing system, should be removed in this *I–V* curve. Therefore, in our proposed algorithm, only consecutive data points with low change of slope are used to determine I_{sc} and R_{sh}, thereby excluding curvature the low-voltage region, as shown in Fig. 5(b). Then, using the selected linear data points, we correct the nonlinear region, as shown in Fig. 5(c). This method can automatically find the appropriate current and voltage values that define the linear part of *I–V* curves at low voltages.

V_{oc} and R_s are similarly calculated from the linear part in the *I–V* curve with voltage higher than that of the maximum power point. Here, we consider a regression model with voltage as the dependent variable and current as the independent variable. Let the change rate of current be the difference between two consecutive data points of currents divided by the current with lower voltage. Thus, we set a critical value and select the data points that have change rate of current larger than the critical value consecutively. According to the definition of V_{oc} and R_s, V_{oc} is estimated by the linear intercept, where $I = 0$, and R_s is estimated by the slope. Note that the number of selected data points for the fit of V_{oc} and R_s is typically 50–55 on the 500-point splined *I–V* curve.

Finally, P_{mp} is calculated by finding the maximum product of current and voltage for each of the $500\, I, V$ data point pairs, without fitting of the spline. FF is calculated as follows:

$$FF = \frac{P_{mp}}{I_{sc} \times V_{oc}} . \tag{3}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

4   IEEE JOURNAL OF PHOTOVOLTAICS
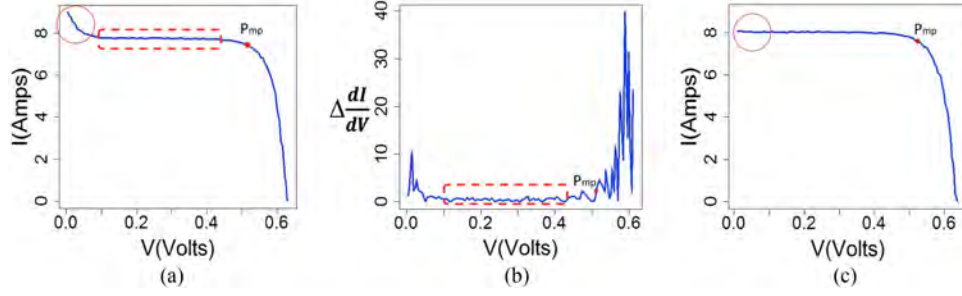

Fig. 5. (a) I–V curve showing curvature in the low-voltage region of the curve due to a module connection problem. (b) Change of slope has great variation in the front, remains stable for the linear part near I_{sc}, and then increases sharply. (c) Using the data-driven method, we are able to correct the I–V curve with curvature.

In addition, the repeatabilities are calculated for the I–V features using a resampling method. For each iteration, we randomly select 90% of the data points and apply the extraction method to obtain I–V feature results. We repeat this 10 000 times to obtain 10 000 values of each extracted I–V feature. The standard deviation (SD) of all 10 000 iterations is calculated for each I–V feature, and the repeatability is defined as 100% − the SD%.

The data-driven I–V feature extraction method and functions are available as a free open-source R package, easily downloaded from the Comprehensive R Archive Network [34].

B. I–V Measurement and Datasets

In order to validate the data-driven I–V feature extraction method, we first use data simulated with the single-diode model to compare the given and extracted I–V features. We then, demonstrate the method on the following three time-series I–V curve datasets from different sources, each with a different number of data points (or observations) in the I–V curve.

The Fraunhofer-ISE dataset [35], [36] consists of I–V time series from eight PV modules across three different locations, with two modules on Mount Zugspitze (in Germany, abbreviated as UFS, in the ET climate zone), three modules in Gran Canaria (in Spain, abbreviated as GC, in the BWh climate zone), and three modules in the Negev Desert (in Israel, abbreviated as NEG, in the BSh climate zone). Climate zones are classified by the Köppen–Geiger climate zone system [37]. BWh represents a hot desert climate, BSh is a hot semiarid climate, and ET is polar climate [38]–[40]. Depending on the module, we have data for three to six years of outdoor exposure with power readings taken every 2–3 min and I–V curve measured every 10 min. The UFS data start from 2012, while data for GC and NEG start from 2010. There are the total of 2.2 million I–V curves, and each single I–V curve has 40–42 data points. These I–V curves were acquired using an ESL Solar 500 tracer made by ET Instrumente [41] in ambient conditions with varying irradiance and temperature.

The second dataset of I–V curves is from the SDLE SunFarm, a 1-acre outdoor test facility on the Case Western Reserve University campus in Cleveland, OH, USA, where we have 122 individual PV power plants with microinverters and 32 PV modules connected to a DayStar Multitracer for acquisition of

I–V and P_{mp} time series with power readings taken every 1 min and I–V curve measured every 10 min [42]. This dataset has I–V curves from a standard multicrystalline silicon aluminum back-surface field (Al-BSF) module and a passivated emitter and rear cell (PERC) monocrystalline silicon module, with name-plate wattages of 279 and 315, respectively. The I–V curves in this dataset have 180–200 data points and are acquired using a DayStar Multitracer [43] in ambient conditions with varying irradiance and temperature. In this paper, we randomly select one I–V curve from this dataset to demonstrate our method.

The third dataset of I–V curves includes three different brands of monosilicon Al-BSF modules, with wattages of 285, 280, and 285, undergoing an accelerated indoor sequential exposure test consisting of 500 h of damp heat exposure, followed by 1000 cycles of dynamic mechanical loading (DH + DML sequential test), which is done stepwise to a total exposure of 4000 h of Damp Heat [44], [45]. In this dataset, each of the I–V curves consists of 3600–3800 data points. These I–V curves were acquired using a SPIRE 4600SLP flash tester [46] at standard test conditions (STC) (1 sun and 25°C). In this paper, we randomly select one I–V curve from this dataset to demonstrate our method.

III. RESULTS

In this section, we conduct a simulation study to validate the data-driven I–V feature extraction method on I–V curves generated with the single-diode model. We then apply the method to real-world I–V curves described above, acquired by different I–V scanning equipment, which produce different numbers of data points for each I–V curve.

A. I–V Curve Simulation Study

The single-diode model assumes that the dark current can be described by a single exponential dependence modified by the diode ideality factor n [47]. The current–voltage relationship is given by

$$I = I_{ph} - \frac{V + IR_s}{R_{sh}} - I_0 \left[\exp\left(\frac{V + IR_s}{nV_{th}}\right) - 1 \right] \quad (4)$$

where V and I are terminal voltage in volts and current in amperes, $I_{ph}(\approx I_{sc})$ is the photogenerated current, I_0 is the diode

TABLE I
PERCENT ERROR BETWEEN *I*–*V* FEATURE EXTRACTED RESULTS AND *I*–*V*
FEATURE SET VALUES BASED ON DIFFERENT NOISE LEVELS

Noise SD (A)	I_{sc}	R_{sh}	V_{oc}	R_s
0	0 %	0.13 %	0.09 %	61.67 %
	(99.99%)	(99.99%)	(99.98%)	(99.98%)
0.005	0 %	0.14 %	0.09 %	64.79 %
	(99.99%)	(99.95%)	(99.97%)	(99.97%)
0.010	0.02 %	2.99 %	0.10 %	66.66 %
	(99.98%)	(99.96%)	(99.97%)	(99.96%)
0.015	0.02 %	31.88 %	0.11 %	66.67 %
	(99.97%)	(99.95%)	(99.98%)	(99.97%)
0.020	0.03 %	51.16 %	0.12 %	67.71 %
	(99.97%)	(99.94%)	(99.97%)	(99.95%)

The repeatabilities are listed in parentheses.

TABLE II
AVERAGE PERCENT DIFFERENCE OF *I*–*V* FEATURE EXTRACTED RESULTS
FROM FRAUNHOFER-ISE LABORATORY REPORTED VALUES FOR OVER
2 200 000 *I*–*V* CURVES BY MODULE

	I_{sc}	R_{sh}	V_{oc}	R_s	P_{mp}	FF
GC1	0.57%	18.50%	0.29%	23.38%	1.90%	1.03%
GC2	1.06%	NA	0.39%	24.57%	5.00%	5.24%
GC3	8.52%	44.60%	0.65%	26.00%	3.17%	5.72%
NEG1	0.85%	26.00%	0.19%	16.00%	0.56%	0.59%
NEG2	0.30%	28.21%	0.16%	17.23%	2.62%	2.14%
NEG3	0.34%	NA	0.04%	19.24%	0.18%	0.44%
UFS1	6.36%	27.15%	2.22%	19.81%	4.60%	5.62%
UFS2	1.35%	21.20%	1.07%	21.93%	9.96%	9.74%
ALL	2.0%	24.65%	0.50%	21.10%	3.43%	2.90%

reverse saturation current, and V_{th} is the thermal voltage. It is well known that (4) is an implicit transcendental equation, which may not be solved explicitly in general for I and V using common elementary functions [48]. Therefore, one approach for exact explicit analytical solutions for I and V can be expressed using the Lambert W function, which is defined as the solution to the equation $W(x)\exp[W(x)] = x$, [6], [7] as follows:

$$I = \frac{(I_{ph} + I_0) - \frac{V}{R_{sh}}}{1 + \frac{R_s}{R_{sh}}}$$
$$- \frac{nV_{th}}{R_s} W\left[\frac{I_0 R_s}{nV_{th}(1 + \frac{R_s}{R_{sh}})} \exp\left(\frac{V + (I_{ph} + I_0)R_s}{nV_{th}(1 + \frac{R_s}{R_{sh}})}\right)\right] \quad (5)$$

and

$$V = (I_{ph} + I_0)R_{sh} - I(R_s + R_{sh})$$
$$- nV_{th} W\left[\frac{I_0 R_{sh}}{nV_{th}} \exp\left(\frac{(I_{ph} + I_0 - I)R_{sh}}{nV_{th}}\right)\right] \quad (6)$$

where

$$I_0 = \frac{I_{ph}}{\exp(\frac{V_{oc}}{N} - 1)} \quad (7)$$

and W represents the Lambert W function.

To illustrate the robustness of our data-driven I–V feature extraction method, we generate an I–V curve with 1000 observations (data points) based on the single-diode model and then use the algorithm to calculate I–V parameters including I_{sc}, V_{oc}, R_{sh}, and R_s. Let N_c be the number of cells, which is included in the V_{th} in (4), and $Temp$ denote the temperature. Setting $N_c = 60$, $Temp = 25\,°C$, $V_{oc} = 40.20$ V, $n = 1.5$, $I_{sc} = 8$ A, $R_{sh} = 600\,\Omega$, and $R_s = 0.48\,\Omega$, an I–V curve is generated from (5) using a sequence of 1000 points in V from 0 to V_{oc}. To the I values, we add random noise, which follows a normal distribution with zero mean and different SD, listed in Table I. Here, we generate I–V curves with noise levels between 0 and 0.02 A, which are typical noise levels for real-world I–V curves.

Table I shows the percent error and repeatability of four extracted I–V parameters for I–V curves generated using the diode model with the different levels of noise. We observe that the percent errors for I_{sc} and V_{oc} are low, which indicates that the data-driven I–V feature extraction method performs very well

in feature estimation for I_{sc} and V_{oc}. For R_{sh}, as the noise level increases, so does the percent error, with accuracy significantly decreasing at 0.015 A of noise on the simulated curve. For R_s, since the extracted values are higher than the set values, as has been demonstrated previously [26], the percent error is more than 61%. However, all extracted values are calculated with repeatability greater than 99.9%. Therefore, the data-driven I–V feature extraction method is a robust, practical, and easily implemented parameter extraction procedure for I–V curves.

B. Real-World I–V Curve Examples

1) Time-Series I–V Curves From the Fraunhofer-ISE Outdoor Dataset: We apply the data-driven feature extraction method to the dataset from Fraunhofer-ISE consisting of over 2 200 000 I–V curves. This dataset does not include nighttime values and was not filtered for this analysis. Table II shows the average percent difference of I–V feature extracted results to the reported values for each module. Note that there is no reported values for R_{sh} in GC2 and NEG3. The percent difference is small generally for I_{sc}, V_{oc}, P_{mp}, and FF, with only two modules with large difference.

2) Single I–V Curves From Various Sources: Fig. 6 shows examples of splined (red) and original (black) I–V curves from three real-world datasets, each measured with unique equipment and having a different number of datapoints and inherent noise. The I–V features extracted from these three curves are given along with the accompanying reported values in Table III.

The I–V curve in Fig. 6(a) was selected from the Fraunhofer-ISE dataset, from module GC1. The temperature and irradiance at the time of measurement were $23.6\,°C$ and 205.3 W/m^2, respectively. The I–V curve in Fig. 6(b) from the SDLE SunFarm was recorded for a 60-cell PERC module at 563.27 W/m^2 irradiance and $45.37\,°C$ temperature. The I–V curve in Fig. 6(c) was taken on a SPIRE 4600SLP flash tester at STC for a commercial module that had undergone damp heat + dynamic mechanical loading indoor accelerated testing. For the three I–V curves shown in Fig. 6, the extracted I–V feature values from our proposed method agree with reported values, and with greater than 99.9% reliability in all cases.

IV. DISCUSSION

A data-driven I–V feature extraction method to extract the solar cell I–V feature parameters has been developed. While

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.
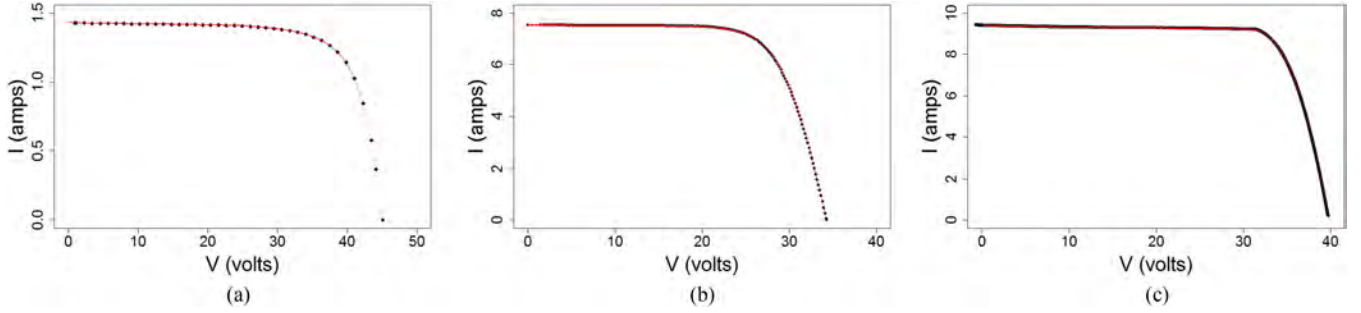
6

IEEE JOURNAL OF PHOTOVOLTAICS

Fig. 6. (a) I–V curve with 41 data points from the Fraunhofer-ISE outdoor dataset. (b) I–V curve with 186 data points from CWRU-SDLE SunFarm. (c) I–V curve with 3694 data points from the DH + DML Indoor Accelerated Test. The corresponding 500 generated I–V curve data point smoothing splines are shown in red.

TABLE III
I–V Feature Extracted Result With Repeatabilities for Three Single I–V Curves From Different Sources and the I–V Tracer Equipment Reported Values

Type	I_{sc} (A)	R_{sh} (Ω)	V_{oc} (V)	R_s (Ω)	$P_{mp}(W)$	FF (%)
			Fraunhofer-ISE Laboratory			
Extracted	1.429	1303.78	45.058	2.449	47.399	73.61
	(100%)	(99.94%)	(100%)	(99.99%)	(100%)	(100%)
Reported	1.429	1326.58	45.051	2.103	47.400	73.60
			CWRU-SDLE SunFarm			
Extracted	7.547	314.028	34.275	0.663	182.562	70.58
	(100%)	(99.93%)	(99.96%)	(99.91%)	(99.96%)	(99.95%)
Reported	7.541	NA	34.265	NA	182.535	70.64
			DH+DML Indoor Accelerated Test			
Extracted	9.409	108.013	39.84	0.488	292.114	77.93
	(100%)	(99.93%)	(99.96%)	(99.91%)	(99.96%)	(99.97%)
Reported	9.409	107.495	39.82	0.465	292.390	78.02

previous literature typically used the diode model or a combination of diode model and statistical methods to extract I–V features, they did not consider the occurrence of multiple steps in I–V curves [27] or the curvature that appears in I–V curves caused by measurement inaccuracies during the I–V curve tracer as it sweeps the voltage. Our proposed data-driven I–V feature extraction method makes corrections for the curvature issue and extracts the I–V features using computationally efficient data-driven algorithm, which enables analysis of massive numbers of I–V curves as are acquired as time-series datasets.

A. Accuracy and Repeatability of Extracted I–V Features

To illustrate the accuracy of our proposed data-driven I–V feature extraction method, we conducted a simulation study using diode-model-generated curves. The repeatabilities of all extracted I–V features are greater than 99.9%, as shown in Table I, which indicates that the data-driven I–V feature extraction method is robust in feature estimation. In the simulation study, the extracted I_{sc}, R_{sh}, and V_{oc} are very accurate compared with the values set in the single-diode model for curve generation. Note that the extracted R_{sh} becomes inaccurate when the noise reaches 0.015. Meanwhile, the value of the extracted R_s is approximately 62% higher than the set value. The percent deviation of the extracted R_s from its true value changes with the ideality factor of the diode model, indicating that the slope near V_{oc} is highly dependent on the cell recombination rates.

However, because the data-driven feature extraction method is highly repeatable using our automated algorithm, values of R_s produced this way are intercomparable. Care should be taken, however, in interpreting the absolute values of the extracted value of R_s, as this is an amalgamation of the actual R_s and recombination influences.

B. Robustness of Parameter Extraction From Different I–V Curve Sources

The time-series I–V data from the Fraunhofer-ISE dataset showed good agreement between extracted and reported values for most I–V features across $2\,200\,000$ I–V curves. Extracted R_{sh} and R_s exhibited systematic differences from reported values, as expected based on prior studies comparing linear methods for determining these values, as discussed earlier.

Certain modules had high percent difference for other parameters. One reason for particular modules' high percent difference may be due to atypically shaped I–V curves that are not adequately handled by traditional feature extraction methods. For example, a module with poor electrical connection with an I–V curve, as shown in Fig. 5, would have consistently larger reported I_{sc} and lower reported R_{sh} than those obtained with our algorithm. We suspect this is the case for modules GC3 and UFS1. The difference in P_{mp} may be the result of splining, as the reported result uses the measured data points (40–42 data points) in I–V curves, while we use 500 data points, which are closer to underlying I–V curves of the module.

The quality of I–V curve data has a strong influence on extracting the I–V features. In the Fraunhofer-ISE dataset, I–V curves have only 40–42 data points each, leading to inaccuracy in estimating R_{sh}, R_s, and P_{mp} on the original curves. We use a smoothing spline function to fit this data and generate 500 data points from this curve in order to make the result more accurate and repeatable. For the I–V curve data from the DH + DML Indoor Accelerated Test, which has 3600–3800 data points, we still generate 500 data points and found that the repeatability is 99.98%. Therefore, using 500 data points is sufficient for accurate extracted I–V features from a range of data sources. In addition, I–V curves with different numbers of observations make it hard to set a uniform criteria in the function (i.e., the critical value to find the linear region) and would be problematic

when dealing with a large number of I–V curves. By splining a curve to 500 data points, we can use uniform criteria and, thus, apply our algorithm more broadly. The repeatability for I_{sc}, V_{oc}, R_s, R_{sh}, P_{mp}, and FF is greater than 99.9% for all curves tested here; therefore, the technique is robust for highly varied sources of I–V curves.

Many I–V tracer systems also report values of the I–V features, but without disclosing the algorithms used to determine these parameters. This is the case for the Spire, DayStar Multitracer, and the ESL systems used as the data acquisition sources in this paper. This leads to the inherent obfuscation of the meaning and accuracy of reported feature parameters from these pieces of equipment. By using a common analytical package based in open-source algorithms and codes, one is able to analyze I–V curves from diverse instruments and arrive at I–V feature parameters with a common basis. This is an example of strong scientific advantages of open-source software, codes, and algorithms [49], [50].

C. Computational Efficiency I–V Feature Extraction

For the Fraunhofer-ISE dataset, there are a total of 2.2 million I–V curves [35], [36]. The computation of extracted I–V features took approximately 3 h using Simple Linux Utility for Resource Management computing resource on a single machine with specifications: a compute node of High Performance Computing server has Intel(R) Xeon(R) CPU E5-24500 @ 2.10-GHz processor, 24-GB memory, and 12 CPU cores × 2.69 GHz.

V. CONCLUSION

In this paper, we have developed a data-driven I–V feature extraction method to extract features from I–V curves and calculate the repeatabilities of each I–V feature. Three different datasets have been used to demonstrate how this method can be applied in practice. Moreover, we have conducted a simulation study to illustrate the accuracy and reproducibility of the extracted I–V features by generating I–V curves from the single-diode model. Our proposed method performs very well in I–V feature estimation for I_{sc}, R_{sh}, and V_{oc}, while the estimation of R_s shows predictable error. All values are estimated with very high repeatability. Therefore, the data-driven I–V feature extraction method is an accurate, robust, and fast parameter extraction procedure for characterizing large volumes of PV module I–V data.

ACKNOWLEDGMENT

REFERENCES

[1] R. H. French *et al.*, "Degradation science: Mesoscopic evolution and temporal analytics of photovoltaic energy materials," *Current Opin. Solid State Mater. Sci.*, vol. 19, no. 4, pp. 212–226, 2015.

[2] E. L. Meyer and E. E. van Dyk, "Assessing the reliability and degradation of photovoltaic module performance parameters," *IEEE Trans. Rel.*, vol. 53, no. 1, pp. 83–92, Mar. 2004.

[3] E. E. van Dyk and E. L. Meyer, "Analysis of the effect of parasitic resistances on the performance of photovoltaic modules," *Renew. Energy*, vol. 29, no. 3, pp. 333–344, Mar. 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148103002507

[4] E. E. van Dyk, A. R. Gxasheka, and E. L. Meyer, "Monitoring current-voltage characteristics and energy output of silicon photovoltaic modules," *Renew. Energy*, vol. 30, no. 3, pp. 399–411, Mar. 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148104002447

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

[6] A. Jain and A. Kapoor, "Exact analytical solutions of the parameters of real solar cells using Lambert W-function," *Sol. Energy Mater. Sol. Cells*, vol. 81, no. 2, pp. 269–277, Feb. 2004. [Online]. Available: https://doi.org/10.1016/j.solmat.2003.11.018

[7] A. Jain, S. Sharma, and A. Kapoor, "Solar cell array parameters using Lambert W-function," *Sol. Energy Mater. Sol. Cells*, vol. 90, no. 1, pp. 25–31, Jan. 2006. [Online]. Available: https://doi.org/10.1016/j.solmat.2005.01.007

[8] A. Ortiz-Conde, F. J. García Sánchez, and J. Muci, "New method to extract the model parameters of solar cells from the explicit analytic solutions of their illuminated I-V characteristics," *Sol. Energy Mater. Sol. Cells*, vol. 90, no. 3, pp. 352–361, Feb. 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0927024805001170

[9] C. Zhang, J. Zhang, Y. Hao, Z. Lin, and C. Zhu, "A simple and efficient solar cell parameter extraction method from a single current-voltage curve," *J. Appl. Phys.*, vol. 110, no. 6, Sep. 2011, Art. no. 064504. [Online]. Available: https://aip.scitation.org/doi/abs/10.1063/1.3632971

[10] K. Ishaque, Z. Salam, and Syafaruddin, "A comprehensive MATLAB Simulink PV system simulator with partial shading capability based on two-diode model," *Sol. Energy*, vol. 85, no. 9, pp. 2217–2227, 2011.

[11] M. Haouari-Merbah, M. Belhamel, I. Tobías, and J. M. Ruiz, "Extraction and analysis of solar cell parameters from the illuminated current-voltage curve," *Sol. Energy Mater. Sol. Cells*, vol. 87, no. 1, pp. 225–233, May 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0927024804003411

[12] L. Sandrolini, M. Artioli, and U. Reggiani, "Numerical method for the extraction of photovoltaic module double-diode model parameters through cluster analysis," *Appl. Energy*, vol. 87, no. 2, pp. 442–451, Feb. 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306261909003092

[13] F. Ghani, G. Rosengarten, M. Duke, and J. K. Carson, "The numerical calculation of single-diode solar-cell modelling parameters," *Renew. Energy*, vol. 72, pp. 105–112, Dec. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148114003759

[14] R. Khezzar, M. Zereg, and A. Khezzar, "Comparative study of mathematical methods for parameters calculation of current-voltage characteristic of photovoltaic module," in *Proc. Int. Conf. Elect. Electron. Eng.*, Nov. 2009, pp. I-24–I-28.

[15] J. Del Cueto, S. Rummel, B. Kroposki, C. Osterwald, and A. Anderberg, "Stability of CIS/CIGS modules at the outdoor test facility over two decades," in *Proc. 33rd IEEE Photovolt. Spec. Conf.*, 2008, pp. 1–6.

[16] N. Kato *et al.*, "Degradation analysis of dye-sensitized solar cell module after long-term stability test under outdoor working condition," *Sol. Energy Mater. Sol. Cells*, vol. 93, no. 6, pp. 893–897, Jun. 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0927024808003644

[17] S. Kichou *et al.*, "Characterization of degradation and evaluation of model parameters of amorphous silicon photovoltaic modules under outdoor long term exposure," *Energy*, vol. 96, pp. 231–241, Feb. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544215016965

[18] S. Kichou *et al.*, "Study of degradation and evaluation of model parameters of micromorph silicon photovoltaic modules under outdoor long term exposure in Jaén, Spain," *Energy Convers. Manag.*, vol. 120, pp. 109–119, Jul. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0196890416303430

[19] M. H. Ali, A. Rabhi, A. E. Hajjaji, and G. M. Tina, "Real time fault detection in photovoltaic systems," *Energy Procedia*, vol. 111, pp. 914–923, Mar. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1876610217302874

[20] L. S. Bruckman *et al.*, "Statistical and domain analytics applied to PV module lifetime and degradation science," *IEEE Access*, vol. 1, pp. 384–403, 2013.

[21] W.-H. Huang *et al.*, "netSEM: Network Structural Equation Modeling," Jun. 2018. [Online]. Available: https://CRAN.R-project.org/package= netSEM

[22] D. Pysch, A. Mette, and S. W. Glunz, "A review and comparison of different methods to determine the series resistance of solar cells," *Sol. Energy Mater. Sol. Cells*, vol. 91, no. 18, pp. 1698–1706, 2007.

[23] L. H. I. Lim, Z. Ye, J. Ye, D. Yang, and H. Du, "A linear method to extract diode model parameters of solar panels from a single I-V curve," *Renew. Energy*, vol. 76, pp. 135–142, Apr. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148114007332

[24] A. J. Bühler, F. Perin Gasparin, and A. Krenzinger, "Post-processing data of measured I-V curves of photovoltaic devices," *Renew. Energy*, vol. 68, no. Suppl. C, pp. 602–610, Aug. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148114001396

[25] D. Sera and R. Teodorescu, "Robust series resistance estimation for diagnostics of photovoltaic modules," in *Proc. 35th Annu. Conf. IEEE Ind. Electron. Soc.*, 2009, pp. 800–805.

[26] K. Tada, "What do apparent series and shunt resistances in solar cell estimated by I–V slope mean? Study with exact analytical expressions," *Phys. Status Solidi (a)*, vol. 215, 2018, Art. no. 1800448.

[27] T. J. Peshek *et al.*, "Insights into metastability of photovoltaic materials at the mesoscale through massive I–V analytics," *J. Vacuum Sci. Technol. B, Nanotechnol. Microelectron.: Mater., Process., Meas., Phenom.*, vol. 34, no. 5, 2016, Art. no. 050801.

[28] A. Bellini, S. Bifaretti, V. Iacovone, and C. Cornaro, "Simplified model of a photovoltaic module," in *Proc. Appl. Electron.*, Sep. 2009, pp. 47–51.

[29] Solmetric, "Guide to interpreting IV curve measurements of PV arrays," 2010. [Online]. Available: http://resources.solmetric.com/get/Guide%20 to%20Interpreting%20I-V%20Cu%rves.pdf

[30] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R* (Springer Texts in Statistics), 1st ed. New York, NY, USA: Springer, Aug. 2013. [Online]. Available: http://www-bcf.usc.edu/gareth/ISL/index.html

[31] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Found. Statist. Comput., 2018. [Online]. Available: https://www.R-project.org/

[32] V. M. Muggeo, "Segmented: An R package to fit regression models with broken-line relationships," *R News*, vol. 8, no. 1, pp. 20–25, 2008.

[33] V. M. Muggeo, "Package 'segmented'," *Biometrika*, vol. 58, pp. 525–534, 2017.

[34] W.-H. Huang *et al.*, "ddiv: Data Driven I-V feature extraction," Sep. 2018. [Online]. Available: https://CRAN.R-project.org/package=ddiv

[35] J. Liu *et al.*, "Cross-correlation analysis of the indoor accelerated and real world exposed photovoltaic systems across multiple climate zones," in *Proc. IEEE 7th World Conf. Photovolt. Energy Convers.*, 2018, pp. 3949–3954, doi: 10.1109/PVSC.2018.8547840.

[36] M. Wang *et al.*, "Evaluation of photovoltaic module performance using novel data-driven I-V feature extraction and Suns-VOC determined from outdoor time-series I-V Convers," in *Proc. IEEE 7th World Conf. Photovolt. Energy Convers.*, 2018, pp. 778–783, doi: 10.1109/PVSC.2018.8547772.

[37] F. Rubel and M. Kottek, "Observed and projected climate shifts 1901–2100 depicted by world maps of the Koppen-Geiger climate classification," *Meteorologische Zeitschrift*, vol. 19, no. 2, pp. 135–141, Apr. 2010.

[38] M. C. Peel, B. L. Finlayson, and T. A. McMahon, "Updated world map of the Köppen-Geiger climate classification," *Hydrol. Earth Syst. Sci.*, vol. 11, no. 5, pp. 1633–1644, Oct. 2007. [Online]. Available: https://www.hydrol-earth-syst-sci.net/11/1633/2007/

[39] F. Rubel, K. Brugger, K. Haslinger, and I. Auer, "The climate of the European Alps: Shift of very high resolution Köppen-Geiger climate zones 1800–2100," *Meteorol. Zeitschrift*, vol. 26, pp. 115–125, 2017. [Online]. Available: http://www.schweizerbart.de/papers/metz/detail/prepub/87237/The_climate%_of_the_European_Alps_Shift_of_very_hig?af=crossref

[40] C. Bryant, N. R. Wheeler, F. Rubel, and R. H. French, "KGC: Koeppen-Geiger climatic zones," Nov. 2017. [Online]. Available: https://cran.r-project.org/web/packages/kgc/index.html

[41] ET Instrumente, "Series ESL-Solar 500," 2018. [Online]. Available: https://et-instrumente.de/index.php/en/products2/overview-electronic-lo%ads/pv-modul-test-unit

[42] Y. Hu *et al.*, "A nonrelational data warehouse for the analysis of field and laboratory data from multiple heterogeneous photovoltaic test sites," *IEEE J. Photovolt.*, vol. 7, no. 1, pp. 230–236, Jan. 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7763779/

[43] DayStar, "Multi-Tracer," 2018. [Online]. Available: http://www.daystarpv.com/multitracer3.html

[44] *Terrestrial Photovoltaic (PV) Modules—Design Qualification and Type Approval—Part 2: Test Procedures*, IEC 61215-2:2016, 2016. [Online]. Available: https://webstore.iec.ch/publication/24311

[45] *Cyclic (Dynamic) Mechanical Load Test—Photovoltaic (PV) Modules*, IEC TS 62782:2016, 2018. [Online]. Available: https://webstore.iec.ch/publication/24310

[46] Eternal Sun, "Spi-Sun Simulator 4600slp," 2018. [Online]. Available: https://www.spiresolar.com/products/previous-models/spi-sun-simulat-or-4%600slp/

[47] G. K. Singh, "Solar power generation by PV (photovoltaic) technology: A review," *Energy*, vol. 53, pp. 1–13, May 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544213001758

[48] E. Kreyszig, *Advanced Engineering Mathematics*, 10th ed. Hoboken, NJ, USA: Wiley, Aug. 2011.

[49] D. C. Ince, L. Hatton, and J. Graham-Cumming, "The case for open computer programs," *Nature*, vol. 482, no. 7386, pp. 485–488, Feb. 2012. [Online]. Available: http://www.nature.com/nature/journal/v482/n7386/full/nature10836.html

[50] J. S. S. Lowndes *et al.*, "Our path to better science in less time using open data science tools," *Nature Ecol. Evol.*, vol. 1, no. 6, Jun. 2017, Art. no. 0160. [Online]. Available: https://www.nature.com/articles/s41559-017-0160