Quantifying and Reducing Curve-Fitting Uncertainty in I_{SC}

Preprint

Mark Campanelli and Keith Emery
National Renewable Energy Laboratory

Benjamin Duck
CSIRO Energy Flagship

Presented at the 42nd IEEE Photovoltaic Specialists Conference
New Orleans, Louisiana
June 14–19, 2015

© 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

M. Campanelli, B. Duck and K. Emery, "Quantifying and reducing curve-fitting uncertainty in Isc," Photovoltaic Specialist Conference (PVSC), 2015 IEEE 42nd, New Orleans, LA, 2015, pp. 1-6. doi: 10.1109/PVSC.2015.7356127

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Conference Paper
NREL/CP-5J00-63602
September 2015

Contract No. DE-AC36-08GO28308

NOTICE

The submitted manuscript has been offered by an employee of the Alliance for Sustainable Energy, LLC (Alliance), a contractor of the US Government under Contract No. DE-AC36-08GO28308. Accordingly, the US Government and Alliance retain a nonexclusive royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for US Government purposes.

This report was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or any agency thereof.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Available electronically at SciTech Connect http:/www.osti.gov/scitech

Available for a processing fee to U.S. Department of Energy and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
OSTI http://www.osti.gov
Phone: 865.576.8401
Fax: 865.576.5728
Email: reports@osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312
NTIS http://www.ntis.gov
Phone: 800.553.6847 or 703.605.6000
Fax: 703.605.6900
Email: orders@ntis.gov

Cover Photos by Dennis Schroeder: (left to right) NREL 26173, NREL 18302, NREL 19758, NREL 29642, NREL 19795.

NREL prints on paper that contains recycled content.

Quantifying and Reducing Curve-Fitting Uncertainty in I_{sc}

Mark Campanelli*, Benjamin Duck†, Keith Emery*

*National Renewable Energy Laboratory, Golden, CO, 80401, USA
†CSIRO Energy Flagship, Mayfield West NSW, 2304, Australia

***Abstract*—Current-voltage (I-V) curve measurements of photovoltaic (PV) devices are used to determine performance parameters and to establish traceable calibration chains. Measurement standards specify localized curve fitting methods, e.g., straight-line interpolation/extrapolation of the I-V curve points near short-circuit current, I_{sc}. By considering such fits as statistical linear regressions, uncertainties in the performance parameters are readily quantified. However, the legitimacy of such a computed uncertainty requires that the model be a valid (local) representation of the I-V curve and that the noise be sufficiently well characterized. Using more data points often has the advantage of lowering the uncertainty. However, more data points can make the uncertainty in the fit arbitrarily small, and this fit uncertainty misses the dominant residual uncertainty due to so-called model discrepancy. Using objective Bayesian linear regression for straight-line fits for I_{sc}, we investigate an evidence-based method to automatically choose data windows of I-V points with reduced model discrepancy. We also investigate noise effects. Uncertainties, aligned with the *Guide to the Expression of Uncertainty in Measurement* (GUM), are quantified throughout.**

***Index Terms*—Bayesian inference, data window selection, evidence, linear regression, measurement uncertainty analysis, model discrepancy, noise model, uncertainty quantification.**

I. Introduction

Short-circuit current, denoted as I_{sc}, is a key performance parameter for photovoltaic (PV) devices and is also commonly used in PV measurement calibration chains. This practice imparts importance to the proper quantification of uncertainty in I_{sc}, which is often inferred from a measured current-voltage (I-V) curve. Measurement standards (e.g., [1]) suggest straight-line fitting methods for determining I_{sc}. Furthermore, the slope of this fitted line is often used to assess shunt resistance issues.

The problem of model discrepancy—e.g., a statistical model with an inadequate functional relationship or poor noise representation—can invalidate an uncertainty computation based on a statistical linear regression [2]. Even with significant model discrepancy, more data points can make the fit uncertainty arbitrarily small, in which case the fit uncertainty misses the dominant residual uncertainty due to model discrepancy [2]. In the metrology field, this is often called *dark uncertainty* [3]. Expanding upon earlier work [4], we investigate a method for reducing model discrepancy and the associated dark uncertainty. Here, we do not consider all possible uncertainty sources, only those due to fitting a straight line to data with noise in the ordinate.[1]

[1]We distinguish the more specific term *straight-line fit* from the more general term *linear regression*. The latter refers to any model that is linear with respect to the fitting parameters, such as the coefficients in a polynomial.

The uncertainty analysis we develop aligns with the *Guide to the Expression of Uncertainty in Measurement* (GUM) [5]. In the probabilistic GUM framework, a state-of-knowledge distribution (SoKD) is the fundamental description of an uncertain quantity of interest [6]. A quantity of interest is thus viewed as a random variable that typically is described fully by a probability density function (PDF), with a summary given by its expected value (or mean), called the *measured value*, and its standard deviation, called the *standard uncertainty*. If these summary values do not exist for the distribution, then alternatives such as a mode and a 95% coverage interval can summarize the SoKD.

There can be numerous benefits of using local linear models instead of global non-linear physical models. For example, considerable cell mismatch in a PV module may suggest using a local straight-line fit at I_{sc} instead of a non-linear single- or double-diode model, which may poorly represent the entire I-V curve in this situation. In general, linear models allow computationally straightforward and efficient regressions to data (e.g., using linear algebra to compute a least-squares fit). Herein, we work within the framework of *objective Bayesian linear regression* (o-BLR), which yields SoKDs with analytic forms that do not require a sampling algorithm to compute and for which the so-called model evidence is readily computable.

II. Evidence-Based Fitting for I_{sc}

In this section, we describe an automated computational workflow that aims to reduce model discrepancy in a straight-line fit of an I-V curve at I_{sc}. The purpose is to estimate I_{sc} with reliably quantified uncertainty. This uncertainty is reduced by attempting to take a maximal window of I-V data points about I_{sc} within the limits of validity of a local straight-line model to the I-V curve at I_{sc}. Because the valid quantification of uncertainty requires a sufficiently valid model, we attempt to reduce the dark uncertainty from model discrepancy by maximizing the model *evidence*, which is also known as the *marginal likelihood* [7].

The evidence is readily computed in the context of o-BLR, and o-BLR also naturally provides SoKDs that quantify the uncertainty in the model parameters. The case for I_{sc} is particularly straightforward because the intercept of the straight-line model is I_{sc}. After reviewing the relevant theory of o-BLR, we discuss how the window of I-V data points taken near I_{sc} is selected so as to maximize the evidence. In general, larger data windows reduce the computed uncertainty in I_{sc}, but the evidence is used to avoid including data points for which a straight-line model is clearly inadequate. Lastly,

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

examples show the workflow applied to different I-V curves with particular characteristics that affect window selection.

A. Objective Bayesian Linear Regression (o-BLR)

o-BLR shares many common mathematical features with linear least squares (LLS) and maximum likelihood (ML) estimation. However, one should recognize that the fundamental result of o-BLR is a *posterior* joint SoKD for the model parameters. In this case, the *straight-line parameters* are the intercept a_0 and slope a_1 of the statistical model relating current to voltage, i.e.,

$$I_k = a_0 + a_1 V_k + \epsilon_{I_k}(\sigma^2), \quad k = 1, \ldots, K, \tag{1}$$

where K is the total number of I-V curve data pairs (V_k, I_k) used to estimate the straight-line parameters. In particular, the I-intercept a_0 is the short-circuit current I_{sc} (where $V = 0$).

The *statistical parameter* σ^2 in (1) specifies the variance for the distribution from which the additive noise ϵ_{I_k} is realized in the kth observed current value I_k. In this work, we also take σ^2 as unknown and to be inferred from the I-V data, which reflects the state of affairs for many I-V measurement systems that are under statistical process control but that are measuring a significant variety of PV devices. In the absence of system-specific information, this noise is assumed to be independent and identically distributed (i.i.d.) between data points according to a zero-mean normal distribution, i.e.,

$$\widetilde{\epsilon}_{I_k}(\sigma^2) \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\mu_{\widetilde{\epsilon}_{I_k}} = 0, \Sigma_{\widetilde{\epsilon}_{I_k}} = \sigma^2\right), \quad k = 1, \ldots, K,$$

where $\mu_{\widetilde{\epsilon}_{I_k}}$ denotes the mean and $\Sigma_{\widetilde{\epsilon}_{I_k}}$ denotes the variance-covariance matrix.[2] (Here, both are scalars.) This assumption is typically reasonable for measurements using Xenon-arc-lamp based solar simulation in which an imperfect light-level correction is applied to each current value [8]. The error in the voltage data (potentially non-random) is assumed to be negligible by comparison, which for many PV devices is a reasonable assumption sufficiently far away from V_{oc} and when the light-level corrections are sufficiently small [9].

For notational purposes, model (1) is rewritten in matrix-vector form. Specifically,

$$\mathbf{I} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}_{\mathbf{I}}, \quad \widetilde{\boldsymbol{\epsilon}}_{\mathbf{I}} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\widetilde{\boldsymbol{\epsilon}}_{\mathbf{I}}} = \mathbf{0}, \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\epsilon}}_{\mathbf{I}}} = \sigma^2 \boldsymbol{\mathcal{I}}_K\right), \tag{2}$$

where $\mathbf{I} = (I_1, \ldots, I_K)^{\mathrm{T}}$ is the vector of current data points, $\mathbf{V} = (V_1, \ldots, V_K)^{\mathrm{T}}$ is the vector of voltage data points,

$$\mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{V} \end{bmatrix} = \begin{bmatrix} 1 & V_1 \\ \vdots & \vdots \\ 1 & V_K \end{bmatrix}$$

is the so-called *design matrix*, $\boldsymbol{\beta} = (a_0, a_1)^{\mathrm{T}}$ is the vector of straight-line parameters, and $\boldsymbol{\mathcal{I}}_K$ is the K-dimensional identity matrix used to specify the covariance matrix $\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\epsilon}}_{\mathbf{I}}}$ under the i.i.d. assumption. We denote a vector of realized model parameter values as $\boldsymbol{\theta} = (a_0, a_1, \sigma^2)^{\mathrm{T}} = (\boldsymbol{\beta}^{\mathrm{T}}, \sigma^2)^{\mathrm{T}}$, with corresponding random vector $\widetilde{\boldsymbol{\theta}} = (\widetilde{a}_0, \widetilde{a}_1, \widetilde{\sigma}^2)^{\mathrm{T}} = (\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}, \widetilde{\sigma}^2)^{\mathrm{T}}$.[3]

[2]We distinguish random variables/vectors by using tilde notation $\widetilde{\cdot}$.

[3]The transpose operator $^{\mathrm{T}}$ is often omitted for better readability.

As a point of reference, we recall the standard LLS (equivalently, ML) estimate of $\boldsymbol{\beta}$ given the data, namely,

$$\widehat{\boldsymbol{\beta}} := (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\,\mathbf{I}. \tag{3}$$

In addition, the degrees-of-freedom (DoF) adjusted residual sum of squares is an estimate of σ^2 in model (2), namely,

$$\widehat{\sigma}^2 := (\mathbf{I} - \mathbf{X}\,\widehat{\boldsymbol{\beta}})^{\mathrm{T}}(\mathbf{I} - \mathbf{X}\,\widehat{\boldsymbol{\beta}})/\nu, \tag{4}$$

where $\nu := K - P$ is the DoF and P is the number of model parameters to estimate, excluding σ^2. Clearly, the number of data points K must exceed the number of straight-line parameters $P = 2$, so that $\nu = K - P \geq 1$. In addition, the matrix inverse in definition (3) requires that $\mathbf{X}$ have full rank, which happens when at least two I-V data points have distinct voltages. $\widehat{\boldsymbol{\theta}} = (\widehat{a}_0, \widehat{a}_1, \widehat{\sigma}^2) = (\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2)$ denotes the combined model parameters estimate.

For model (2), we now perform the standard o-BLR with non-informative, improper prior SoKD for the parameters [7], namely,

$$\pi_{\widetilde{\boldsymbol{\theta}}}(\boldsymbol{\theta}) = 1/\sigma^2. \tag{5}$$

Conditional on the current data $\mathbf{I}$ taken at voltages $\mathbf{V}$, the PDF for the posterior joint SoKD is given by the following factorized and analytical form

$$f_{\widetilde{\boldsymbol{\theta}}|\mathbf{I},\mathbf{V}}(\boldsymbol{\theta}) = f_{\widetilde{\boldsymbol{\beta}}|\sigma^2,\mathbf{I},\mathbf{V}}(\boldsymbol{\beta})\, f_{\widetilde{\sigma}|\mathbf{I},\mathbf{V}}(\sigma^2), \tag{6}$$

with posterior joint normal distribution for the straight-line parameters given conditionally on $\widetilde{\sigma}^2 = \sigma^2$ and the data by

$$\widetilde{\boldsymbol{\beta}}|\sigma^2, \mathbf{I}, \mathbf{V} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\widetilde{\boldsymbol{\beta}}|\sigma^2,\mathbf{I},\mathbf{V}} = \widehat{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\beta}}|\sigma^2,\mathbf{I},\mathbf{V}} = \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\right),$$

and marginal posterior distribution for the noise variance given by

$$\widetilde{\sigma}^2 \sim \mathcal{IG}\left(\alpha_1 = \nu/2, \alpha_2 = \widehat{\sigma}^2\,\nu/2\right),$$

where $\mathcal{IG}$ specifies the inverse-gamma distribution with shape parameter α_1 and scale parameter α_2.

One can show that marginalizing out $\widetilde{\sigma}^2$ gives a shifted and scaled multivariate t-distribution for $\widetilde{\boldsymbol{\beta}}$ with ν DoF [7], [10], namely,

$$\widetilde{\boldsymbol{\beta}} \sim t_\nu\left(\boldsymbol{\mu}_{\widetilde{\boldsymbol{\beta}}} = \widehat{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\beta}}} = \widehat{\sigma}^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\right),$$

with scale matrix $\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\beta}}}$.[4] Further marginalizing out $\widetilde{a}_1$ gives a shifted and scaled univariate t-distribution for $\widetilde{a}_0$ with ν DoF [7], [10], namely,

$$\widetilde{a}_0 \sim t_\nu\left(\mu_{\widetilde{a}_0} = \widehat{\beta}_1, \Sigma_{\widetilde{a}_0} = \left[\widehat{\sigma}^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\right]_{1,1}\right), \tag{7}$$

where the $[\cdot]_{1,1}$ notation denotes the $1,1$ entry in $\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\beta}}}$ [10]. The slope parameter $\widetilde{a}_1$ has an analogous marginal posterior distribution. In general, parameters $\widetilde{\boldsymbol{\beta}} = (\widehat{a}_0, \widehat{a}_1)$ are jointly distributed in their posterior SoKD and thus not independent.

Importantly, the marginal distribution for $\widetilde{a}_0$ is a GUM-compatible SoKD for I_{sc} that provides an estimate for, and quantifies the uncertainty of, I_{sc}. If the DoF $\nu = K - P \geq 2$,

[4]$\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\beta}}}$ is *not* the covariance matrix for the straight-line parameters [10].

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

then the expected value of this SoKD is well-defined as $\mu_{\widetilde{a}_0}$. Otherwise, if $\nu = K - P = 1$, then $\mu_{\widetilde{a}_0}$ is still the well-defined mode of the SoKD. In either case, $\mu_{\widetilde{a}_0}$ can be taken as the measured value for I_{sc}. If $\nu = K - P \geq 3$, then the standard uncertainty for I_{sc} is well-defined and given by $\sigma_{\widetilde{a}_0} = \sqrt{\frac{\nu}{\nu-2} \Sigma_{\widetilde{a}_0}}$. Thus, at least five data points ($K = 5$) are needed for the straight-line model ($P = 2$) in order for the standard uncertainty to be well defined. Otherwise, if $\nu = K - P = 1$ or $\nu = K - P = 2$, then $I_{95} = [I_{95,\text{min}}, I_{95,\text{max}}]$, the 95% coverage interval (centered about $\mu_{\widetilde{a}_0}$ for the symmetric shifted and scaled t-distribution), can still be computed as the expanded uncertainty from the SoKD.[5] In all cases we take the relative expanded uncertainty to be

$$U_{95} = \frac{(I_{95,\text{max}} - I_{95,\text{min}})/2}{\mu_{\widetilde{a}_0}} = \frac{I_{95,\text{max}} - I_{95,\text{min}}}{I_{95,\text{max}} + I_{95,\text{min}}}.$$

The SoKD for $\widetilde{a}_0$ is a shifted and scaled univariate t-distribution with ν DoF and with positive scale parameter $\Sigma_{\widetilde{a}_0} = \widehat{\sigma}^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}{}_{1,1}$.[6] A reduction in the magnitude of this scale parameter gives a reduction in the uncertainty in I_{sc}. Recalling (4), the magnitude of $\widehat{\sigma}^2$ depends directly on the sum of squared residuals $(\mathbf{I} - \mathbf{X}\widehat{\boldsymbol{\beta}})^{\mathrm{T}}(\mathbf{I} - \mathbf{X}\widehat{\boldsymbol{\beta}})$ and inversely on the DoF ν. Because the factor ν appears in the denominator of (4), a reduction in uncertainty can result from increasing the number of data points, as long as the sum of squares of the residuals and the 1,1-component of the inverse design matrix are bounded or do not grow too quickly with additional data points. As a secondary effect, ν also acts as a shape parameter, so that increasing the number of data points reduces uncertainty by reducing the heavy tails of the t-distribution to the limiting tails of a Gaussian distribution as $\nu \to \infty$.

In the examples presented here, the $1,1$-component of the inverse design matrix strictly decreases with additional data points from larger data windows, so that an increase in the sum of squared residuals must grow sufficiently quickly with increasing data windows to cause larger uncertainty in $\widetilde{a}_0$. In some cases, the sum of squared residuals could grow sufficiently slowly with additional data so that the uncertainty is reduced with more data, *regardless of the validity of the straight-line fit* [2]. Taking all these considerations into account for a given measured I-V curve, it may be advantageous to take larger data windows with more points in order to reduce the parameter uncertainty in the fit, as long as the straight-line model remains valid over the larger window.

Finally, we note that these o-BLR results readily extend to polynomial regressions of arbitrarily higher degree, with the caveat that the choice of polynomial basis can become quite important for the stability of numerical computations.

B. Evidence-Based Data Windowing

The previous results for o-BLM derive from the application of Bayes' theorem using an improper prior SoKD. Succinctly and intuitively, Bayes' theorem transforms the prior SoKD of the model parameters into the posterior SoKD of the parameters by a probabilistic "re-weighting" according to the *likelihood* function. The likelihood is derived from model (2) and quantifies the likelihood of a particular parameter being the true value, given a particular realization of the data.

The likelihood function is determined by considering the probability (density) of observing current data $\mathbf{I}$ at voltages $\mathbf{V}$, given any valid choice for the parameter $\boldsymbol{\theta}$ in the model parameter space $\boldsymbol{\Theta}$. For model (2), we have

$$L(\boldsymbol{\theta}; \mathbf{I}, \mathbf{V}) = f_{\tilde{\mathbf{I}}|\boldsymbol{\theta},\mathbf{V}}(\mathbf{I}). \tag{8}$$

where $f_{\tilde{\mathbf{I}}|\boldsymbol{\theta},\mathbf{V}}$ is the PDF for the joint normal distribution of the current data $\mathbf{I}$ conditional on $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta}$ and on $\mathbf{V}$, namely,

$$\tilde{\mathbf{I}}|\boldsymbol{\theta}, \mathbf{V} \sim \mathcal{N} \;\; \mu_{\tilde{\mathbf{I}}|\boldsymbol{\theta},\mathbf{V}} = \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\tilde{\mathbf{I}}|\boldsymbol{\theta},\mathbf{V}} = \sigma^2 \boldsymbol{\mathcal{I}}_K \quad .$$

Given likelihood (8), prior (5), and posterior (6), applying Bayes' theorem gives

$$f_{\widetilde{\boldsymbol{\theta}}|\mathbf{I},\mathbf{V}}(\boldsymbol{\theta}) = \frac{L(\boldsymbol{\theta}; \mathbf{I}, \mathbf{V})\, \pi_{\widetilde{\boldsymbol{\theta}}}(\boldsymbol{\theta})}{\pi_{\widetilde{\mathbf{I}}|\mathbf{V}}(\mathbf{I})} = \frac{L(\boldsymbol{\theta}; \mathbf{I}, \mathbf{V})\, \pi_{\widetilde{\boldsymbol{\theta}}}(\boldsymbol{\theta})}{M}. \tag{9}$$

For realized current values $\mathbf{I}$ at voltages $\mathbf{V}$, integration of (9) allows the model evidence M to be computed as

$$M = \pi_{\widetilde{\mathbf{I}}|\mathbf{V}}(\mathbf{I}) = \int_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L(\boldsymbol{\theta}; \mathbf{I}, \mathbf{V})\, \pi_{\widetilde{\boldsymbol{\theta}}}(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}.$$

Numerically, M is a positive number that provides the necessary re-scaling so that the posterior distribution integrates to a total probability of one. However, the evidence derives its name from the fact that it quantifies how well the chosen model (with the specified prior SoKD on the model parameters) explains the observed data [7]. Larger evidence indicates a more adequate model, and it is often used for model comparison, model selection, or model averaging [7].

Conveniently, the evidence can be computed in o-BLR *without* integration. Solving (9) for the constant M gives

$$M = L(\boldsymbol{\theta}; \mathbf{I}, \mathbf{V})\, \pi_{\widetilde{\boldsymbol{\theta}}}(\boldsymbol{\theta}) \;\; f_{\widetilde{\boldsymbol{\theta}}|\mathbf{I},\mathbf{V}}(\boldsymbol{\theta}), \tag{10}$$

which can be evaluated at *any* $\boldsymbol{\theta}$ to compute M. For convenience, we evaluate the evidence at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. Because of potentially large values, we log-transform (10) in computations.

By maximizing the evidence, we investigate the selection of I-V data windows that reduce model discrepancy in the straight-line fit at I_{sc}. Starting with an I-V dataset ordered by voltage, a *core window* of I-V points is selected by taking the three I-V points with voltages closest to zero, including potentially negative voltages. We then grow the window in both positive and negative directions, one point at a time. Considering all combinations of such windows creates a 2-D discrete optimization space. To reduce the computational burden, data points with voltage above the voltage at the maximum power data point can be eliminated from consideration. This also avoids portions of the I-V curve with the highest propensity for systematic voltage errors from I-V curve corrections [8]. After choosing the appropriate data window that maximizes $\ln M$,

[5] In this application, $\mu_{\widetilde{a}_0}$ and $I_{95,\text{min}}$ are typically positive.

[6] Some univariate t-distribution definitions equivalently refer to the positive square root of $\Sigma_{\widetilde{a}_0}$ as the scale parameter.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

we then directly compute using (7) the measured value of, and the uncertainty in, $I_{\rm sc}$. This procedure can be automated, and the following examples suggest that the procedure is effective, although perhaps not fully optimal.

C. First I_{sc} Fitting Example

Figure 1 shows the results of applying this fitting procedure to I-V curve data that were synthetically generated with a non-ideal single-diode model for two cells in series, each with a bypass-diode and with a photocurrent mismatch that appears in forward-bias. This example has 51 equally spaced voltages and current values more common to PV devices with larger area cells. We see that maximizing the evidence properly excludes the two "knees" in the curve closest to $I_{\rm sc}$ that deviate from a straight line. This example shows that the window that corresponds to the maximum evidence is not necessarily the same window that corresponds to the minimum uncertainty. However, for this dataset, larger evidence does generally correspond to lower uncertainty.

For the data window that maximizes the evidence, $\boldsymbol{\theta} = (\widehat{a}_0, \widehat{a}_1, \widehat{\sigma}^2) = (5.984\ \text{A}, -0.1272\ \text{A/V}, 0.003720\ \text{A}^2)$. Thus, the measured value of $I_{\rm sc}$ is $\mu_{I_{\rm sc}} = 5.984$ A, whereas the true value is known to be 5.995 A, a relative error of -0.1778%. The relative expanded uncertainty $U_{95} = 0.6408\%$ corresponds to the interval $I_{95} = [5.946, 6.023]$, which covers the true value. Here, the minimum relative expanded uncertainty over all data windows is $0.5313\% < 0.6408\%$. Figure 2 shows the posterior SoKD for $I_{\rm sc}$, a shifted and scaled univariate t-distribution with $\nu = 18$ DoF. The slope a_1 could be analyzed similarly. Relative to $I_{\rm sc}$, the estimated standard deviation of the noise is $100 \cdot \frac{\widehat{\sigma}}{|\widehat{a}_0|} = 1.019\%$, whereas the true value is known to be 1%.

Figure 1 also shows that the uncertainty in the core window is significantly larger, even though this small window can be argued, *a priori*, to have minimal model discrepancy. Furthermore, the uncertainty is increased by taking data points in reverse-bias where a bypass diode is active or in forward bias into the mismatch feature. This is because the resulting increase in the sum of squared residuals dominates any uncertainty reduction due to the larger DoF. We also note the somewhat erratic structure of the graph of the coefficient of determination goodness-of-fit metric for LLS, denoted by R^2, as compared to the more smoothly structured log evidence graph. In general, the value of R^2 can be quite sensitive when fitting data that describe a nearly horizontal line. In this case, the maximal R^2 corresponds to a window that includes the non-straight-line mismatch feature. The core data window has a similarly large R^2, and is observed to be the global maximum for some synthetic data realizations.

The results above are merely one snapshot of a stochastic process that generates I-V curve data. Because synthetic data are used in this example, we can further consider various aspects of the statistical performance of the procedure using Monte Carlo (MC) simulation. The MC sampling enables many realizations of the random current noise in the I-V curve data. We consider two extreme window selections: (1) the smallest core window vs. (2) the larger maximum-evidence window.

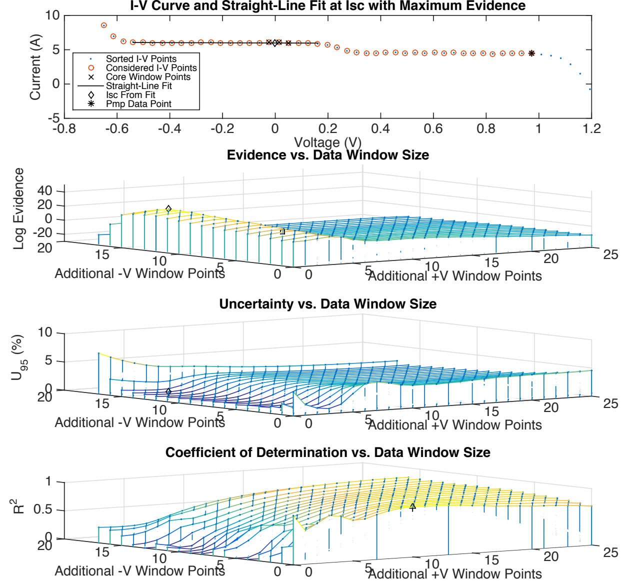


Fig. 1. Results for the First $I_{\rm sc}$ Fitting Example using synthetic I-V curve data. *Top:* A straight-line fit for $I_{\rm sc}$ that only extends over the data window that maximizes the evidence M. *Top-mid:* $\ln M$ as a function of window size, giving a 2-D optimization space. The core window corresponds to $(0, 0)$. *Bot-mid:* The corresponding relative expanded uncertainty U_{95} for the various windows. In the middle two plots, the diamond marks the window with the maximum evidence and the square marks the window with the minimum uncertainty. *Bot:* The coefficient of determination R^2 for linear least squares fits for each data window. The triangle marks the window with maximum R^2.

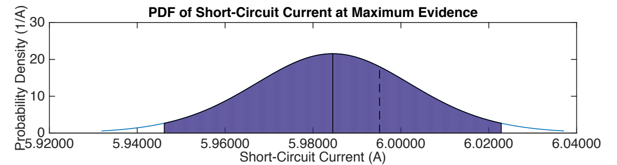


Fig. 2. The posterior SoKD for $I_{\rm sc}$ (a shifted and scaled univariate t-distribution) for the First $I_{\rm sc}$ Fitting Example using synthetic data. A dashed vertical line marks the known true value. A solid vertical line marks the mean (equal to the mode) and the area corresponding to the 95% coverage interval is shaded, representing a 95% state-of-knowledge probability.

We compare the average U_{95} for $I_{\rm sc}$ for the two windows by taking 100 realizations of the I-V curve data with the same voltages each time. The average U_{95} for the core window is 6.109%, while the average U_{95} for the maximum-evidence window is 0.6861%. Thus, a significant reduction in uncertainty is achieved, on average, by taking a larger window. However, for this sampling the proportion of times that the I_{95} coverage interval contains the true $I_{\rm sc}$ is 0.96 for the core window and 0.80 for the maximum-evidence window.[7]

[7] Here, a frequentist metric judges the coverage interval, which is technically a Bayesian credible interval. In this o-BLR setting, one anticipates commensurate frequentist performance in the absence of model discrepancy [11].

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

This suggests that the maximum-evidence windows may be somewhat too large and therefore include I-V points where model curvature is non-negligible. This is further supported by the average positive (negative) bias in the residuals at the leftmost (rightmost) data point of the maximum-evidence window, which is not present in the core window. Here, the residuals are defined as the observed current minus the straight-line fit current using the mode of the joint posterior SoKD for $\widetilde{a}_0$ and $\widetilde{a}_1$. Thus, we hypothesize that an optimal window, which minimizes uncertainty with negligible model discrepancy, lies somewhere in between the core and maximum-evidence windows.

We also use synthetic data to examine the effect of current scaling. Figure 3 shows the analogous results for an I-V curve with about 2% of the short-circuit current, which is more typical of a $4\ \text{cm}^2$ reference cell. Here, the magnitude of the additive noise is scaled so that it is still 1% of the true I_{sc}.[8] This figure shows a maximum-evidence data window that is clearly too large, with the effect of lowering the measured value for I_{sc}. The error in this value is -2.506%, which is considerably outside the coverage interval corresponding to $U_{95} = 1.928\%$. As the current is scaled lower, we observe that the global maximum evidence corresponding to a better window becomes a local maximum. We currently cannot explain this phenomenon. However, this suggests that using the local maximum evidence closest to the core window may be effective when selecting or constraining the data window.

Lastly, we use synthetic data to examine the effect of the noise level in the current. Figure 4 shows the results of changing the noise level on the maximum-evidence window. For this particular realization of the random noise, we see that the maximum-evidence window grows slightly larger before decreasing. As one might expect for smaller noise levels, it appears that the evidence can better distinguish the onset of model discrepancy. However, larger (and potentially unrealistic) noise levels tend to decrease the window size in a conservative fashion. Further analysis will be required to better understand these effects over a wide variety of I-V curve datasets and noise realizations. We note that the density and spacing of the voltage points as well as the accuracy of the noise model (e.g., i.i.d. normal) can also affect such results.

D. Second I_{sc} Fitting Example

Figure 5 shows the results of applying this fitting procedure to a measured I-V curve with lower current and a bypass-diode feature in reverse bias, in addition to the usual forward-bias diode feature. We see that maximizing the evidence includes a large portion of the flat part of the I-V curve, while properly excluding the two "knees" in the curve. Note that the window with maximum evidence is not the same window that gives the minimum uncertainty. However, for this dataset, larger evidence does generally correspond to lower uncertainty.

For the data window that maximizes the evidence, $\boldsymbol{\theta} = (\widehat{a}_0, \widehat{a}_1, \widehat{\sigma}^2) = (0.01258\ \text{A}, -2.044 \times 10^{-4}\ \text{A/V}, 1.546 \times 10^{-9}\ \text{A}^2)$. Thus, the measured value of I_{sc} is $\mu_{I_{\text{sc}}} = 0.01258$ A. The relative expanded uncertainty $U_{95} = 0.1441\%$ corresponding to the interval $I_{95} = [0.01256, 0.01260]$. Here, the minimum relative uncertainty over all data windows is $0.05319\% < 0.1441\%$. When considering the evidence as a goodness-of-fit metric, the window giving the smallest uncertainty is not preferred to the window giving the maximal evidence. Figure 6 shows the posterior SoKD for I_{sc}, a shifted and scaled univariate t-distribution with $\nu = 62$ DoF. The slope a_1 can be analyzed similarly. Relative to I_{sc}, the estimated standard deviation of the noise is $100 \cdot \frac{\widehat{\sigma}}{|\widehat{a}_0|} = 0.3126\%$. Comparing evidence to R^2 as goodness-of-fit metrics, one observes agreement or disagreement, depending on the window.

[8] For comparison purposes, we continue to consider two mismatched cells wired in series with the same exact voltages.

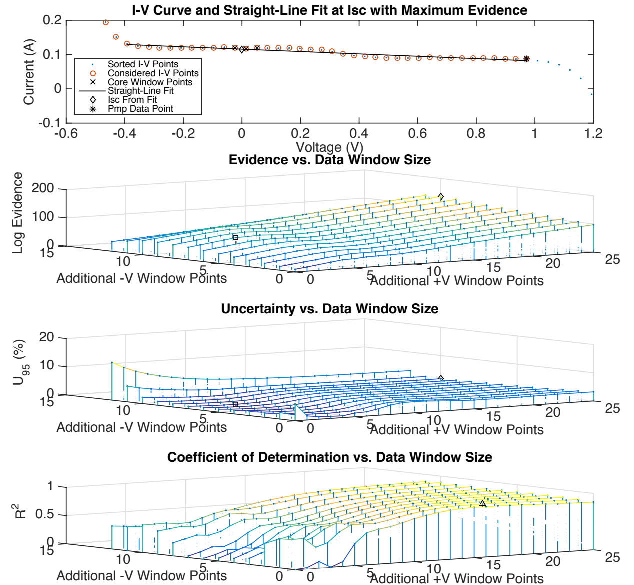
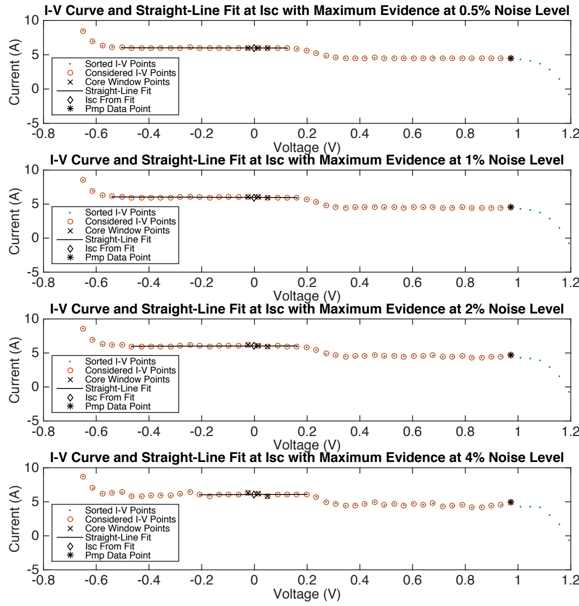


Fig. 3. Results for the First I_{sc} Fitting Example using synthetic I-V curve data with significantly lower current. *Top:* The maximum-evidence window is too large. *Top-mid:* $\ln M$ as a function of window size, notice the local maximum in evidence near the window with minimum uncertainty. *Bot-mid:* The corresponding relative expanded uncertainty U_{95} for the various windows. In the middle two plots, the diamond marks the window with the maximum evidence and the square marks the window with the minimum uncertainty. *Bot:* The coefficient of determination R^2 for linear least squares fits for each data window. The triangle marks the window with maximum R^2.

III. Conclusion

We use model evidence in the o-BLR framework to investigate an automated local straight-line regression to determine I_{sc} that reduces uncertainty while avoiding significant model discrepancy. The maximum-evidence approach appears promising and better than the R^2 goodness-of-fit metric. However, the diversity of PV devices, measurement systems, and measurement factors that affect the regression are challenging to explore exhaustively. Furthermore, model evidence is traditionally used for model comparison/selection, and its use here might benefit from a more rigorous framing in these terms.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.



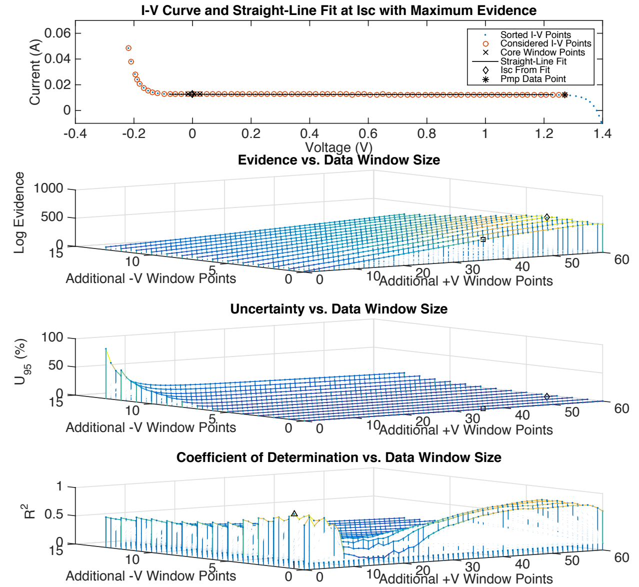


Fig. 4. Results for the First I_{sc} Fitting Example using synthetic I-V curve data with different noise levels. The standard deviation of current noise relative to I_{sc} is— *Top:* 0.5%, *Top-mid:* 1%, *Bot-mid:* 2%, and *Bot:* 4%. For this particular realization of the random noise, the maximum-evidence window first increases and then decreases in size with increasing noise level.

Fig. 5. Results for the Second I_{sc} Fitting Example using measured I-V curve data. *Top:* A straight-line fit for I_{sc} that only extends over the data window that maximizes the evidence M. *Top-mid:* $\ln M$ Log evidence as a function of window size, giving a 2-D optimization space. The core window corresponds to $(0, 0)$. *Bot-mid:* The corresponding relative expanded uncertainty U_{95} for the various windows. In the middle two plots, the diamond marks the window with the maximum evidence and the square marks the window with the minimum uncertainty. *Bot:* The coefficient of determination R^2 for linear least squares fits for each data window. The triangle marks the window with maximum R^2.

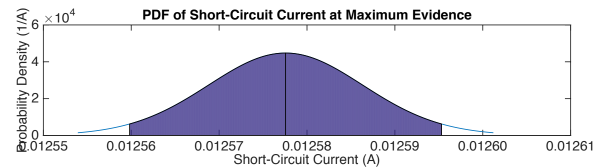


Fig. 6. The posterior SoKD for I_{sc} (a shifted and scaled univariate t-distribution) for the Second I_{sc} Fitting Example using measured data. The true value is unknown. A solid vertical line marks the mean (equal to the mode) and the area corresponding to the 95% coverage interval is shaded, representing a 95% state-of-knowledge probability.

Intuitively, the method is comparing models for which data points outside the chosen window have no quantitative bearing on the likelihood function. This apparently corresponds to an “improper” likelihood function that would be susceptible to delicate mathematical issues. A more rigorous development may explain the current-scalability issue and suggest a more optimal window selection for minimal fit uncertainty with negligible model discrepancy. Finally, we have provided a rigorous framework for analyzing other measurements and approaches, such as local polynomial fits to maximum power and increased I-V curve sampling near points of interest.

Acknowledgment

This work was supported by the U.S. Department of Energy under Contract No. DE-AC36-08-GO28308 with the National Renewable Energy Laboratory.

References

[1] ASTM Standard E1036-12, *Standard Test Methods for Electrical Performance of Nonconcentrator Terrestrial Photovoltaic Modules and Arrays Using Reference Cells*. West Conshohocken, PA, USA: ASTM International, 2012.

[2] J. Brynjarsdóttir and A. O’Hagan, “Learning about physical parameters: the importance of model discrepancy,” *Inverse Problems*, vol. 30, 2014.

[3] M. Thompson and S. L. R. Ellison, “Dark uncertainty,” *Accreditation and Quality Assurance*, vol. 16, October 2011.

[4] K. A. Emery and C. R. Osterwald, “PV Performance Measurement Algorithms, Procedures, and Equipment,” in *Proc. of the 21st Photovoltaic Specialists Conference*. IEEE, 1990.

[5] Joint Committee for Guides in Metrology (BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML), *Evaluation of measurement data—An Introduction to the “Guide to the expression of uncertainty in measurement” and related documents, JCGM 104:2009*. Sèvres, France: International Bureau of Weights and Measures (BIPM), 2009.

[6] R. Kacker, K.-D. Sommer, and R. Kessel, “Evolution of Modern Approaches to Express Uncertainty in Measurement,” *Metrologia*, vol. 44, pp. 513–29, 2007.

[7] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, 2nd ed. Boca Raton, FL: Chapman & Hall/CRC, 2004.

[8] M. Campanelli, K. Emery, R. Elmore, and B. Zaharatos, “Uncertainty Analysis for Maximum Power at SRC Using Hierarchical Monte Carlo Simulation,” in *Proc. of the 40th Photovoltaic Specialists Conference*. IEEE, 2014.

[9] M. Campanelli and K. Emery, “Device-Dependent Light-Level Correction Errors in Photovoltaic I-V Performance Measurements,” in *Proc. of the 39th Photovoltaic Specialists Conference*. IEEE, 2013.

[10] S. Kotz and S. Nadarajah, *Multivariate t Distributions and Their Applications*. New York, NY, USA: Cambridge University Press, 2004.

[11] E. T. Jaynes, *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*. Dordrecht: D. Reidel, 1976, ch. Confidence Intervals vs Bayesian Intervals, pp. 175–257.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.