Noise-Scaled Euclidean Distance: A Metric for Maximum Likelihood Estimation of the PV Model Parameters

Efstratios Batzelis, *Senior Member, IEEE*, José M. Blanes, F. Javier Toledo, and Vicente Galiano

***Abstract*—This article revisits the objective function (or metric) used in the extraction of photovoltaic (PV) model parameters. A theoretical investigation shows that the widely used current distance (CD) metric does not yield the maximum likelihood estimates (MLE) of the model parameters when there is noise in both voltage and current samples. It demonstrates that the Euclidean distance (ED) should be used instead, when the voltage and current noise powers are equal. For the general case, a new noise-scaled Euclidean distance (NSED) metric is proposed as a weighted variation of ED, which is shown to fetch the MLE of the parameters at any noise conditions. This metric requires the noise ratio (i.e., ratio of the two noise variances) as an additional input, which can be estimated by a new noise estimation (NE) method introduced in this study. One application of the new metric is to employ NSED regression as a follow-up step to existing parameter extraction methods toward fine-tuning of their outputs. Results on synthetic and experimental data show that the so-called NSED regression "add-on" improves the accuracy of five such methods and validate the merits of the NSED metric.**

***Index Terms*—Euclidean distance (ED), fitting, noise extraction (NE), orthogonal distance, parameter estimation, parameter extraction, parameter identification, photovoltaic (PV) model, regression.**

I. Introduction

ONE of the most long-standing research problems in the field of photovoltaic (PV) generators is the extraction of the PV model parameters. This research topic, known as parameters extraction, identification, or estimation, refers to identifying the equivalent circuit parameters of the PV cell or module, which are not usually available in the manufacturer's datasheet. This remains to this day an active research area [1], with a myriad of relevant methods in the literature, greatly varying in complexity, input data, and their very nature.

The input data employed in these methods is usually either a few *key points* or several samples of the current-voltage characteristic in the first quadrant (*I–V curve*). The former class involves the coordinates of noteworthy operating points [e.g., maximum power point (MPP), short-circuit (SC) current, open-circuit (OC) voltage] and possibly a few additional inputs. The relevant parameter extraction methods usually form a system of equations, solved either numerically (*numerical* methods) [2], [3] or explicitly after simplifications (*explicit* methods) [4]–[6]. All these variants are relatively simple, but their accuracy is sensitive on the quality of their limited inputs.

The alternatives that employ the entire I–V curve dataset are generally considered more complex, but also more reliable [7]. This cluster, referred to as *optimization* methods (or curve fitting, or metaheuristic), aims to fit the PV model on the I–V dataset by minimizing an objective function [1], [8]–[14]. Although there is abundant literature on the minimization algorithm employed, most commonly an evolutionary-based technique [7], [10]–[14], the objective function itself has been overlooked and taken for granted in most cases. In the majority of relevant methods, the objective function (or metric, or fitness function) is based on the PV current, expressed in various forms: most commonly root mean square error (RMSE) [8], [14], sum of squared errors (SSE) [9], [10], and absolute error (AE) [13]. It is indicative that almost all methods reviewed in [7] adopt such a current-based criterion, either one of the above metrics or a normalized/mean variation.

However, although current regression is intuitive and simple, it is not necessarily the ideal tool to extract the model parameters [1]. In fact, it is known since 1986 that a current-based metric skews the results in favor of the high-voltage region at the expense of the remaining part of the curve [15]. This arises from the fact that measured I–V curves feature distortion (noise) in both current and voltage samples, which effectively renders this a *multivariate noise* problem (or error-in-variables) according to statistics theory [16]. In such problems, the fundamental assumption of ordinary least squares (OLS) optimization for observation error only on the dependent variable, i.e., the PV current, does *not* hold true. This challenge is reaffirmed by more recent studies as well [14], [17], [18].

Manuscript received November 2, 2021; revised January 20, 2022 and February 18, 2022; accepted March 4, 2022. Date of publication April 1, 2022; date of current version April 21, 2022. The work of Efstratios Batzelis was supported by the Royal Academy of Engineering under the Engineering for Development Research Fellowship scheme under Grant RF2018191886. The work of F. Javier Toledo was supported by the Ministerio de Ciencia e Innovación of Spain under Grant PGC2018-097960-B-C21, in part by the government of the Valencian Communityunder Grant PROMETEO/2021/063, and in part by the European Union (ERDF, "A way to make Europe"). The work of Vicente Galiano was supported by the Spanish Ministry of Science, Innovation, and Universities under Grant RTI2018-098156-B-C54 and in part by the FEDER funds (MINECO/FEDER/UE). *(Corresponding author: Efstratios Batzelis.)*

Efstratios Batzelis is with the University of Southampton, SO17 1BJ Southampton, U.K. (e-mail: e.batzelis@soton.ac.uk).

José M. Blanes, F. Javier Toledo, and Vicente Galiano are with Miguel Hernández University, 03202 Elche, Spain (e-mail: jmblanes@umh.es; javier.toledo@umh.es; vgaliano@umh.es).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JPHOTOV.2022.3159390.

Digital Object Identifier 10.1109/JPHOTOV.2022.3159390

For this reason, quite a few alternative metrics have been proposed in the past. The area between the measured and reconstructed I–V curve is adopted as objective function in [15] and as an evaluation criterion in [6] and [19]. Other alternatives include: the power absolute error in [10], a metric based on the derivative of power with respect to voltage in [11], a combination of the SC, OC, and MPP absolute errors in [12], and a hybrid current-voltage criterion in [14]. Although these studies back the superiority of their metrics via results, there is no theoretical justification on their optimality.

Another noteworthy alternative employs the *Euclidean distance* (ED) (or orthogonal distance), i.e., the shortest distance of the samples to the curve. Minimization of the ED metric [Euclidean distance regression (EDR) or orthogonal Regression] is a total least squares problem. In linear models, EDR is proven to have an explicit solution and fetch the maximum likelihood estimates (MLE) of the parameters at normally distributed noise conditions with known noise variances [20], [21]. The MLE of the parameters are those values that make the observed data most likely and are widely seen among the most reliable estimations of the original/true parameters. However, in nonlinear models, such as the single-diode PV model, EDR can only lead to an approximation of the MLEs and its application is not a trivial task. Boggs *et al.* [22] introduced ODRPACK in the 1990s, [23], a FORTRAN-based software to compute EDR in nonlinear models by successive numerical iteration steps based on Levenberg–Marquardt. However, ODRPACK is not specifically designed for PV models, which entails potential computational challenges in accommodating the implicit model equations, and is not able to calculate the noise variances that are quite challenging to capture [21].

A thorough literature survey reveals that there is only a handful of research papers that apply ED-based regression on PV models [1], [17], [18], [24]. Burgers *et al.* [17] were the first to discuss in 1996 the merits of EDR over the conventional current-based regression, but the proposed implementation simplifies the ED to a weighted current-based metric. This ED approximation is also adopted in [18]. Another study [1], employs ODRPACK for the PV parameters extraction, but it omits the noise variances. Similarly, Diantoro *et al.* [24] briefly mentioned application of orthogonal regression but does not disclose the formulation details. In fact, this investigation concludes that the EDR subject has been treated only superficially in the field of PV modeling, without fully exploring its appropriateness for the parameters extraction or giving the implementation details.

This article reexamines the application of ED in the field of PV modeling using the MLE statistics theory. The widely used current distance (CD) (or vertical distance) metric is reassessed and shown that it yields estimates skewed by noise in case of both voltage and current distortion. Instead, a mathematical analysis articulated in engineering language demonstrates that the ED metric approximates the preferable MLE of the model parameters, as long as the noise power in voltage and current are equal. For the general multivariate noise case, the *noise-scaled Euclidean distance* (NSED) metric is introduced, applied for the first time to this problem, based on a weighted version of ED that is scaled on the noise ratio of voltage and current.

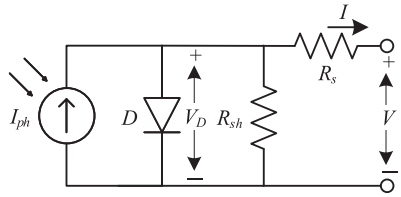


Fig. 1. Equivalent circuit of the single-diode PV model.

A theoretical assessment proves that only NSED among the aforementioned metrics fetches the MLE of the parameters in the general noise case. The proposed NSED formulation includes also a new problem-specific equation for the calculation of the ED of a sample to the curve based on the single-diode PV model.

Furthermore, the calculation of the noise ratio has not been treated in any of the aforementioned studies [1], [17], [18], [24], although it is a requirement for NSED and any other noise-weighted metric. This article proposes also a new noise extraction (NE) method to identify the noise variables using only the I–V curve as input when the voltage is sampled linearly, i.e., the voltage samples are equidistant.

Simulation and experimental results on the NREL dataset [25] validate the superiority of NSED over the common CD and unscaled ED metrics, giving rise to many applications in the field of PV modeling. One such application discussed later is the enhancement of existing parameter extraction methods by applying NSED regression (NSEDR) as a second-stage step toward fine-tuning of their parameter estimations. The so-called *NSEDR add-on* is shown to improve the accuracy of five methods examined from the literature.

The rest of this article is organized as follows. Section II provides the mathematical formulation of the different metrics considered, which are then theoretically analyzed in Section III. The proposed NE method is described in Section IV, which provides the noise ratio required for the NSEDR add-on discussed in Section V. Section VI performs a detailed comparison on synthetic and experimental data to validate the superiority of the NSED metric and the usefulness of the NSEDR add-on. Finally, Section VII concludes this article.

II. Metrics for the Extraction of the Photovoltaic Model Parameters

The widely used single-diode PV model of Fig. 1 involves five parameters: 1) the photocurrent I_{ph}, 2) the diode's saturation current I_s, 3) the diode's modified ideality factor a, 4) the series resistance R_s, and 5) shunt resistance R_{sh} [4]. This model can sufficiently describe any PV generator, as long as there is uniformity in the structural characteristics and operating conditions. The model's current-voltage $(I$–$V)$ equation is given in implicit form (I and V in both sides of the equation) by

$$I = I_{ph} - I_s\left(e^{\frac{V+IR_s}{a}} - 1\right) - \frac{V + IR_s}{R_{sh}} \tag{1}$$

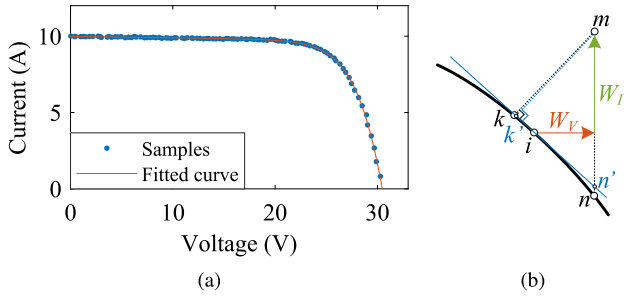


Fig. 2. Examples of (a) fitting the PV model on samples and (b) a noisy sample m projected vertically n and orthogonally k on the curve.

and in explicit formulation (I and V separated in the two sides of the equation) by

$$I = \frac{R_{sh}I_{ph} - V}{R_s + R_{sh}} - \frac{a}{R_s} W\left\{\frac{R_s R_{sh} I_s}{a(R_s + R_{sh})} e^{\frac{R_{sh}(R_s I_{ph} + V)}{a(R_s + R_{sh})}}\right\} = f(V, \boldsymbol{p}) \quad (2)$$

where $\boldsymbol{p} = [I_{ph}, I_s, a, R_s, R_{sh}]$ is the parameters vector and $W\{x\}$ stands for the Lambert W function [5], [26], [27]. Equation (1) requires numerical/iterative solution, whereas (2) allows for more straightforward evaluation. Numerical implementations of the Lambert W function are readily available in computational platforms, such as MATLAB [28] and SciPy [29], while cost-efficient series expansions may also be found in [30].

The five parameters of the model are not freely available, but they have to be extracted either from datasheet information or a measured I–V curve. The optimization methods act upon the entire I–V curve and aim to identify $\boldsymbol{p}$ by fitting the PV model $I = f(V, \boldsymbol{p})$ on a set of (V_m, I_m) samples, as illustrated in Fig. 2(a). These samples always feature a certain level of observation error (measurement noise) induced mainly by sensors tolerance and measurement errors. This fitting is an optimization problem seeking to find $\boldsymbol{p}$ that minimizes the objective function.

The notations used hereinafter are shown in Fig. 2(b). Every original (true) sample i is distorted by additive noise in both voltage W_V and current W_I, so that the reading m is given by $(V_m, I_m) = (V_i + W_V, I_i + W_I)$. Throughout the article, the indices i and m refer to the original and distorted sample, and the notation W_X to noise variables not to be confused with the Lambert W function $W\{x\}$. The noise variables are assumed to be normally distributed with zero mean value and positive variance, i.e., $W_V \sim \mathcal{N}(0, \sigma_V^2)$ and $W_I \sim \mathcal{N}(0, \sigma_I^2)$. The noise and original samples are considered uncorrelated variables, while the ratio of σ_V over σ_I has particular significance and is denoted as *noise ratio* $\sigma_n = \sigma_V / \sigma_I$.

It is worth noting that these assumptions are realistic, but may not always hold true. For example, the measurement noise in voltage and current may exhibit some correlation in case of a common distortion source or the single-diode model adopted may not be the exact physical equivalent for some PV modules and conditions. The analysis that follows is subject to these assumptions. The remainder of this section outlines the formulation of the CD, ED, and NSED metrics, including all calculation steps. The appropriateness of these metrics is theoretically assessed subsequently in Section III.

A. Current Distance Metric

The majority of regression methods adopt the RMSE or SSE of the PV current [8]–[10], [14], which corresponds to the standard OLS minimization. These two metrics are equivalent optimizationwise, so the latter is selected here for the analysis. The sum of squared current deviations of every sample from the curve (i.e., the vertical distances mn in Fig. 2) is denoted here as the CD metric and is given by

$$S_{CD} = \sum \left[f(V_m, \boldsymbol{p}) - I_m\right]^2. \quad (3)$$

Please note that the summation index is implied here and throughout the article. Minimizing S_{CD} is referred to as current distance regression (CDR), while (2) can be used for $f(V_m, \boldsymbol{p})$ for explicit formulation. A similar expression can be derived in terms of voltage, although it is not often used in this problem.

B. Euclidean Distance Metric

The ED in PV modeling is either simplified or not explicitly calculated in [1], [17], [18], and [24]. This section provides this formulation with more detail. ED is the shortest distance of the sample to the model [i.e., distance mk in Fig. 2(b)], and the ED metric is the sum of squared such distances

$$S_{ED} = \sum (V_m - V_k)^2 + (I_m - I_k)^2 \quad s.t. \quad (V_k, I_k) = \underset{I = f(V, \boldsymbol{p})}{\text{proj}} (V_m, I_m) \quad (4)$$

where the notation $\underset{I=f(V,\boldsymbol{p})}{\text{proj}} (V_m, I_m)$ represents the closest point of the I–V curve to the sample (V_m, I_m) (shortest distance), i.e., the closest orthogonal projection of the sample on the model $I = f(V, \boldsymbol{p})$.

The calculation of projection (V_k, I_k) is not a straightforward task because of the nonlinear nature of the PV model. Orthogonal projection implies that the tangent line at k will be perpendicular to the line mk, that is

$$\left.\frac{dI}{dV}\right|_k = -\frac{V_m - V_k}{I_m - I_k} \quad (5)$$

where the derivative of current w.r.t. voltage is given by [31]

$$\left.\frac{dI}{dV}\right|_k = -\frac{R_{sh}I_s e^{\frac{V_k + I_k R_s}{a}} + a}{R_s R_{sh} I_s e^{\frac{V_k + I_k R_s}{a}} + a(R_s + R_{sh})}. \quad (6)$$

The solution (V_k, I_k) is then found by numerically solving a system of two equations: the model (1) and the slope equation derived by substituting (6) into (5).

Alternatively, this problem can be reduced to a single equation if the model diode voltage $V_D = V_k + I_k R_s$ is used as an auxiliary variable (see Fig. 1). Writing the model (1) and derivative

(6) in terms of V_D yields

$$I_k = I_{ph} - I_s\left(e^{\frac{V_D}{a}} - 1\right) - \frac{V_D}{R_{sh}} \tag{7}$$

$$\left.\frac{dI}{dV}\right|_k = -\frac{R_{sh}I_s e^{\frac{V_D}{a}} + a}{R_s R_{sh} I_s e^{\frac{V_D}{a}} + a(R_s + R_{sh})} \tag{8}$$

which can be then substituted into (5) to yield an expression with only V_D as unknown

$$\begin{aligned}&\left[I_{ph} - I_s e^{\frac{V_D}{a}} + I_s - \frac{V_D}{R_{sh}} - I_m\right] \times \\ &\quad\left[R_s + (1+R_s^2)\left(\frac{1}{R_{sh}} + \frac{I_s}{a}e^{\frac{V_D}{a}}\right)\right] \\ &= (V_D - V_m - I_m R_s)\left[1 + R_s\left(\frac{1}{R_{sh}} + \frac{I_s}{a}e^{\frac{V_D}{a}}\right)\right].\end{aligned} \tag{9}$$

Therefore, to find the projection k one has to numerically solve (9) for V_D, and then use it to get I_k from the explicit (7), and, finally, find $V_k = V_D - I_k R_s$. It is worth noting that (9) may have theoretically more than one solution in some cases. Initializing V_D with $V_m + I_m R_s$ usually yields the correct solution becuase of the proximity of the samples to the curve.

C. Noise-Scaled Euclidean Distance Metric

The NSED metric proposed in this article is a generalization of ED aiming to equalize the noise impact in voltage and current. One option to that end is to normalize (or scale) the voltage on σ_V and the current on σ_I. Alternatively, it is simpler to scale only the voltage on the noise ratio $\sigma_n = \sigma_V/\sigma_I$ for the same effect. This operation, referred to thereafter simply as *model scaling*, is essentially a transformation of the PV model $I = f(V, \boldsymbol{p})$ to the equivalent $I = \tilde{f}(v, \boldsymbol{\rho})$, having

$$v = V/\sigma_n \tag{10}$$

$$\boldsymbol{\rho} = [I_{ph}, I_s, a/\sigma_n, R_s/\sigma_n, R_{sh}/\sigma_n]. \tag{11}$$

In other words, the scaled model uses the same fundamental (1) but with modified parameters and the sampled voltages V_m are normalized on the noise ratio, i.e., $v_m = v_i + w_V = V_i/\sigma_n + W_V/\sigma_n$. The new voltage noise $w_V = W_V/\sigma_n$ is a scaled version of the original voltage noise W_V. The gain from this transformation is that the noise variance in the voltage and current become now equal, as shown below using the expectation operator $\mathbf{E}(.)$

$$\mathbf{E}(w_V^2) = \mathbf{E}\left(\frac{W_V^2}{\sigma_n^2}\right) = \mathbf{E}\left(\frac{W_V^2}{\sigma_V^2}\sigma_I^2\right) = \sigma_I^2 = \mathbf{E}(W_I^2). \tag{12}$$

This allows an EDR acting upon the scaled model to fetch the MLE of the original parameters, explained in detail later.

Therefore, the proposed NSED metric is finally defined as

$$\begin{aligned}S_{\text{NSED}} &= \sum (v_m - v_k)^2 + (I_m - I_k)^2 \\ s.t.\ \ (v_k, I_k) &= \underset{I=\tilde{f}(v,\boldsymbol{\rho})}{\text{proj}}\ (v_m, I_m)\end{aligned} \tag{13}$$

where v_m and $\boldsymbol{\rho}$ are given by eqs. 10 and 11 and the projection k by solving (9) as explained in Section II-B. Please note that application of the NSED metric requires the value of the noise ratio σ_n as an additional input, as compared with the CD and ED metrics discussed previously. Minimization of this metric is denoted hereinafter as *NSED regression* (NSEDR).

III. Maximum Likelihood Estimates Assessment of the Three Metrics

This section sheds some light into why some metrics perform better than others in approximating the MLE of the model parameters and why the NSED metric is optimal in the MLE sense. The following analysis considers the equivalent linearized PV model, according to which the model around a point i is represented by the tangent line at that point [see Fig. 2(b)]. This simplifies the PV model in the vicinity of i to

$$I = f(V, \boldsymbol{p})|_i \approx \lambda_i V + \beta_i. \tag{14}$$

Please note that the coefficients λ_i and β_i differ for every point i. This intervention allows approximation of the vertical distance mn by mn' and the orthogonal distance mk by mk' required for the analysis below [see Fig. 2(b)]. It is worth noting that the concept of linearization is widely used in many scientific fields (e.g., small-signal stability analysis) and remains valid as long as the linearized region is reasonably small; this holds true in this application with noisy samples sufficiently close to the I–V curve.

A. CDR Performance

The CDR entails minimization of the CD metric given in (3), which is approximated using the linearized PV model

$$\begin{aligned}S_{CD} &\approx \sum [I_m - \lambda_i V_m - \beta_i]^2 \\ &= \sum [I_i + W_I - \lambda_i(V_i + W_V) - \beta_i]^2 = S_{CD}^0 + A + B\end{aligned} \tag{15}$$

where S_{CD}^0 is the CD metric as if there was no noise in the measurements, i.e., the nonnoisy CD sum, and the terms A and B are the remainders because of noise

$$S_{CD}^0 = \sum [I_i - \lambda_i V_i - \beta_i]^2 \tag{16}$$

$$A = 2\sum [I_i - \lambda_i V_i - \beta_i]\,[W_I - \lambda_i W_V] \tag{17}$$

$$B = \sum [W_I - \lambda_i W_V]^2. \tag{18}$$

Since the noise variables W_V and W_I have zero mean value and they are uncorrelated to the original sample (V_i, I_i), A gets zero on average ($\mathbf{E}(A) = 0$) and $\mathbf{E}(B) = N\sigma_I^2 + N\sigma_V^2\mathbf{E}(\lambda_i^2)$, where N is the number of samples (details in Appendix A). Therefore, the expected CD sum equals the expected nonnoisy sum plus some noise-inflicted terms that involve the slopes λ_i, i.e. they are functions of the model parameters:

$$\mathbf{E}(S_{CD}) \approx \mathbf{E}(S_{CD}^0) + N\sigma_I^2 + N\sigma_V^2\mathbf{E}(\lambda_i^2). \tag{19}$$

It is apparent that the *nonnoisy* sum S_{CD}^0 can be minimized to exactly zero for the original parameters, i.e., the true parameters

are perfectly attained when there is no noise in measurements. The objective of the OLS minimization is to identify the MLE of these parameters by minimizing the actual *noisy* sum S_{CD}. However, to achieve this both sums need to get their minimum value for the same parameters $\boldsymbol{p}$, i.e., their partial derivatives w.r.t. $\boldsymbol{p}$ have to be zero for the same solution. This does *not* hold true in (19), where the last term involves the slopes λ_i that depend on $\boldsymbol{p}$, which implies that the two sums have different partial derivatives and minima.

Therefore, the CD metric cannot fetch the MLE of the original parameters in the case of multivariate noise. It is worth noting, however, that if the voltage noise is neglected ($\sigma_V = 0, \sigma_n = 0$), the last term in (19) is nullified and the aforementioned condition holds then true. Consequently, the CD metric will yield the MLE of the original parameters if and only if there is noise only on the current samples (zero noise ratio), which is not often the case with measured curves.

B. EDR and NSEDR Performance

Using the linearized PV model, the generalized NSED metric of (13) is approximated by a simple explicit expression

$$S_{\text{NSED}} \approx \sum \frac{[I_m - \lambda_i v_m - \beta_i]^2}{1+\lambda_i^2} \Rightarrow$$

$$\mathbf{E}(S_{\text{NSED}}) = \mathbf{E}(S^0_{\text{NSED}}) + N\mathbf{E}(W_I^2)\mathbf{E}\left(\frac{1}{1+\lambda_i^2}\right) + N\mathbf{E}(w_V^2)\mathbf{E}\left(\frac{\lambda_i^2}{1+\lambda_i^2}\right) \tag{20}$$

following the same derivation steps as for the CD sum (see Appendix B). In the case of NSED, the noise impact has been equalized and $\mathbf{E}(W_I^2) = \mathbf{E}(w_V^2) = \sigma_I^2$ from (12), which makes (20)

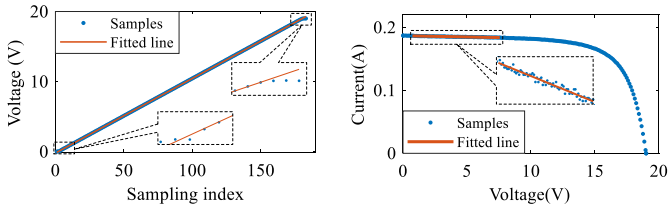
$$\mathbf{E}(S_{\text{NSED}}) \approx \mathbf{E}(S^0_{\text{NSED}}) + N\sigma_I^2\mathbf{E}\left(\frac{1+\lambda_i^2}{1+\lambda_i^2}\right) = \mathbf{E}(S^0_{\text{NSED}}) + N\sigma_I^2. \tag{21}$$

The resulting (21) clearly shows that the expected noisy and nonnoisy sums have the same partial derivatives w.r.t. the parameters (the term $N\sigma_I^2$ is a constant), and, thus, they get minimized at the same point. Therefore, the expected solution found by NSEDR is the same as if there was not any noise, i.e. the NSEDR fetches the MLE of the true parameters.

Furthermore, it is worth noting that if the regular *unscaled* EDR is adopted, the residual sum from (20) becomes

$$\mathbf{E}(S_{ED}) \approx \mathbf{E}(S^0_{ED}) + N\sigma_I^2 + N(\sigma_V^2 - \sigma_I^2)\mathbf{E}\left(\frac{\lambda_i^2}{1+\lambda_i^2}\right). \tag{22}$$

It is apparent that the last term in (22) gets zero when the noise powers in voltage and current are equal ($\sigma_I = \sigma_V$, $\sigma_n = 1$), in which case EDR becomes optimal in the MLE sense. However, even in the general unequal-noise case, a comparison of the last terms in (22) and (19) reveals that the dependence on the parameters is much lower in the ED than in the CD case.



(a) (b)

Fig. 3. Example from the NREL dataset. (a) Voltage with sampling index. (b) I–V curve.

In fact, the main conclusion from this linear analysis is that the proposed NSED metric is the most reliable method to reach to the MLE of the original parameters. If the model scaling step needs to be avoided for whatever reason, the unscaled ED metric (implied $\sigma_n = 1$) should be preferred over the conventional CD metric (implied $\sigma_n = 0$). These theoretical findings are further validated with simulation and experimental results in Section VI.

IV. Noise Extraction Method

Application of the proposed NSED metric of (13) requires knowledge of the noise ratio σ_n. This is an additional input compared with the unscaled ED and CD metrics, which is not generally known and needs to be calculated. When the voltage and current sensors used in the I–V curve recordings are available, it is a straightforward task to measure the noise variables σ_V and σ_I and find $\sigma_n = \sigma_V/\sigma_I$. In the general case, however, the only information available to extract σ_n is the measured (V_m, I_m) dataset. This is managed by the two-step NE method proposed, as explained below.

A. Voltage Noise Estimation

The extraction of the voltage noise σ_V is based on the observation that the voltage is usually sampled linearly during the I–V curve sweep, i.e., the voltage samples are equidistant. Such an example is depicted in Fig. 3(a) referring to an I–V curve from the NREL dataset examined in Section VI. This is the common case nowadays and corresponds to the state-of-the-art *I–V* tracers, although not generally true.

In that case, σ_V can be extracted by fitting a linear model $V = \alpha j + \beta$ to the samples, where j is the sampling index and α, β coefficients, as shown in Fig. 3(a). Because here there is noise only on voltage (the sampling index is a nonnoisy independent variable), this is a linear OLS problem with an exact solution

$$\sigma_V = \frac{\|\boldsymbol{Y} - \boldsymbol{X}\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{Y}\|_2}{N} \tag{23}$$

where $\boldsymbol{Y} = [V_1, \ldots, V_N]^\top$ and $\boldsymbol{X} = \begin{bmatrix} j_1 & \cdots & j_N \\ 1 & \cdots & 1 \end{bmatrix}^\top$ are the input matrices, and N the number of samples.

Application of (23) may require some processing of the input V_m dataset to remove possible outliers and distortions, as commonly done in regression problems. For example, in Fig. 3(a) the very first and last samples do not closely follow the

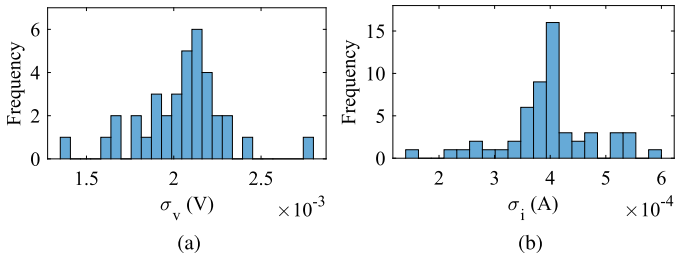


Fig. 4. Histograms of extracted noise variables for the example of Fig. 3. (a) σ_V from 36 different fitting regions. (b) σ_I from 55 fitting regions.

linear-sampling rule. This could be done by explicitly selecting a fitting region, e.g., 10%–90% of the original range.

However, a more automated and reliable approach that copes with any kind of unpredictable distortions is the *median technique* adopted in this article. According to this, the voltage noise σ_V^r is extracted many times for different fitting ranges r (subsets of the entire dataset, possibly overlapping) and then the median value median(σ_V^r) is selected. The premise is that the median will correspond to the undistorted fittings, which will be much more in numbers than the distorted ones. Fig. 4(a) shows the histogram of the extracted σ_V^r values by applying (23) to the example of Fig. 3 for 36 fitting regions (10%–90% of the initial range with 10% resolution). The selected median value of 2.1 mV avoids overestimation from distortions and underestimation from very short fitting regions.

B. Current Noise Estimation

The extraction of the current noise level σ_I is based on the observation that the I–V curve is approximately linear in the short-circuit region [see Fig. 3(b)]. In that sense, fitting a linear model $I = \alpha V + \beta$ on that region using an OLS formulation similar to (23) is a simple solution. However, this is a multivariate noise problem and a total least squares approach should be followed instead. Fortunately, this is a very simple linear regression problem with an exact solution

$$\text{min: } \sum (I_m - \alpha V_m - \beta)^2 - \alpha^2 \sigma_V^2 N \tag{24}$$

$$\sigma_I = \sqrt{\frac{\|\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{p}}\|_2^2 - \|[N\sigma_V \; 0]\hat{\boldsymbol{p}}\|_2^2}{N}} \tag{25}$$

where now $\boldsymbol{Y} = [I_1, \ldots, I_N]^\top$ and $\boldsymbol{X} = \begin{bmatrix} V_1 & \cdots & V_N \\ 1 & \cdots & 1 \end{bmatrix}^\top$ are the inputs, and $\hat{\boldsymbol{p}} = (X^\top X - \begin{bmatrix} N\sigma_V^2 & 0 \\ 0 & 0 \end{bmatrix})^{-1} X^\top Y$ the parameter estimates. Please note that these equations require σ_V as input.

Identifying an appropriate fitting region here is more challenging, since the selected fitting region needs to be "sufficiently linear" apart from undistorted and not too short. A fixed fitting region, e.g., 5%–40% of the voltage range may not always comply with these requirements. This is where the median technique facilitates automatic identification of the fitting region and σ_I value. In the example of Fig. 3(b), applying (25) to 55 regions (0%–50% of voltage range with 5% resolution) yields the histogram of Fig. 4(b), which shows that the median value of 396 μA is a reliable σ_I estimation.

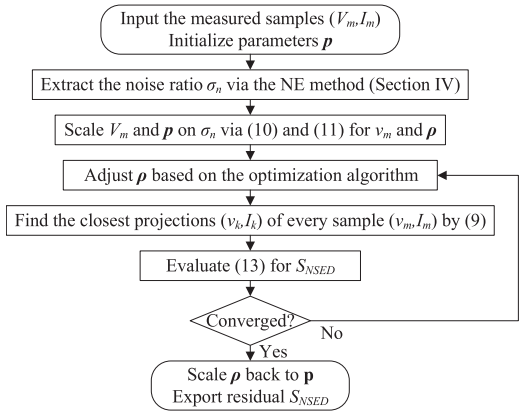


Fig. 5. Flowchart of the proposed NSEDR method.

To summarize, the complete NE method involves first extraction of σ_V through (23) and then identification of σ_I via (25) to get $\sigma_n = \sigma_V/\sigma_I$. The multifitting median approach is optional in case of undistorted (V_m, I_m) datasets. Please note that the NE method is only applicable when the I–V curve is sampled linearly in voltage, i.e., in equidistant voltage points. Extension of this algorithm to uneven voltage spacing is possible if the sampling mechanism is known *a priori* (e.g., quadratic relation between voltage step and index), in which case the aforementioned steps can be modified accordingly.

V. Application of Noise-Scaled Euclidean Distance Regression as an Add-On to Existing Parameter Extraction Methods

The full picture in applying NSEDR is given in the flowchart of Fig. 5. Once the noise ratio is extracted and the model scaling is performed, an optimization algorithm identifies the model parameters by minimizing the NSED. At every iteration, the projections of all samples on the curve need to be updated depending on the current parameters. Upon convergence, the calculated model parameters are scaled back to their original form using a reversed version of (11).

The optimization algorithm is deliberately left generic in Fig. 5. For the results of Section VI, the Levenberg–Marquardt has been employed, but any other numerical or heuristic optimization method will do equally well. The proposed NSEDR does not deal with the problem of global optimality (i.e., multiple local minima), which is the focus of other metaheuristic methods [10]–[14]; it requires an initial $\boldsymbol{p}$ vector that lies sufficiently close to the global optimum to guarantee convergence to the MLE of the true parameters. In that sense, the optimization technique does not matter in this application.

This hints to one possible application of NSEDR as a second-stage step in existing parameter extraction methods toward fine-tuning of their results. In that sense, any parameter extraction method can be enhanced by using its outputs to initialize and apply NSEDR, so that the final values enjoy the benefits of the

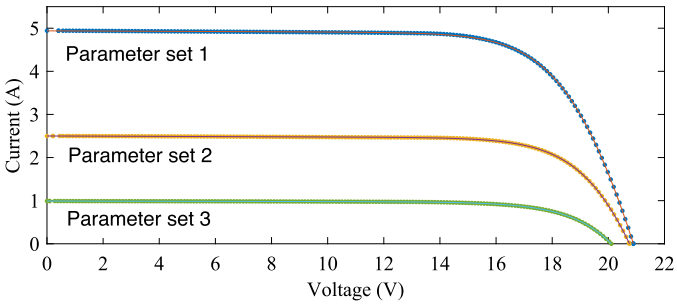


Fig. 6. Sampled and reconstructed I–V curves for the three parameter sets taken from the NREL dataset.

TABLE I
PARAMETER SETS USED AS SYNTHETIC DATA OF THIS SECTION

Set	Irrad. (W/m^2)	Temp. (^{o}C)	I_{ph} (A)	$ln(I_s)$ (ln(A))	a (V)	R_s (Ω)	R_{sh} (Ω)
Set 1	1,003	35.4	4.942	-15.508	1.222	0.246	387.0
Set 2	509	26.8	2.501	-15.997	1.228	0.228	442.2
Set 3	220	21.0	0.991	-14.418	1.398	0.039	844.4

NSED metric. This application, denoted here as *NSEDR add-on*, is fully universal and can be applied to any extracted parameter set, with particular significance when there is considerable noise in both voltage and current samples. See Appendix C for access to the implementation code.

VI. RESULTS AND DISCUSSION

This section assesses the proposed NSED metric and NE method by a series of tests on synthetic data and experimental measurements. First, the NE method is applied to outdoor measurements provided by NREL, then different metrics are compared on synthetic data, and, finally, the NSEDR add-on is tested on five parameter extraction methods from the literature.

The methods employed are: 1) the series-parallel resistance (SPR) method from [5], 2) the oblique asymptote (OA) method from [6], 3) the two-step linear least squares (TSLLS) method from [8], 4) the reduced-form (RF) method form [2], and 5) the co-content (CC) method from [9]. These techniques are indicatively chosen from the enormous pool of numerical and explicit parameter extraction methods only for the purpose of validating the NSEDR add-on. This investigation is by no means a comparison of these methods, which falls out of the scope of this article.

The experimental dataset used in this section was kindly provided by NREL [25] and involves hundreds of thousands of I–V curves from different PV modules at various conditions. This dataset was also employed to produce the synthetic data employed later: Three I–V curves were selected from the Cocoa_mSi460A8 module corresponding to typical high, medium, and low irradiance conditions, whose parameters were then extracted via the TSLLS method (see Fig. 6). The resulting parameter sets 1–3 are given in Table I and are used thereafter whenever synthetic data is required. The saturation current I_s is treated in logarithmic form in Table I and elsewhere in this section to facilitate comparison, as its estimation tends to vary orders of magnitude. Please note that the objective here is to get realistic reference parameters and the extraction method does not really matter; other methods could do equally well in place of TSLLS in this step.

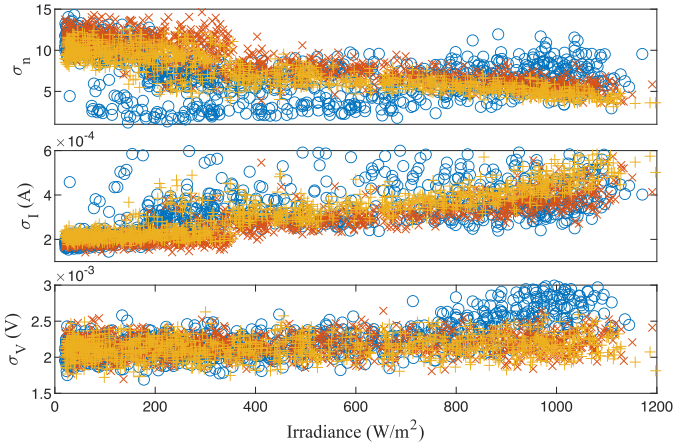


Fig. 7. Extracted noise attributes σ_n, σ_I, and σ_V for three PV modules from the NREL dataset against irradiance variation (Cocoa location: mSi460A8 in ○, mSi0188 in ×, and mSi0166 in +).

TABLE II
NOISE STATISTICS EXTRACTED FROM COCOA_MSI460A8

Noise conditions	σ_n	σ_I (μA)	σ_V (mV)
Low noise ratio	1.91	1247	2.38
Normal noise ratio	7.84	266	2.09
High noise ratio	15.74	142	2.23

A. Validation of the NE Method

The NE method is first applied to three PV modules from the NREL dataset to investigate the noise levels and patterns, and then its accuracy is validated against synthetic data.

1) Noise in NREL Dataset: Three PV modules were selected from the NREL dataset, namely, mSi460A8, mSi0188, and mSi0166 from Cocoa location. The NE method was then applied to 1000 randomly selected curves of these modules to extract all three noise attributes σ_n, σ_I, and σ_V. The results are shown in Fig. 7 plotted against irradiance variation.

Clearly the results for the three modules (colored markers) match very well, which indicates consistency. The noise ratio varies broadly within 3 and 15, exhibiting an increasing trend with falling irradiance. Conversely, the opposite seems to hold true for the current noise, its values ranging mainly from 150 to 500 μA. The voltage noise is rather constant within a 1.5–3-mV range regardless of the conditions.

Table II provides these numbers for three representative conditions (low, normal, and high noise ratio) recorded for Cocoa_mSi460A8, which are used later in the section. The main conclusion from this investigation is that the noise attributes are weather-dependent, rather than constant, and a method to

TABLE III
ESTIMATION ERRORS IN NE

Samples per curve	RMS of the estimation errors (%)		
	σ_n	σ_I	σ_V
100	10.87	9.80	5.99
200	7.21	6.49	3.75
1000	3.67	3.27	1.51

estimate their values is necessary. It is apparent that this dependence has clear patterns and could be possibly modeled mathematically, but this is left for future work.

2) Validation on Synthetic Results: The performance of the proposed NE method can only be evaluated on synthetic data, since the NREL dataset or other similar databases do not provide the actual noise values. To this end, parameter set 1 from Table I is used to produce I–V curves, which are then modified by superimposing artificial noise to emulate measured characteristics. 1000 such curves are produced, featuring normally distributed noise $W_V \sim \mathcal{N}(0, \sigma_V^2)$ and $W_I \sim \mathcal{N}(0, \sigma_I^2)$, but with different voltage and current noise variances at every curve. The noise variances are considered random variables in this experiment, taken uniformly distributed within the ranges seen in the NREL dataset, i.e., $\sigma_V \sim \mathcal{U}(1.5, 3)$mV and $\sigma_I \sim \mathcal{U}(150, 500)\mu$A; this way, the resulting dataset involves a wide range of noise conditions and constitutes a reliable testbench for the assessment of the NE method. In the initial test the I–V curve is sampled in 100 points, and then the experiment is repeated for 200 and 1000 samples per curve resolutions.

Table III shows the rms of the resulting noise estimation errors. The 100 samples/curve resolution leads to moderate errors, which are deemed acceptable given the absence of other noise estimation alternatives. It is interesting to see that as the sampling resolution increases, the errors decrease substantially; this reaffirms that the more data points the better for the NE method, like most statistical techniques. It should be noted that most state-of-the-art *I–V* tracers sample nowadays in the 100–200 range, such as in the NREL dataset. The main conclusion here is that although the NE method could benefit from improvement, it is already trustworthy and reliable enough for application in the NSED metric.

B. Assessment of the NSED Metric

This section evaluates the significance of the NSED metric over the CD and ED alternatives. For this purpose, the respective regressions NSEDR, CDR, and EDR are applied on synthetic data to capture their accuracy in approximating the model parameters. This synthetic dataset is created using the parameter set 1 from Table I to construct an I–V curve with 100 samples, which is then distorted by noise 1000 times to produce an equal number of noise-induced curves. This process is repeated for low, normal, and high noise ratio conditions according to Table II. The known original parameters (set 1) are used to initialize the regressions and calculate the estimation errors, which would not be possible if measured datasets were used.

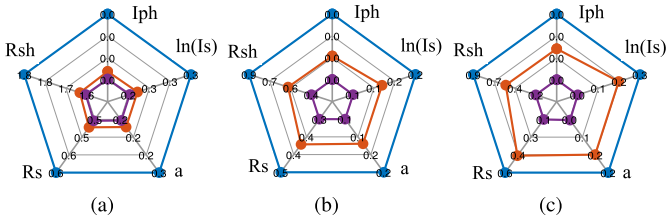


Fig. 8. Spider graphs of the estimation errors in % (CDR in blue, EDR in red, NSEDR in purple) for (a) low, (b) normal, and (c) high noise ratio.

Table IV reports the errors in estimating each of the five parameters by CDR, EDR, and NSEDR, as well as the residual sum after the minimization in each case (bold font indicates lowest value). The Benchmark parameters are found by applying ED regression with perfect knowledge of the noise ratio and initialized from the (known) true parameters; this "idealized NSEDR" is used as a baseline for accuracy and it has no other practical use apart from representing the theoretical upper bound of NSEDR performance. Looking at the parameter estimation errors, NSEDR always yields the lowest values at all noise conditions, which are found only slightly worse than the Benchmark's. This is more clearly illustrated in the spider graphs of Fig. 8 (each axis being the error in a different parameter), which clearly show that NSEDR is superior in estimating all five parameters; the improvement against CDR and EDR is stronger at higher noise ratios, when their inherent assumption of $\sigma_n = 0$ or 1, respectively, is weaker. Between the last two, EDR performs better since assuming equal noise in voltage and current is "more valid" than neglecting the former entirely.

These findings are further justified by comparing the residual sums after the minimization, as calculated using the CD, ED and NSED metrics. As expected, CDR yields the lowest CD sum, EDR the lowest ED sum, and NSEDR the lowest NSED sum, which confirms that all these regressions have done their job perfectly well; it is the objective function employed really that makes the difference. It is also worth noting that the NE-estimated noise ratios in NSEDR are very close to the true values given in the Benchmark in all cases. As indication of the benefit brought by more accurate parameters, the MPP power estimation error is also given in this table, calculated numerically through the standard system of two equations (for details see [31]). The results reaffirm once again the previous observations on the superiority of NSED.

Table IV also shows the computational performance of these regressions in the form of average iterations and mean execution time. The *lsqnonlin* MATLAB function has been used in all implementations with the same convergence tolerances. Clearly, CDR is consistently the most cost-efficient regression owing to its simple formulation. EDR exhibits also consistent performance, but with about double iterations and triple time compared with CDR. The calculation cost of NSEDR is somewhat more stochastic, with higher values at low and normal noise ratios, but lower values at high noise ratio. As expected, NSEDR is more computationally intensive than CDR.

TABLE IV
PERFORMANCE OF DIFFERENT METRICS ON SYNTHETIC DATA

Method	Parameter estimation errors (%)					MPP power error (%)	Mean residual sum			Noise ratio	Computation	
	I_{ph}	$ln(I_s)$	a	R_s	R_{sh}		CD	ED	NSED		Iterat.	Time (ms)
Low noise ratio												
(Benchmark)	(7.112E-3)	(0.225)	(0.201)	(0.499)	(1.621)	(8.93E-03)	(1.7140E-3)	(1.3579E-3)	(1.2151E-3)	(1.91)	(5.3)	(9.0)
CDR	7.687E-3	0.305	0.272	0.642	1.824	11.19E-3	**1.7000E-3**	1.3600E-3	1.2207E-3	0	**2.3**	**3.1**
EDR	7.185E-3	0.237	0.211	0.520	1.644	9.27E-3	1.7057E-3	**1.3566E-3**	1.2159E-3	1	4.6	8.4
NSEDR	**7.118E-3**	**0.227**	**0.202**	**0.503**	**1.625**	**8.94E-3**	1.7149E-3	1.3582E-3	**1.2151E-3**	1.99	5.4	9.2
Normal noise ratio												
(Benchmark)	(1.606E-3)	(0.093)	(0.082)	(0.280)	(0.377)	(3.55E-3)	(1.0929E-3)	(6.7074E-4)	(2.5975E-4)	(7.87)	(11.0)	(16.2)
CDR	3.229E-3	0.240	0.213	0.531	0.927	7.69E-3	**1.0652E-3**	6.6712E-4	2.7056E-4	0	**2.2**	**2.8**
EDR	2.188E-3	0.161	0.142	0.399	0.583	5.18E-3	1.0708E-3	**6.6287E-4**	2.6334E-4	1	4.3	7.5
NSEDR	**1.608E-3**	**0.093**	**0.082**	**0.280**	**0.377**	**3.56E-3**	1.0937E-3	6.7119E-4	**2.5981E-4**	8.00	10.7	15.7
High noise ratio												
(Benchmark)	(7.608E-4)	(0.005)	(1.4E-11)	(0.119)	(0.173)	(3.24E-3)	(1.1581E-3)	(6.8661E-4)	(1.3882E-4)	(15.79)	(1.3)	(3.9)
CDR	3.142E-3	0.254	0.226	0.565	0.940	8.07E-3	**1.1094E-3**	6.7254E-4	1.5344E-4	0	2.2	**2.8**
EDR	1.883E-3	0.169	0.150	0.423	0.542	5.33E-3	1.1154E-3	**6.6780E-4**	1.4351E-4	1	4.3	7.6
NSEDR	**7.629E-4**	**0.006**	**0.003**	**0.119**	**0.174**	**3.25E-3**	1.1578E-3	6.8651E-4	**1.3882E-4**	15.65	**1.4**	4.0

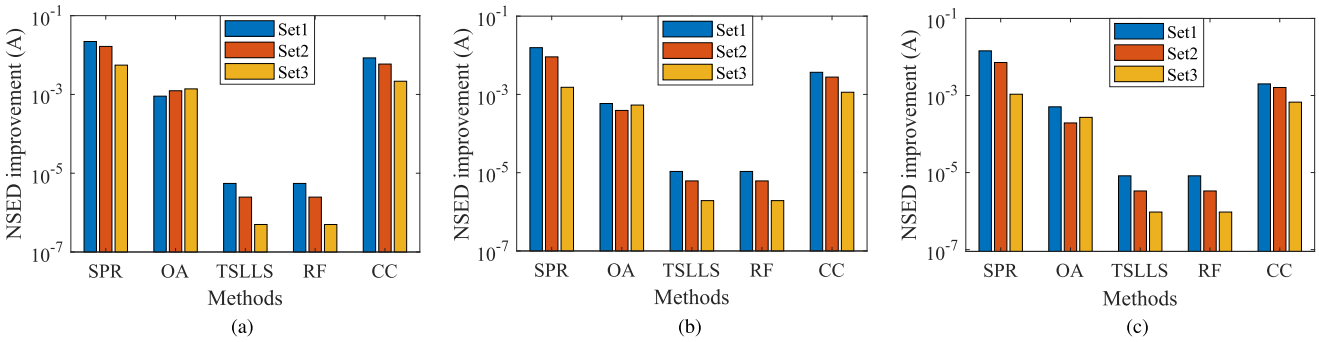


Fig. 9. Barcharts of the NSED improvement brought by enhancing the study-case methods. Average values over 1000 runs for the three parameter sets in logarithmic scale. (a) Low noise ratio. (b) Normal noise ratio. (c) High noise ratio.

This investigation corroborates the theoretical analysis of the previous sections that minimizing the NSED metric leads to more accurate parameters, even with the imperfect noise ratio given by the NE method. If the noise estimation needs to be avoided, the ED metric should be used instead, rather than the conventional CD objective function. These accuracy improvements come at the cost of reasonably higher computational burden.

C. Application of the NSEDR Add-On

This section evaluates the application of NSEDR as a second-stage fine-tuning step to existing parameter extraction methods, first using synthetic data and then the NREL dataset.

1) Validation on Synthetic Data: The five study-case literature methods SPR, OA, TSLLS, RF, and CC are applied to the synthetic dataset of the previous section (set 1, 1000 curves, normal noise ratio), and their results are then optimized via the NSEDR add-on. Table V shows the estimation errors (bold font indicates lowest values), where again Benchmark stands for an "idealized NSEDR" (perfect knowledge of the noise ratio, initialized from the true parameters), used only as an accuracy baseline.

TABLE V
ESTIMATION ERRORS BY FIVE PARAMETER EXTRACTION METHODS WITH AND WITHOUT THE NSEDR ADD-ON (NORMAL NOISE RATIO)

Method	Parameter estimation errors (%)					MPP err. (%)	Mean res. NSED
	I_{ph}	$ln(I_s)$	a	R_s	R_{sh}		
(Bench.)	(1.65E-3)	(0.096)	(0.084)	(0.282)	(0.384)	(3.64E-3)	(2.593E-4)
SPR	6.39E-2	14.702	16.846	38.590	-	2.98E-2	1.591E-2
SPR*	1.69E-3	0.108	0.095	0.318	0.396	3.65E-3	2.595E-4
OA	7.52E-3	1.103	0.926	3.119	4.799	1.40E-2	8.451E-4
OA*	1.73E-3	0.125	0.109	0.363	0.410	3.66E-3	2.599E-4
TSLLS	3.16E-3	0.243	0.216	0.547	0.911	7.39E-3	2.701E-4
TSLLS*	**1.66E-3**	**0.099**	**0.087**	**0.289**	**0.389**	**3.64E-3**	**2.593E-4**
RF	3.16E-3	0.243	0.216	0.547	0.911	7.39E-3	2.701E-4
RF*	**1.66E-3**	**0.099**	**0.087**	**0.289**	**0.389**	**3.64E-3**	**2.593E-4**
CC	2.75E-3	2.002	1.706	0.949	0.745	4.73E-1	3.933E-3
CC*	1.77E-3	0.134	0.117	0.391	0.419	3.65E-3	2.600E-4

* With NSEDR add-on.

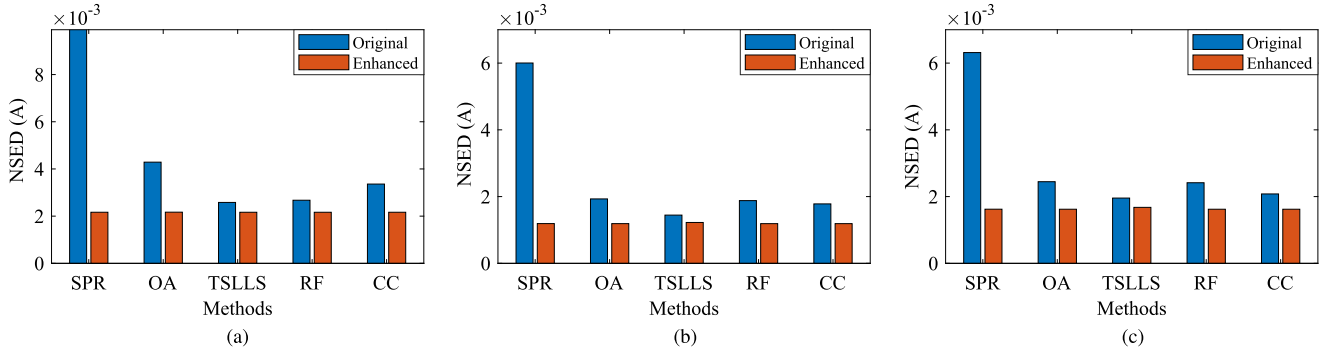


Fig. 10. Barcharts of the average residual NSED using the original and enhanced versions of the study-case methods for 1000 curves of three PV modules from the NREL dataset. (a) mSi460A8. (b) mSi0188. (c) mSi0166.

TABLE VI
OPTIMALITY INDEX FOR THREE NREL PV MODULES

Module	Optimality index (%)									
	SPR	SPR*	OA	OA*	TSLLS	TSLLS*	RF	RF*	CC	CC*
mSi460A8	0	98.3	0	98.0	0	98.4	0	98.4	0	98.3
mSi0188	0	99.7	0	99.9	0.5	96.5	0	100	0	99.9
mSi0166	0	99.9	0	99.9	0.6	96.0	0	99.9	0	100
Time (ms)	0.01	276	0.03	99	3,639	3,755	1,760	1,978	10	85

* With NSEDR add-on

Clearly, the estimation errors are reduced in all parameters when applying the add-on (indicated with *) for every method. Among the enhanced variations (*), the errors are very close to each other and only marginally higher than the Benchmark's values, with the TSLLS* and RF* yielding entirely identical results. Very similar observations are extracted for the resulting MPP power estimation error. This shows that the NSEDR add-on not only improves the accuracy of every method to which it was applied, but leads also to the global optimum if the original method has converged to the global optimality region.

These findings are further validated for the other parameter sets of Table I in Fig. 9. The barcharts illustrate the NSED improvement from enhancing each method (reduction of residual NSED after applying the add-on) for set 1–3 at the three noise ratios. Apparently, the NSED improvement varies with the parameter set and noise, while some methods are boosted more than others, but the improvement is always positive, i.e., all methods are benefited by the add-on at least by a little.

2) Validation on the NREL Dataset: The study-case methods are applied to 1000 randomly selected curves from the three NREL modules, after removing a few distorted instances. Since the true model parameters are not available here, NSED serves as the accuracy indicator, i.e., the lower the better. Fig. 10 illustrates the results in the form of bar charts, showing the residual NSED at the three modules for every method before and after the enhancement. It is apparent that all methods get boosted by the add-on, yielding afterward all very similar NSED (red bars) for the same module.

These findings are also given in Table VI using the *optimality index*, defined as the frequency of the resulting NSED being the lowest among all methods within a reasonable tolerance (i.e., 95% means that this method yields the lowest NSED with a 1E-6 tolerance in 95% of the cases). What Table VI really shows is that, once enhanced by the NSEDR add-on, all five methods converge to the same solution 96%–100% of the times, which is strong indication of consistency. The small differences in the optimality index relate to cases where the original method gets trapped to a local optimum.

The last row in Table VI provides the average execution time of each method. Expectedly, the time of the original methods varies greatly, with the explicit SPR and OA being the fastest by far. The enhanced versions (*) are burdened by tens to hundreds of additional milliseconds, which may be a major or minor extra cost depending on the application. Once enhanced, simple methods like SPR and OA cannot be considered explicit anymore, but they yield accuracy very similar to more complex methods at a lower total execution time.

In fact, the main conclusion here is that the proposed NSEDR add-on works fine with measured data, and that simple parameter extraction methods with this follow-up step may perform outstandingly regardless of their simplifications. A more thorough investigation and comparison of literature methods falls out of the scope of this article.

VII. CONCLUSION

The theoretical analysis and experimental validation of this article show that the proposed NSED metric is superior to the commonly used CD or ED alternatives in fetching the MLE of the PV model parameters at multivariate noise. To get the noise ratio required in NSED, the proposed NE method is easy to apply and reliable for I–V curves with linear voltage sampling. When the noise ratio cannot be obtained, the ED metric should be preferred over the default CD option. The additional complexity brought by ED and NSED is reflected in the execution time. An application of the new metric has been introduced as a fine-tuning step in other parameter extraction methods, which is demonstrated to always improve their accuracy by at least a little.

This new metric has many applications in the field of PV modeling. Apart from the NSEDR add-on step examined in this article, the NSED metric can be directly integrated into other parameter extraction methods, by replacing the metric used; this could potentially lead to improved accuracy and lower complexity compared with the former two-stage process, an investigation left for future work. Future directions also include extension of the NSED metric to more sophisticated PV models (e.g., double-diode model) and integration to parameter extraction methods with multiple sampled I–V curves at different operating conditions. Furthermore, the noise attributes show dependence on irradiance and temperature with clear patterns, which may be worth investigating and formulating. On a relevant note, the NE method would benefit from improvement, especially in I–V curves with a few data points and nonlinear voltage sampling.

APPENDIX

A. Derivation Steps of the CD Sum

In the CD sum, the term A from (17) can be written as

$$A = 2\sum W_I \underbrace{[I_i - \lambda_i V_i - \beta_i]}_{\Lambda_1} - 2\sum W_V \underbrace{\lambda_i [I_i - \lambda_i V_i - \beta_i]}_{\Lambda_2}$$

$$\Rightarrow \mathbf{E}(A) = 2N\mathbf{E}(W_I\Lambda_1) - 2N\mathbf{E}(W_V\Lambda_2) \quad (26)$$

where Λ_1 and Λ_2 are auxiliary variables that depend only on the original sample i; N is the number of samples; and $\mathbf{E}(.)$ denotes the expectation operator. The assumption that noise and sample are uncorrelated implies that W_I and Λ_1 are uncorrelated, as well as W_V and Λ_2, thus $\mathbf{E}(W_I\Lambda_1) = \mathbf{E}(W_I)\mathbf{E}(\Lambda_1)$ and $\mathbf{E}(W_V\Lambda_2) = \mathbf{E}(W_V)\mathbf{E}(\Lambda_2)$. Accounting also for the zero mean value of the noise variables $\mathbf{E}(W_I) = \mathbf{E}(W_V) = 0$ yields that $\mathbf{E}(A)$ gets effectively zero.

Similar steps for the B term from (18) lead to

$$B = \sum W_I^2 - 2\sum \lambda_i W_I W_V + \sum \lambda_i^2 W_V^2 \Rightarrow$$

$$\mathbf{E}(B) = N\mathbf{E}(W_I^2) - 2N\mathbf{E}(\lambda_i W_I W_V) + N\mathbf{E}(\lambda_i^2 W_V^2)$$

$$= N\mathbf{E}(W_I^2) - 2N\mathbf{E}(\lambda_i)\mathbf{E}(W_I)\mathbf{E}(W_V) + N\mathbf{E}(\lambda_i^2)\mathbf{E}(W_V^2)$$

$$= N\sigma_I^2 - 0 + N\sigma_V^2\mathbf{E}(\lambda_i^2). \quad (27)$$

B. Derivation Steps of the NSED Sum

Following a similar methodology to CD, the NSED sum from (20) can be expressed as

$$S_{\text{NSED}} \approx \sum \frac{[I_m - \lambda_i v_m - \beta_i]^2}{1+\lambda_i^2} = S^0_{\text{NSED}} + C + D \quad (28)$$

where

$$S^0_{\text{NSED}} = \sum \frac{[I_i - \lambda_i v_i - \beta_i]^2}{1+\lambda_i^2} \quad (29)$$

$$C = 2\sum \frac{(I_i - \lambda_i v_i - \beta_i)(W_I - \lambda_i w_V)}{1+\lambda_i^2} \quad (30)$$

$$D = \sum \frac{(W_I - \lambda_i w_V)^2}{1+\lambda_i^2}. \quad (31)$$

Because of the zero mean value of the noise variables and the uncorrelation between samples and noise, $\mathbf{E}(C) = 0$ for the same reasons that $\mathbf{E}(A) = 0$, and $\mathbf{E}(D)$ becomes

$$\mathbf{E}(D) = N\mathbf{E}(W_I^2)\mathbf{E}\left(\frac{1}{1+\lambda_i^2}\right) + N\mathbf{E}(w_V^2)\mathbf{E}\left(\frac{\lambda_i^2}{1+\lambda_i^2}\right) \quad (32)$$

following the same steps as for $\mathbf{E}(B)$.

C. Implementation Code

The implemented algorithms for the NE and NSEDR functions are freely.[1] For any inquiries, please contact Dr. Efstratios Batzelis.[2]

REFERENCES

[1] M. B. Campanelli and B. H. Hamadani, "Calibration of a single-diode performance model without a short-circuit temperature coefficient," *Energy Sci. Eng.*, vol. 6, no. 4, pp. 222–238, 2018.

[2] A. Laudani, F. Mancilla-David, F. Riganti-Fulginei, and A. Salvini, "Reduced-form of the photovoltaic five-parameter model for efficient computation of parameters," *Sol. Energy*, vol. 97, pp. 122–127, Nov. 2013.

[3] M. Villalva, J. Gazoli, and E. Filho, "Comprehensive approach to modeling and simulation of photovoltaic arrays," *IEEE Trans. Power Electron.*, vol. 24, no. 5, pp. 1198–1208, May 2009.

[4] E. I. Batzelis and S. A. Papathanassiou, "A method for the analytical extraction of the single-diode PV model parameters," *IEEE Trans. Sustain. Energy*, vol. 7, no. 2, pp. 504–512, Apr. 2016.

[5] S. Cannizzaro, M. C. Di Piazza, M. Luna, and G. Vitale, "PVID: An interactive Matlab application for parameter identification of complete and simplified single-diode PV models," in *Proc. IEEE 15th Work. Control Model. Power Electron.*, 2014, pp. 1–7.

[6] F. Toledo and J. M. Blanes, "Geometric properties of the single-diode photovoltaic model and a new very simple method for parameters extraction," *Renewable Energy*, vol. 72, pp. 125–133, Dec. 2014.

[7] S. Li, W. Gong, and Q. Gu, "A comprehensive survey on meta-heuristic algorithms for parameter extraction of photovoltaic models," *Renewable Sustain. Energy Rev.*, vol. 141, May 2021, Art. no. 110828.

[8] F. J. Toledo, J. M. Blanes, and V. Galiano, "Two-step linear least-squares method for photovoltaic single-diode model parameters extraction," *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6301–6308, Aug. 2018.

[9] A. Ortiz-Conde, F. 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.

[10] R. C. M. Gomes, M. A. Vitorino, M. B. D. R. Corrêa, D. A. Fernandes, and R. Wang, "Shuffled complex evolution on photovoltaic parameter extraction: A comparative analysis," *IEEE Trans. Sustain. Energy*, vol. 8, no. 2, pp. 805–815, Apr. 2017.

[11] T. Sudhakar Babu, K. Priya, N. Rajasekar, and K. Balasubramanian, "An innovative method for solar PV parameter extraction for double diode model," in *Proc. Annu. IEEE India Conf.*, New Delhi, India, 2016, pp. 1–6.

[12] B. Subudhi and R. Pradhan, "Bacterial foraging optimization approach to parameter extraction of a photovoltaic module," *IEEE Trans. Sustain. Energy*, vol. 9, no. 1, pp. 381–389, Jan. 2018.

[13] M. F. AlHajri, K. M. El-Naggar, M. R. AlRashidi, and A. K. Al-Othman, "Optimal extraction of solar cell parameters using pattern search," *Renewable Energy*, vol. 44, pp. 238–245, Aug. 2012.

[14] J. Su, Y. Zhang, C. Zhang, T. Gu, and M. Yang, "Parameter extraction of photovoltaic single-diode model using integrated current-voltage error criterion," *J. Renewable Sustain. Energy*, vol. 12, no. 4, Aug. 2020, Art. no. 043704.

[15] D. S. H. Chan, J. R. Phillips, and J. C. H. Phang, "A comparative study of extraction methods for solar cell model parameters," *Solid State Electron.*, vol. 29, no. 3, pp. 329–337, Mar. 1986.

[1] [Online]. Available: https://pvmodel.umh.es/

[2] [Online]. Available: e.batzelis@soton.ac.uk

[16] W. A. Fuller, *Measurement Error Models*. New York, NY, USA: Wiley, 2006.
[17] A. R. Burgers, J. A. Eikelboom, A. Schonecker, and W. C. Sinke, "Improved treatment of the strongly varying slope in fitting solar cell I-V curves," in *Proc. Conf. Rec. IEEE Photovolt. Specialists Conf.*, 1996, pp. 569–572.
[18] A. Ghoneim *et al.*, "Analysis of performance parameters of amorphous photovoltaic modules under different environmental conditions," *Energy Sci. Technol.*, vol. 2, no. 1, pp. 43–50, Aug. 2011.
[19] F. J. Toledo and J. M. Blanes, "Analytical and quasi-explicit four arbitrary point method for extraction of solar cell single-diode model parameters," *Renew. Energy*, vol. 92, pp. 346–356, Jul. 2016.
[20] I. Markovsky and S. Van Huffel, "Overview of total least-squares methods," *Signal Process.*, vol. 87, no. 10, pp. 2283–2302, 2007.
[21] R. J. Carroll and D. Ruppert, "The use and misuse of orthogonal regression in linear errors-in-variables models," *Amer. Stat. Assoc.*, vol. 50, no. 1, pp. 1–6, 1996.
[22] P. T. Boggs and J. E. Rogers, "Orthogonal distance regression," Center Comput. Appl. Math., US Dept. Commerce, Washington, DC, USA, NISTIR 89–4197, 1990.
[23] P. T. Boggs, R. H. Byrd, J. E. Rogers, and R. B. Schnabel, "User's reference guide for ODRPACK version 2.01 software for weighted orthogonal distance regression," Nat. Inst. Standards Technol., US Dept. Commerce, Gaithersburg, MD, USA, NISTIR 4834, 1992.
[24] M. Diantoro *et al.*, "Shockley's equation fit analyses for solar cell parameters from I-V curves," *Int. J. Photoenergy*, vol. 2018, Apr. 2018, Art. no. 9214820.
[25] B. Marion *et al.*, "New data set for validating PV module performance models," in *Proc IEEE 40th Photovolt. Specialist Conf. Denver*, 2014, pp. 1362–1366.
[26] E. I. Batzelis, I. A. Routsolias, and S. A. Papathanassiou, "An explicit PV string model based on the lambert W function and simplified MPP expressions for operation under partial shading," *IEEE Trans. Sustain. Energy*, vol. 5, no. 1, pp. 301–312, Jan. 2014.
[27] Y. Mahmoud and E. F. El-Saadany, "Fast power-peaks estimator for partially shaded PV systems," *IEEE Trans. Energy Convers.*, vol. 31, no. 1, pp. 206–217, Mar. 2016.
[28] MATLAB, "W Lambert function in MATLAB," [Online]. Available: https://uk.mathworks.com/help/symbolic/lambertw.html
[29] C. Hansen, "Lambert W implementation in scipy." 2021. [Online]. Available: https://github.com/pvlib/pvlib-python/blob/v0.9.0/pvlib/ivtools/sdm.py#L1211
[30] E. I. Batzelis, G. Anagnostou, C. Chakraborty, and B. C. Pal, "Computation of the lambert W function in photovoltaic modeling," in *Proc. ELECTRIMACS*, 2020, pp. 583–595.
[31] E. I. Batzelis, G. E. Kampitsis, S. A. Papathanassiou, and S. N. Manias, "Direct MPP calculation in terms of the single-diode PV model parameters," *IEEE Trans. Energy Convers.*, vol. 30, no. 1, pp. 226–236, Mar. 2015.