STATISTICAL MODELING OF PHOTOVOLTAIC DEVICE PERFORMANCE

by

Brian Zaharatos

A thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of PhD (Applied Mathematics and Statistics).

Golden, Colorado

Date _____

Signed: _____
Brian Zaharatos

Signed: _____
Dr. Luis Tenorio
Thesis Advisor

Signed: _____
Dr. Paul Constantine
Thesis Advisor

Golden, Colorado

Date _____

Signed: _____
Dr. Willy Hereman
Professor and Applied Mathematics and Statistics Department Head
Department of Applied Mathematics and Statistics

ABSTRACT

Accurate performance modeling of photovoltaic (PV) devices is important for increasing the usage, reliability, and bankability of solar energy. For example, performance modeling of PV devices can help reduce costs and spur deployment by giving confidence to system engineers, customers, lenders, and investors. The performance of a PV device is typically determined by measuring its current-voltage (I-V) characteristics. These measurements, along with a performance model such as the single-diode model, are used to estimate important device performance parameters such as the maximum power output under a range of operating conditions. Many possible approaches to estimating these parameters have been explored in the PV literature. However, to the best of our knowledge, a statistical approach to PV performance modeling using measured I-V characteristics and the single-diode model has not yet been studied. In this thesis, we study several aspects of parameter estimation for the single-diode model, often within a statistical framework, in order to provide more reliable estimates of PV performance parameters.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I would like to thank my PhD advisors Luis Tenorio and Paul Constantine, National Renewable Energy Lab (NREL) mentor Mark Campanelli, and committee members William Navidi, and Aaron Porter; their guidance and support has been invaluable during each step of my graduate work. I feel humbled and honored to have had the chance to work with each of these individuals. Much of the work in this dissertation was financially supported by the U.S. Department of Energy and NREL (under Contract No. DE-AC36-08-GO28308); I am extremely grateful for this financial support and would like to thank everyone involved in securing this funding (including Mark Campanelli, Luis Tenorio, and Keith Emery). I would also like to thank Mahadevan Ganesh, Jahi Simbai, and Carl Mitcham for helping me secure funding for my first year at Colorado School of Mines. I will be forever grateful for their hard work.

I feel extraordinarily lucky to have had such a strong group of mentors during my academic career. It is hard to express how thankful that I am to have worked with Mark Campanelli. Mark has all of the qualities that a great mentor ought to have: he is ambitious, excited about his work, patient, intelligent, and has great intuition. I can only hope that the future brings more collaboration with Mark. In addition to my PhD and NREL mentors, I would like to thank several other fantastic mentors that have helped me develop intellectually, professionally, and personally. First, I would like to thank Michael Dickson (University of South Carolina) for giving me my first real exposure to clear and rigorous philosophical thinking. I am a better thinker and writer because of Michael's guidance. Second, I would like to thank Carl Mitcham. Carl has helped me in more ways that I can articulate. Because of Carl, I was able to complete my MA in philosophy (from the University of South Carolina), become a 'published philosopher', and start a career of interdisciplinary academic work. Third, I would like to thank Robert Tims for being an exemplar first mentor,

giving me my first internship, and my first exposure to industrial and applied mathematics. Finally, I would like to thank Andy Bardwell of Bardwell Consulting for providing me with consulting projects that helped focus my interests in statistics. I hope that one day I can be as helpful to others as these individuals have been to me.

In addition to those who have helped me develop professionally, I would like to thank my family and friends for their support. I would be nowhere without the constant support of my mother; I am honored to call such a strong, loving, and thoughtful woman 'Mom'. Thanks to my brother, Gary, for the laughs and companionship. And last, but certainly not least, thanks to Britney, for your support and for all the little ways that you make life so much better.

CHAPTER 1

INTRODUCTION

The most common way of converting energy from the sun into electricity is through the use of photovoltaic (PV) cells. PV cells are most often constructed of semi-conductors, such as crystalline silicon or thin film materials (e.g., cadmium telluride (CdTe), copper indium gallium selenide (CIGS) and amorphous silicon (a-Si)), which absorb radiation from the sun and emit electrons to be harnessed as usable electricity. Instantaneous electrical power and total energy yield are key performance metrics for PV. Because the electrical yield from a single PV cell is insufficient for many purposes, many cells are usually wired together in series and packaged into *modules* that are suitable for outdoor use. Modules themselves may be arranged in larger circuits to make an *array*. Fundamentally, a PV power *system* encompasses a durable electro-optical device and an electrical and mechanical support structure.

PV devices have many advantages over other forms of energy production technologies. For example, solar energy is indefinitely renewable; that is, the potential for solar energy conversion exists as long as light from the sun reaches the Earth's surface. PV devices are also one of the most environmentally sustainable energy sources, since, compared to fossil fuels, they contribute negligibly to the greenhouse effect, pollution, and climate change.[1] Finally, the cost associated with solar energy is rapidly declining; some analysts predict that the cost of solar energy will be competitive with traditional energy sources by 2025 [2]. Good characterization of PV device performance helps reduce costs and spur deployment by giving confidence to system engineers, customers, lenders, and investors. In order to keep solar energy competitive, new and better ways of characterizing performance are necessary.

[1]This claim is thought to be true even when considering the mining of materials needed for PV devices; see [1].

1.1 Background and Literature Review

The performance of a PV device is determined by measuring its current-voltage (I-V) characteristics (to be defined in Section 1.1.1) and using those measurements to estimate key performance parameters, such as the maximum power output of the device. Temperature and spectral irradiance are key environmental factors that affect device performance. To standardize performance comparisons, one often attempts to collect the I-V data at standard reporting conditions (SRC), defined as a standard spectral irradiance of 1000 W/m^2—under the standard spectrum defined in ASTM G173-03 [3]—and a temperature of 25°C. However, even in controlled laboratory conditions, achieving SRC is difficult. Frequently, the lamp used to simulate the spectral irradiance deviates slightly from 1000 W/m^2; further, the irradiance and temperature is measured with noise. Thus, a model of PV performance that takes into account these (and other) fluctuations in irradiance and temperature conditions, along with measurement noise, is desirable.

Using I-V data, one can characterize performance at and away from SRC by developing a mathematical model of the PV device. The selection of a PV performance model and its calibration from data are nontrivial. A performance model should (1) be sufficiently good representation of a wide range of PV devices; (2) be simple enough to allow for tractable estimation of the model and performance parameters; (3) make accurate predictions at a number of operating conditions (e.g., irradiance and temperature); (4) allow for statistical modeling of data measured with noise to quantify uncertainty in parameter estimates; and (5) have desirable statistical properties, such as parameter *identifiability*—i.e., two distinct values in the parameter space should not yield the same distribution of observables—and parameter *estimability*—i.e., estimation techniques, such as maximum likelihood estimation, should yield unique parameter estimates. In order to gain a better understanding of PV characterization and performance models that may satisfy (1)—(5), we review some basic PV physics.

1.1.1 PV Physics

When photons from a light source (e.g., the sun or a lamp) hit the photo-active material of a PV cell, electrons are "knocked loose" and guided to form an electric *current* and a *voltage*. A current is the amount of charge passing through a given point per unit time; voltage is the potential difference between two points, i.e., voltage describes the potential for current to flow between these two points. The materials used to create a PV device have particular *resistances*—physical characteristics that resist the flow of electrons.

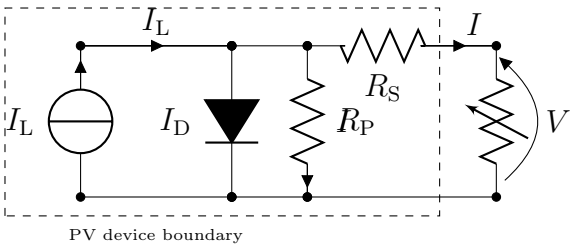


Figure 1.1: The light falling on the PV device, the photocurrent, I_L, enters the PV device, and flows in different directions: the current flowing through the diode is denoted I_D; other current flows through the parallel resistor,R_P; and through the series resistor, R_S.

Conventional PV cells operate as a *diode*. A diode allows current to pass in one *forward* direction, but not in the *reverse* direction (like a "one-way valve"). When light falls on the PV cell, current is generated by the diode. The I-V characteristics are determined by this diode and by the resistances of the semiconductor materials and other components.

The circuit diagram in Figure 1.1 gives a visual representation of these I-V characteristics. The light falling on the PV device generates a photocurrent, I_L, which flows in different directions: some flows through the diode, denoted I_D; some flows through the parallel resistor, denoted I_P (in Figure 1.1, the parallel resistance itself is denoted R_P); and the remaining current passes through the series resistor, denoted I_S (similarly, in Figure 1.1, the series resistance is denoted R_S). By Kirchhoff's law [4], the current output is equal to the (signed) sum of photocurrent, diode current, and parallel resistor current:

$$I = I_L - I_D - I_P. \tag{1.1}$$

The Shockley diode equation [5] is often used to describe the flow of current through the diode in equation (1.1):

$$I_{\mathrm{D}} = I_{\mathrm{S}}\left(e^{\frac{V+IR_{\mathrm{S}}}{N_{\mathrm{S}}nV_{\mathrm{th}}}} - 1\right),$$

where $V_{\mathrm{th}} := k_{\mathrm{B}}\,T/q$ and:

$I_{\mathbf{S}}$ diode reverse saturation current (A)

$N_{\mathbf{S}}$ # of cells connected in series (unit-less)

n ideality factor (unit-less)

$R_{\mathbf{S}}$ device-level series resistance (Ω)

$k_{\mathbf{B}}$ Boltzmann constant ($1.3806504 \times 10^{-23}$ J/K)

q electron charge ($1.602176487 \times 10^{19}$ C)

T junction temperature (K).

The current flowing through the parallel resistor can be described as

$$I_{\mathrm{P}} = \frac{V + IR_{\mathrm{S}}}{R_{\mathrm{P}}},$$

where R_{P} is the parallel (shunt) resistance of device, measured in Ohms (Ω). Finally, putting these components together yields:

$$I = I_{\mathrm{L}} - \underbrace{I_{\mathrm{S}}\left(e^{\frac{V+IR_{\mathrm{S}}}{N_{\mathrm{S}}nV_{\mathrm{th}}}} - 1\right)}_{I_{\mathrm{D}}} - \underbrace{\frac{V + IR_{\mathrm{S}}}{R_{\mathrm{P}}}}_{I_{\mathrm{P}}}. \tag{1.2}$$

In general, the values I_{L}, I_{S}, n, R_{S}, and R_{P} in equation (1.2) are properties of the specific device and are generally functions of temperature (T) and irradiance (E). They are described by the following *auxiliary equations*:

$$
\begin{aligned}
I_{\mathrm{L}} &= \psi_{I_{\mathrm{L}}}(T, E; \boldsymbol{\theta}_{I_{\mathrm{L}}}) \\
I_{\mathrm{S}} &= \psi_{I_{\mathrm{S}}}(T, E; \boldsymbol{\theta}_{I_{\mathrm{S}}}) \\
n &= \psi_{n}(T, E; \boldsymbol{\theta}_{n}) \\
R_{\mathrm{S}} &= \psi_{R_{\mathrm{S}}}(T, E; \boldsymbol{\theta}_{R_{\mathrm{S}}}) \\
R_{\mathrm{P}} &= \psi_{R_{\mathrm{P}}}(T, E; \boldsymbol{\theta}_{R_{\mathrm{P}}}).
\end{aligned}
$$

The vectors $\boldsymbol{\theta}_{I_{\mathrm{L}}}, \boldsymbol{\theta}_{I_{\mathrm{S}}}, \boldsymbol{\theta}_{n}, \boldsymbol{\theta}_{R_{\mathrm{S}}}$, and $\boldsymbol{\theta}_{R_{\mathrm{P}}}$ represent parameters specific to the particular choice of auxiliary equation, and often include parameter values at SRC. For example, $\boldsymbol{\theta}_{R_{\mathrm{P}}}$ may include the component R_{P_0}, the parallel resistance of the device when measured at SRC conditions. There is no consensus on the choice of auxiliary equations; they can be different for different device types, or they may correspond to a greater modeling simplification than another. In this dissertation, we develop a simple set of auxiliary equations that describes the I-V characteristics at $T = T_0 = 25°\mathrm{C}$ and E close to 1000 $\mathrm{W/m^2}$.

1.1.2 Parameter Estimation

In single-diode parameter estimation problems, we are generally interested in estimating the model parameters $\boldsymbol{\theta} = (I_{\mathrm{L}_0}, I_{\mathrm{S}_0}, n_0, R_{\mathrm{S}_0}, R_{\mathrm{P}_0})$. From these parameters, we can estimate the *key performance parameters*: the short circuit current, I_{SC}, the open circuit voltage, V_{OC}, and the maximum power output, P_{max}.[2] Each of these quantities is a parameter in the statistical sense, meaning that, neglecting changes that occur over time (e.g., degradation), each is a fixed quantity that cannot be directly measured.

There are many possible approaches to estimating the parameters of the single-diode model. The authors of [6] and [7] provide an overview of estimation techniques found in the PV literature. Broadly, these techniques fall into two categories: approaches that use limited data, for example, approximations of I_{SC} and V_{OC} at SRC (such approximations are often used because they are readily given by the PV device's manufacturer) and approaches that

[2]Note that in some cases, we estimate the short circuit current, I_{SC_0}, directly, rather than I_{L_0}.

use data from a full range I-V curve. In order to have a sufficiently large set of equations from which to estimate the parameters, approaches in the former category (e.g., [8], [9], [10]) use information contained on a PV manufacturer's datasheet. Approaches in the latter category (e.g., [11], [12], [13]) have a system of nonlinear equations as large as the sample size of the I-V data. Both approaches use either deterministic iterative root finding or optimization methods—e.g., using Newton's method [9], genetic algorithms [14], particle swarm optimization [15], or sequential techniques [6]—to solve for the parameters.

To the best of our knowledge and based on the reviews in [6] and [7], statistical approaches to estimation of model parameters from noisy measurement data would be novel in the field of PV characterization and would provide important advantages over the current practices. First, statistical approaches to parameter estimation allow us to quantify our uncertainty in the estimates—using a statistical model for measurement noise in the data—by reporting either confidence intervals (under the frequentist paradigm) or credible intervals (under the Bayesian paradigm).

Statistical methods may also provide a "standard" parameter estimation method for PV performance. The authors of [6] write that "without a standard method for estimating model parameters, different modelers may obtain different model parameter values even when the parameter values are derived from a common set of measurements." To some degree, non-uniqueness is to be expected; for example, for a fixed sample size, well-defined maximum likelihood and method of moments estimators can yield different values. However, it is desirable to know whether any particular method yields a unique answer. Non-uniqueness of an estimator may be due to the underlying structure of the model (what we call *non-identifiability* in Chapter 4) or may be an artifact of the estimation method; for example, MLEs are not guaranteed to be unique. In this dissertation, when considering models and performance estimates, we pay special attention to the uniqueness of the maximum likelihood estimator. Specifically, for each model, we study the uniqueness of the MLE for $\boldsymbol{\theta}$ and the key performance parameters ($P_{\max}$, I_{SC}, and V_{OC}); uniqueness of the MLE is desirable because

it (a) implies that the model is free from structural issues and (b) it may circumvent the worries in [6] by providing a standard method for single-diode model parameter estimation within the PV community.

1.2 Chapter Outline

In this dissertation, we study two variations of the single-diode model that are believed to satisfy the properties (1)—(5) listed in Section 1.1 to different degrees. We choose variations of the single-diode model described in Section 1.1.1 because these models are often thought to strike a reasonable balance between simplicity, mathematical tractability, and descriptive power. To the best of our knowledge, the literature on single-diode model parameter estimation does not include a sensitivity analysis of any of the objective functions (i.e., the functions to be minimized) that are used in parameter estimation. Such analyses are important because they help us understand the way that the output of the objective function changes with changes in inputs and whether the parameter estimation problem is well-posed. Further, in the case of ill-posedness, the sensitivity analysis may suggest potential regularization methods. In Chapter 2, we conduct sensitivity analyses on an objective function constructed using a version of the single-diode model that often arises in parameter estimation problems. The goal of this chapter is to gain insight into the structure of this function so that one could construct a more efficient estimation procedure.

In Chapter 3 and Chapter 4 we consider a set of auxiliary equations of the single-diode model that are thought to accurately describe the I-V performance of a large set of PV devices at a fixed temperature of 25°C and at irradiance conditions near 1000 $\mathrm{W/m^2}$. After describing this model, we set out to derive estimates for the model and key performance parameters. The latter are functions of the model parameters. Specifically, we use maximum likelihood estimation (MLE) and Bayesian inference to derive estimates of all of the parameters in question. Because single-diode models are nonlinear and their parameter space is of several dimensions, it is not at all clear whether the model parameters are identifiable or estimable. These properties are important for making statistical inference. For example, if a model's

parameters are not identifiable, then two (or more) distinct values in the parameter space will generate the same data distribution. In such a case, parameter estimation is not well-defined because it's not clear which of the two values in the parameter space actually did generate the data. For this reason, we attempt to learn about the estimability, and to some extent, the identifiability of the model parameters using the method known as *data cloning* [16]. After exploring these properties, and after exploring some merits and drawbacks of the data cloning algorithm, we calculate estimates of the parameters with quantified uncertainty.

In Chapter 5, in collaboration with Paul Constantine, we conduct a sensitivity analysis of the function $f(\boldsymbol{x}) = P_{\max}$, where the vector $\boldsymbol{x}$ is made up of the five single-diode model parameters described in Chapter 3. In particular, the goal of this work is to determine which input components of $f(\boldsymbol{x})$ have the most impact on changes in the output quantity $P_{\max}$. Knowing which inputs have the most impact on the output is useful because one can then construct a lower dimensional function that approximates $f(\boldsymbol{x}) = P_{\max}$. Such an approximation may render estimates of $P_{\max}$ more tractable as it would decrease the dimensionality of the problem. Alternatively, this analysis may suggest how precisely the input parameters must be specified to ensure accurate predictions of $P_{\max}$.

CHAPTER 2

SENSITIVITY ANALYSIS OF THE SINGLE DIODE MODEL

Abstract

The five parameter single-diode model is often used to characterize the electrical performance of photovoltaic (PV) devices. However, calibrating the single-diode model often presents challenges. Calibration involves the minimization of a five-dimensional, nonlinear objective function, and effectively searching a five-dimensional parameter space for a global minimum can be difficult and costly. Consequently, one might like to know whether dimension reduction is possible. In this paper, we attempt to advance PV calibration efforts by studying an objective function that is minimized in single-diode model calibration problems. Specifically, we study whether this objective function admits an active subspace. Discovering an active subspace can be helpful because it points to the important directions in the parameter space that yield the greatest change in the objective function.

2.1 Introduction

Good characterization of photovoltaic (PV) device performance helps reduce costs and spur deployment by giving confidence to system engineers, customers, lenders, and investors. In order to keep solar energy competitive, new and better ways of characterizing performance are necessary. A typical characterization process involves taking current-voltage (I-V) measurement data from a particular device, and using a mathematical model to infer particular model and performance parameters for the device. For example, for the commonly used single-diode performance model (e.g., [17], [18]), model parameters may include series and parallel resistances (R_{S} and R_{P}, respectively); device performance parameters include the maximum power output, P_{max}, the short circuit current, I_{SC}, and the open circuit voltage V_{OC}. Using a mathematical relationship, such as the single-diode model, to infer the value

of a set of parameters is often called the *calibration problem* or the *parameter estimation problem*.

In general, calibrating a mathematical model reduces to minimizing an objective function over a parameter space. One often assumes that there exists a mathematical relationship—called the *mathematical model*—between measured data and the true, unobserved, and unknown *model parameters*. The mathematical model, $\boldsymbol{f}$, relates the measured data, $\boldsymbol{d}$—which may or may not be measured with noise—to the model parameters, $\boldsymbol{\theta}$. That is, given the true parameters, $\boldsymbol{\theta}^*$, and noise free measurements,

$$\boldsymbol{f}(\boldsymbol{d}; \boldsymbol{\theta}^*) = \mathbf{0}.$$

One hopes to recover a good estimate of $\boldsymbol{\theta}^*$, denoted $\widehat{\boldsymbol{\theta}}$, by minimizing $\|f\|^2$ over the parameter space $\boldsymbol{\Theta}$:

$$\begin{aligned}\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ \|\boldsymbol{f}(\boldsymbol{d};\boldsymbol{\theta})\|^2 &= \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ \underbrace{\log(\|\boldsymbol{f}(\boldsymbol{d};\boldsymbol{\theta})\|^2)}_{\Phi(\boldsymbol{\theta})} \\ &= \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ \Phi(\boldsymbol{\theta}).\end{aligned}$$

In cases where $\boldsymbol{f}$ is nonlinear and $\boldsymbol{\theta}$ is in many dimensions, computing $\widehat{\boldsymbol{\theta}}$ can be difficult. Consequently, one might like to know whether *dimension reduction* is possible. Dimension reduction is important because it allows one to work in a space with a lower number of variables, which (ideally) reduces complexity and computation time. One type of dimension reduction, called *subset dimension reduction*, occurs when one discovers that there is a subset of the components of $\boldsymbol{\theta}$ that contribute most to changes in the *objective function*, $\Phi(\boldsymbol{\theta})$. When this type of dimension reduction is not possible—that is, when $\Phi(\boldsymbol{\theta})$ depends on each component of $\boldsymbol{\theta}$—it is still reasonable to ask whether $\Phi(\boldsymbol{\theta})$ depends on a particular linear combination of the components of $\boldsymbol{\theta}$ [19]. This linear combination approach, called *subspace dimension reduction*, generalizes the subset approach; one can think of subspace dimension reduction as the linear combination that assigns a weight of one to "important" components,

and a weight of zero to all others [20].

In this paper, we attempt to advance PV parameter estimation efforts by studying an objective function that is often minimized in single-diode model calibration problems. To the best of our knowledge, no sensitivity analysis or dimension reduction techniques have been applied in this context. In particular, in Sections 2.2 and 2.3, we describe two methods for analyzing the sensitivity of an objective function $\Phi(\boldsymbol{\theta})$: the Morris one-at-a-time method and the active subspace method. In Section 2.4 we construct an objective function from a typical single-diode model found in the PV literature. Then, in Section 2.5, we apply the methods in Sections 2.2 and 2.3 to the objective function described in Section 2.4.

2.2 Morris One-at-a-Time Method

Before using an optimization procedure (e.g., nonlinear least squares) on an objective function $\Phi(\boldsymbol{\theta})$, with a high dimensional input space, it is important to understand how the output of $\Phi(\boldsymbol{\theta})$ changes with respect to changes in its input, $\boldsymbol{\theta}$; that is, it is important to study the *sensitivity* of $\Phi(\boldsymbol{\theta})$ to changes in $\boldsymbol{\theta}$. The Morris's one-at-a-time (OAT) method is one method for studying such changes. OAT falls in the class of subset dimension reduction techniques because OAT is designed to identify a subset of the components of $\boldsymbol{\theta}$ that cause the most changes in $\Phi(\boldsymbol{\theta})$.

To implement OAT, we assume that the inputs of $\Phi(\boldsymbol{\theta})$ come from a hypercube, $[-1,1]^N$, and we divide this input space into a *p-level grid*, where each input, θ_i takes on values in the set

$$\Omega_i = \{-1, \frac{3-p}{p-1}, \frac{5-p}{p-1}, ..., 1\}.$$

The author of [21], [22] defines the *elementary effects* attributed to the i^{th} input to be

$$d_i(\boldsymbol{\theta}) = \frac{\Phi(\theta_1, \theta_2, ..., \theta_{i-1}, \theta_i + \Delta, \theta_{i+1}, ..., \theta_N) - \Phi(\theta_1, \theta_2, ..., \theta_N)}{\Delta}$$

where $i \in \{1, ..., N\}$, Δ is a predetermined multiple of $p/(p-1)$, and $\boldsymbol{\theta} \in \boldsymbol{\Omega}$, except that $\theta_i \leq 1 - \Delta$. If one *randomly samples*[3] $\boldsymbol{\theta} \in \Omega$ and computes d_i, then d_i is a random quantity with a probability distribution. The author of [22] shows that the moments of the distribution $d_i \sim F_i$ provide information about the importance of θ_i with respect to the output of $\Phi(\boldsymbol{\theta})$. Particularly, a large absolute measure of central tendency (first moment) provides evidence that θ_i has an important global influence on the output of $\Phi(\boldsymbol{\theta})$; a large measure of spread (second moment) provides evidence that there are interactions among inputs. To estimate the moments of F_i, one computes $d_i(\boldsymbol{\theta})$ many times for random draws of $\boldsymbol{\theta}$ and use sample statistics

$$\widehat{\mu}_{d_i} = \frac{1}{K} \sum_{k=1}^{K} d_{i_k}, \tag{2.1}$$

and

$$\widehat{\sigma}^2_{d_i} = \frac{1}{K-1} \sum_{k=1}^{K} (d_{i_k} - \widehat{\mu}_{d_i})^2, \tag{2.2}$$

as estimates of the true mean, μ_{d_i}, and variance, $\sigma^2_{d_i}$, of $d_i \sim F_i$. It has been proposed in [23] that one use $\widehat{\mu}^*_{d_i}$, defined as an estimate of the mean of the distribution, G_i, of $|d_i|$. The estimator $\widehat{\mu}^*_{d_i}$ is thought to be a better estimate of the importance of θ_i because when the distribution F_i contains both positive and negative elements, some important effects cancel each other out when computing $\widehat{\mu}_{d_i}$; such canceling results in a low mean when, in fact, the corresponding effect might be important (i.e., computing $\widehat{\mu}_{d_i}$ can lead to a type II error with respect to the importance of the input θ_i). In Section 2.4, we apply the OAT method to an objective function derived from the single-diode model in Section 2.4.

[3]Sampling can be done in different ways. For the results obtained in Section 2.5, the `MATLAB` function `morris_sa` used Latin hypercube sampling.

2.3 Active Subspace Methods

Active subspace methods differ from OAT in important ways. For one, active subspaces are best categorized as *subspace* rather than *subset* dimension reduction techniques. In subspace dimension reduction, one seeks to find a linear combination of the parameters that accounts for most of the change in $\Phi(\boldsymbol{\theta})$. If an active subspace exists, one may then exploit this structure to more efficiently optimize $\Phi(\boldsymbol{\theta})$.

We begin defining the active subspace of the model $\Phi(\boldsymbol{\theta})$ by defining a matrix, $\boldsymbol{C}$, as the average—with respect to the weight function $\rho(\boldsymbol{\theta})$—of the outer product of the gradient, $\nabla_{\boldsymbol{\theta}}\Phi = [\partial\Phi/\partial\theta_1, ..., \partial\Phi/\partial\theta_N]^T$ with itself:

$$\boldsymbol{C} = \int_{\boldsymbol{\Theta}} (\nabla_{\boldsymbol{\theta}}\Phi)(\nabla_{\boldsymbol{\theta}}\Phi)^T \rho(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{2.3}$$

For the purposes of this application, the function $\Phi(\boldsymbol{\theta})$ is an objective function of some optimization problem, and, in particular, in Section 2.5, it will be an objective function constructed from the single-diode model. In general, one can define the active subspace for a generic differentiable and Lipschitz continuous[4] function that takes a vector of inputs and yields a scalar output [24]. If one interprets the weight function $\rho(\boldsymbol{\theta})$ as a probability density function for $\boldsymbol{\theta}$, then $\nabla_{\boldsymbol{\theta}}\Phi$ can be interpreted as a random vector, and consequently, $\boldsymbol{C}$ can be interpreted as the uncentered covariance matrix of the gradient [20]. However, such interpretations are mainly used to help gain some intuition. Strictly speaking, the function under investigation (and its partial derivatives) is a scalar function of a fixed (nonrandom) vector. $\boldsymbol{C}$ is $N \times N$, is symmetric and positive semidefinite, and thus has a real eigenvalue decomposition,

$$\boldsymbol{C} = \boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}^T, \tag{2.4}$$

[4]This condition ensures that the norm of the gradient is bounded [20].

where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_N)$, $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N \geq 0$, and where $\boldsymbol{W}$ is the $N \times N$ orthogonal matrix whose columns $\{\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_N\}$ are the normalized eigenvectors of $\boldsymbol{C}$. The eigenvectors $\{\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_N\}$ can often be separated into two sets: one set, $\boldsymbol{W}_1 = [\boldsymbol{w}_1, ..., \boldsymbol{w}_n]$, whose corresponding eigenvalues are 'large', and another set, $\boldsymbol{W}_2 = [\boldsymbol{w}_{n+1}, ..., \boldsymbol{w}_N]$, whose corresponding eigenvalues are 'small', with $1 \leq n \leq N$. The *active subspace* is defined as the span of the column vectors in $\boldsymbol{W}_1$. The *active variable* is defined as $\boldsymbol{y} = \boldsymbol{w}_1^T \boldsymbol{\theta}$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

The term 'active subspace' has been adopted from from Russis 2010 Ph.D. dissertation [25]. Lemma 2.2 in [19] helps justify this this terminology. In words, the lemma states that $\Phi(\boldsymbol{\theta})$ changes more, on average (with respect to the weight function $\rho(\boldsymbol{\theta})$), when components of $\boldsymbol{\theta}$ are perturbed along the directions of $\boldsymbol{W}_1$ than along directions of $\boldsymbol{W}_2$.

Discovering an active subspace can be helpful because it points to the 'important' directions in the parameter space that yield the greatest change in $\Phi(\boldsymbol{\theta})$. Unimportant directions—directions that yield little change in $\Phi(\boldsymbol{\theta})$—provide evidence that many values of $\boldsymbol{\theta} \in [-1, 1]^N$ nearly minimize $\Phi(\boldsymbol{\theta})$. Once an active subspace is determined, one may exploit the active subspace to improve or accelerate the optimization method [20].

2.3.1 Estimating An Active Subspace

An active subspace, if one exists, is a property of a function. Its existence is not known a priori; rather, an active subspace must be estimated using computer simulations. In this section we describe a Monte Carlo method used to search for and estimate an active subspace.[5] The algorithm used is as follows [24]:

1. Draw M samples $\boldsymbol{\theta}_j$ from the weight function $\rho(\boldsymbol{\theta})$; often, $\rho(\boldsymbol{\theta})$ is set to a joint uniform distribution on $[-1, 1]^N$.

[5]Strictly speaking, active subspace methods are applied to a deterministic function of deterministic inputs. Using Monte Carlo to estimate $\boldsymbol{C}$ requires random draws of $\boldsymbol{\theta}$ from $\rho(\boldsymbol{\theta})$. Because of this, it is tempting to apply some sort of probabilistic interpretation to active subspace methods. Such interpretations are not necessary; theoretically, one could circumvent random sampling in order to estimate $\boldsymbol{C}$ using quadrature (however, such methods are difficult in high dimensions).

2. Compute $\nabla_{\boldsymbol{\theta}}\Phi_j = \nabla_{\boldsymbol{\theta}}\Phi(\boldsymbol{\theta}_j)$.

3. Approximate

$$\boldsymbol{C} \approx \widehat{\boldsymbol{C}} = \frac{1}{M}\sum_{j=1}^{M}(\nabla_{\boldsymbol{\theta}}\Phi_j)(\nabla_{\boldsymbol{\theta}}\Phi_j)^T. \tag{2.5}$$

4. Compute the real eigenvalue decomposition $\widehat{\boldsymbol{C}} = \widehat{\boldsymbol{W}}\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{W}}^T$.

The resulting matrices $\widehat{\boldsymbol{C}}$, $\widehat{\boldsymbol{\Lambda}}$, and $\widehat{\boldsymbol{W}}$ are *estimates* of the true matrices $\boldsymbol{C}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{W}$, respectively. This eigendecomposition contains the eigenvectors used to construct the estimate of the active variable, $\hat{\boldsymbol{y}} = \hat{\boldsymbol{w}}_1^T\boldsymbol{\theta}$.

2.3.2 Assessing the Variability of the Eigenvalues of $\widehat{C}$

Once estimators are obtained, it is desirable to know how much variability exists in the computation of $\widehat{\boldsymbol{\Lambda}}$ and $\widehat{\boldsymbol{W}}$. Large variability in $\widehat{\boldsymbol{\Lambda}}$ and $\widehat{\boldsymbol{W}}$ would imply that the computational procedure has not converged. Here, we describe a method from [20] to assess the variability in the estimation procedure of $\widehat{\lambda}_i$, i.e., the eigenvalues of $\widehat{\boldsymbol{C}}$ and diagonal elements of $\widehat{\boldsymbol{\Lambda}}$.[6]

1. Begin with the previously computed quantities $\nabla_{\boldsymbol{\theta}}\Phi_j$, $\widehat{\boldsymbol{\Lambda}}$, and $\widehat{\boldsymbol{W}}$, for $j \in \{1, ..., M\}$.

2. Choose the number of bootstrap evaluations, M_{b}. Let $i \in \{1, ..., M_{\mathrm{b}}\}$:

   (a) Draw a sample of size M, say, $\{j_1, ..., j_M\}$, from $\{1, ..., M\}$ with replacement.

   (b) Compute the bootstrap replicate:

$$\widehat{\boldsymbol{C}}_i^* = \frac{1}{M}\sum_{k=1}^{M}(\nabla_{\boldsymbol{\theta}}\Phi_{j_k})(\nabla_{\boldsymbol{\theta}}\Phi_{j_k})^T. \tag{2.6}$$

   (c) Compute the eigendecomposition

$$\widehat{\boldsymbol{C}}_i^* = \widehat{\boldsymbol{W}}_i^*\widehat{\boldsymbol{\Lambda}}_i^*\widehat{\boldsymbol{W}}_i^{*T} \tag{2.7}$$

[6] We note that this method does not make any claims about the *bias* of the estimation procedure. Such claims are currently under investigation.

3. Each $\widehat{\boldsymbol{\Lambda}}_i^*$ and $\widehat{\boldsymbol{W}}_i^*$ are estimators of $\boldsymbol{\Lambda}$ and $\boldsymbol{W}$, respectively. Thus, both $\widehat{\boldsymbol{\Lambda}}_i^*$ and $\widehat{\boldsymbol{W}}_i^*$ have a *distribution*, and we can calculate the 2.5 and 97.5 percentiles of the distributions to obtain a 95% "bootstrap interval" for λ_i and the entries of $\boldsymbol{W}_1$.

The resulting 95% bootstrap intervals can be used to provide evidence of gaps in the estimated eigenvalues. That is, a gap present between intervals for eigenvalues $\widehat{\lambda}_i$ and $\widehat{\lambda}_{i+1}$ provides evidence that the same gap exists among the true eigenvalues λ_i and λ_{i+1}

2.4 The Single Diode Model

In this paper, we focus on objective functions constructed from the PV single-diode model described in [26]. We choose the single-diode model as a performance model because we believe that, for a wide range of PV devices, it allows for a good compromise between accuracy and simplicity [27]. The following single-diode lumped-parameter equivalent-circuit model adequately describes the I-V performance of many series-wired photovoltaic (PV) devices:

$$I = I_{\mathrm{L}} - I_{\mathrm{S}}\left(e^{\frac{V+IR_{\mathrm{S}}}{N_{\mathrm{S}}nV_{\mathrm{th}}}} - 1\right) - \frac{V + IR_{\mathrm{S}}}{R_{\mathrm{P}}}, \tag{2.8}$$

where the thermal voltage is $V_{\mathrm{th}} := k_{\mathrm{B}}\,T/q$, and:

I device terminal current (A)

V device terminal voltage (V)

$I_{\mathbf{L}}$ photocurrent (A)

$I_{\mathbf{S}}$ diode reverse saturation current of device (A)

$N_{\mathbf{S}}$ number of cells connected in series (unit-less)

n ideality factor of device (unit-less)

$R_{\mathbf{S}}$ series resistance of device (Ω)

$R_{\mathbf{P}}$ parallel (shunt) resistance of device (Ω)

$k_{\mathbf{B}}$ Boltzmann constant ($1.3806504 \times 10^{-23}$ J/K)

T junction temperature of device (K)

q electron charge ($1.602176487 \times 10^{19}$ C).

The values I_{L}, I_{S}, n, R_{S}, and R_{P} in equation (5.9) are, in general, functions of the operating conditions temperature (T) and effective irradiance (E). These functions—often called *auxiliary equations*— describe the way in which these values change as T and E deviate from *standard reporting conditions* (SRC), defined as $T_0 = 25^\circ$ C and an irradiance of 1000 W/m^2 under the standard spectrum defined in ASTM G173-03 [3]. If operating conditions are well-controlled, as is sometimes the case in a characterization laboratory setting, then the identify auxiliary equations are thought to be reasonable: $I_{\mathrm{L}} = I_{\mathrm{L}_0}$, $I_{\mathrm{S}} = I_{\mathrm{S}_0}$, $n = n_0$, $R_{\mathrm{S}} = R_{\mathrm{S}_0}$, $R_{\mathrm{P}} = R_{\mathrm{P}_0}$. Here, the $-_0$ notation represents the quantity at SRC. For notational convenience, we let $\mathbf{x} = (I_{\mathrm{L}_0}, I_{\mathrm{S}_0}, n_0, R_{\mathrm{S}_0}, R_{\mathrm{P}_0})$ and refer to $\mathbf{x}$ as the *parameters* of the single-diode model with identity auxiliary equations. The authors of [28] and [29] describe other auxiliary equations that contain different parameters and that are thought to be reasonable when T and E deviate from their values at SRC.

Table 2.1: Two different parameter spaces are used in the Active Subspace analysis.

parameter	$\mathbf{\Theta}_{\text{narrow}}$	$\mathbf{\Theta}_{\text{wide}}$
$\mathbf{x}_1$	$[1/12, 0.5]$	$[1/12, 12]$
$\mathbf{x}_2$	$[10^{-11}, 10^{-7}]$	$[10^{-13}, 10^{-5}]$
$\mathbf{x}_3$	$[1.4, 1.7]$	$[1, 2]$
$\mathbf{x}_4$	$[0, 1]$	$[0, 1]$
$\mathbf{x}_5$	$[50, 5000]$	$[1, 5 \times 10^4]$

In this work, for $i \in \{1, ..., 5\}$, we define the parameter space of $\mathbf{x}$ as

$$\mathbf{X} = \{(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_5) |\ \mathbf{x}_i \in \mathbf{X}_i\}. \tag{2.9}$$

where $\mathbf{X}_i \subseteq \mathbb{R}$. In this work, we repeat our analyses on two distinct parameter spaces. The first parameter space is comprised of a device specific range for each parameter, and we denote $\mathbf{\Theta}_{\text{narrow}}$. In this case, the range for each parameter is small. The second parameter space captures a large number of devices (e.g., all devices in [17]), and is denoted by denoted $\mathbf{\Theta}_{\text{wide}}$. In this case, the range for each parameter is large. Table 2.1 gives both parameter spaces used in these analyses. We choose two different parameter spaces to show that the

choice of parameter space can affect the results of the sensitivity analysis.[7]

For computational reasons, we conduct our sensitivity analyses in a normalized region of the parameter space. Specifically, we work in log space for $\mathbf{x}_2$ and use the domains in Table 2.1 to normalize $\mathbf{x}$ to the hypercube $[-1,1]^5$. The normalized parameters, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, are defined as:

$$\theta_i = 2\left(\frac{x_i - A_i}{B_i - A_i}\right) - 1, \quad i \in \{1,3,4,5\}, \tag{2.10}$$

$$\theta_2 = 2\left(\frac{\log(x_2) - \log(A_2)}{\log(B_2) - \log(A_2)}\right) - 1, \tag{2.11}$$

where A_i and B_i are the lower and upper bounds, respectively, of the domain of the device specific or general parameter space for $\mathbf{x}_i$, $i \in \{1,...,5\}$.

At this point, we are in the position to derive the objective function using the identity auxiliary equations. Specifically, (2.8) implies that

$$I_{\mathrm{L}} - I_{\mathrm{S}}\left(e^{\frac{V+IR_{\mathrm{S}}}{N_{\mathrm{S}}nV_{\mathrm{th}}}} - 1\right) - \frac{V + IR_{\mathrm{S}}}{R_{\mathrm{P}}} - I = 0. \tag{2.12}$$

Thus, given K measurements of I and V and $i \in \{1,...,K\}$, the calibration problem amounts to choosing $\widehat{\boldsymbol{\theta}}$ such that

$$\widehat{\boldsymbol{\theta}} = \ \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ \left[\log\left(\sum_{i=1}^{K} \phi_i(\boldsymbol{\theta})^2\right)\right], \tag{2.13}$$

where $\phi_i(\boldsymbol{\theta})$ is the lefthand side of (2.12) normalized to take inputs $\boldsymbol{\theta} \in [-1,1]^5$, and with $I \to I_i$ and $V \to V_i$. We wish to study the behavior of the function

$$\Phi(\boldsymbol{\theta}) = \log\left(\sum_{i=1}^{K} \phi_i(\boldsymbol{\theta})^2\right). \tag{2.14}$$

It is important to understand the behavior of $\Phi(\boldsymbol{\theta})$. The success of the minimization algorithm depends on whether there is a unique minimizer, the shape of $\Phi(\boldsymbol{\theta})$ around the minimum(s), and whether $\Phi(\boldsymbol{\theta})$ has local minima. The sensitivity analyses described in

[7]We note that the active subspace analysis is also sensitive to the choice of $\rho(\boldsymbol{\theta})$.

Section 2.2 and Section 2.3 will be applied in the following sections to help address these issues.

2.5 Sensitivity Analysis of the Single-Diode Model Objective Functions

In this section, we apply the Morris one-at-a-time method described in Section 2.2 and the active subspace estimation method described in Section 2.3 to the objective function in equation (2.14).

2.5.1 Morris One-At-a-Time

As described in Section 2.2, an estimate of the mean of the distribution of the i^{th} elementary effect, $|d_i| \sim G_i$, reveals information about the importance of θ_i. To implement the OAT method, we use the `MATLAB` function `morris_ma.m` and specify the parameters $p = 100$ and $\Delta = 0.5$. Table 2.2 shows the results of this analysis. We note that, for $\boldsymbol{\theta}_{\text{narrow}}$, this method suggests that θ_1 has the most influence on $\Phi(\boldsymbol{\theta})$. For $\boldsymbol{\theta}_{\text{wide}}$, we see that θ_1 has the largest effect, but that θ_2 and θ_3 also seem to have an effect on $\Phi(\boldsymbol{\theta})$. We now turn to the method of active subspaces to generalize these results from a subset dimension reduction analysis to a subspace dimension reduction analysis.

2.5.2 Active Subspaces

To estimate the active subspace, we draw $M = 10^5$ independent realizations of $\boldsymbol{\theta}$ from a uniform distribution on $[-1, 1]^5$ and compute $\widehat{\boldsymbol{C}} \approx \boldsymbol{C}$. Note that, in order to evaluate $\widehat{\boldsymbol{C}}$, we must evaluate the gradient of $\Phi(\boldsymbol{\theta})$.[8] To do so, we must shift and scale (and transformed by the exponential for I_{S}) to the original domain so that evaluations of the single-diode model makes sense. We repeat the active subspace analysis twice, shifting and scaling using both $\boldsymbol{\Theta}_{\text{narrow}}$ and $\boldsymbol{\Theta}_{\text{wide}}$.

[8]The gradient was calculated analytically, rather than approximated with, for example, finite differences.

Table 2.2: $\widehat{\mu}^*_{d_i}$ is an estimate of the mean of the distribution of the absolute value of the i^{th} elementary effect. A larger value of $\widehat{\mu}^*_{d_i}$ suggests that θ_i has more influence on the output $\Phi(\boldsymbol{\theta})$. We note that, for $\boldsymbol{\theta}_{\text{narrow}}$, this method suggests that θ_1 has the most influence on $\Phi(\boldsymbol{\theta})$. For $\boldsymbol{\theta}_{\text{wide}}$, we see that θ_1 has the largest effect, but that θ_2 and θ_3 also seem to have an effect on $\Phi(\boldsymbol{\theta})$.

parameter	$\widehat{\mu}^*_{d_i}$ $(\boldsymbol{\theta}_{\text{narrow}})$	$\widehat{\mu}^*_{d_i}$ $(\boldsymbol{\theta}_{\text{wide}})$
θ_1	1.50	4.81
θ_2	0.18	2.7
θ_3	6.47×10^{-2}	2.66
θ_4	3.59×10^{-2}	0.23
θ_5	5.19×10^{-4}	6.79×10^{-5}

Narrow Ranges When using $\boldsymbol{\Theta}_{\text{narrow}}$, we find some evidence for the existence of a two-dimensional active subspace. Such evidence is visualized, in part, by constructing *sufficient summary plots*—plots of proposed active variable(s), $\boldsymbol{y} = \boldsymbol{w}_1^T\boldsymbol{\theta}$, against corresponding values of $\Phi(\boldsymbol{\theta})$. Figure 2.1 shows one and two dimensional sufficient summary plots for realizations of $\Phi(\boldsymbol{\theta})$. Specifically, we choose $M = 10^5$ random draws of $\boldsymbol{\theta}$, use these draws to compute $\boldsymbol{y}$ and $\Phi(\boldsymbol{\theta})$, and then plot the pairs $(\boldsymbol{y}, \Phi(\boldsymbol{\theta}))$. In the one-dimensional sufficient summary plot, we see that a rough trend is present in the samples of $\Phi(\boldsymbol{\theta})$ when viewed from the angles defined by the eigenvector $\boldsymbol{w}_1$. Evidence for the existence of a two-dimensional active subspace is contained in the two-dimensional sufficient summary plot—the plot on the right in Figure 2.1. In this plot we see that larger values of the second active variable correspond to smaller values of $\Phi(\boldsymbol{\theta})$. Figure 2.3 shows a plot of a measure of separation

$$\gamma_k = \frac{\widehat{\lambda}_k - \widehat{\lambda}_{k+1}}{\widehat{\lambda}_1}. \tag{2.15}$$

Here, we see a large gap between the first two eigenvalues, but also a gap between the second and third eigenvalues.

Table 2.3 shows the 95% bootstrap confidence intervals for $\widehat{\lambda}_i$ described in Section 2.3. Figure 2.2 shows a plot of the components of the first (left) and second (right) eigenvectors of $\widehat{\boldsymbol{C}}$ with $M = 10^5$ samples of the gradient. The shaded region shows the 95% bootstrap

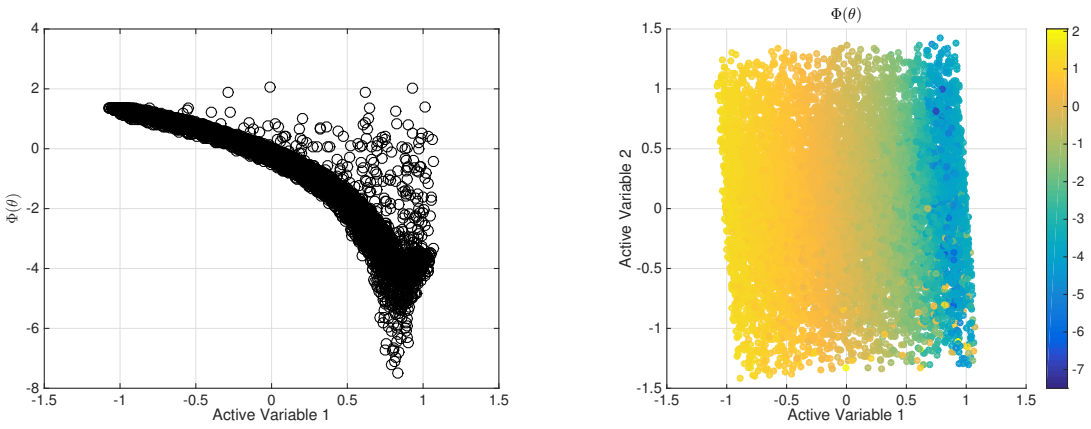


Figure 2.1: The plot on the left is a visualization of the $M = 10^5$ realizations of $\Phi(\boldsymbol{\theta})$ with inputs projected using the one-dimensional active subspace. The plot on the right is a visualization of the $M = 10^5$ realizations of $\Phi(\boldsymbol{\theta})$ with inputs projected using the two-dimensional active subspace. In both cases, the active subspace analysis was conducted using the **narrow** ranges in Table 2.1. Note that there is a rough relationship between the first active variable and $\Phi(\boldsymbol{\theta})$. However, the second active variable also plays an important role in changes in $\Phi(\boldsymbol{\theta})$. Higher values of the second active variable correspond to lower values of $\Phi(\boldsymbol{\theta})$. These plots provide evidence for the existence of a two-dimensional active subspace.

region for the eigenvector components. The narrow shaded regions suggest that there is low variability in the estimator $\widehat{\boldsymbol{C}}$. These values can be used as measures of sensitivity of the output of $\Phi(\boldsymbol{\theta})$ with respect to the input variables $\boldsymbol{\theta}$, whose names are given along the horizontal axis of the figures.

Table 2.3: 95% bootstrap confidence intervals for $\widehat{\lambda}_i$ calculated using $\boldsymbol{\Theta}_{\text{narrow}}$. The narrow ranges provide evidence that there is low variability in $\widehat{\lambda}_i$.

Eigenvalue	Lower 95% limit	Upper 95% limit
$\widehat{\lambda}_1$	19.24	22.54
$\widehat{\lambda}_2$	0.79	0.98
$\widehat{\lambda}_3$	0.17	0.36
$\widehat{\lambda}_4$	4.84×10^{-2}	8.24×10^{-2}
$\widehat{\lambda}_5$	0	0

Wide Ranges. When shifting and scaling with respect to $\boldsymbol{\Theta}_{\text{wide}}$, we find that the conclusions of the active subspace are similar. With these ranges, we find some evidence of a two-

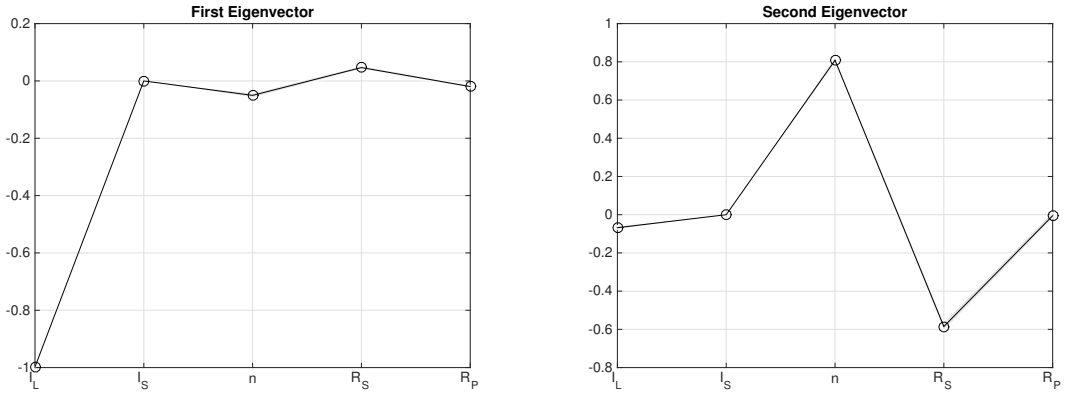


Figure 2.2: Here, the active subspace analysis was conducted using the **narrow** ranges in Table 2.1. The plot on the left shows the components of the first eigenvector of $\widehat{\boldsymbol{C}}$. The plot on the right shows the components of the second eigenvector of $\widehat{\boldsymbol{C}}$. Each were computed with $M = 10^5$ samples of the gradient. In both plots, the shaded region represents the 95% bootstrap region for the eigenvector components. The narrow shaded regions suggest that there is low variability in the estimator $\widehat{\boldsymbol{C}}$. These values can be used as measures of sensitivity of the output of $\Phi(\boldsymbol{\theta})$ with respect to the input variables $\boldsymbol{\theta}$, whose names are given along the horizontal axis of the figures.

dimensional active subspace. Again, this evidence comes in the form of sufficient summary plots. Figure 2.4 shows one and two dimensional sufficient summary plots for $M = 10^5$ realizations of $\Phi(\boldsymbol{\theta})$. In the one-dimensional sufficient summary plot, we see that a trend is present in the samples of $\Phi(\boldsymbol{\theta})$ when viewed from the angles defined by the eigenvector $\boldsymbol{w}_1$. However, in the plot on the right in Figure 2.4, we see some change in $\Phi(\boldsymbol{\theta})$ as a function of the second active variable (as well as the first active variable). This points to the existence of a two-dimensional active subspace.

Table 2.4 shows the 95% bootstrap confidence intervals for $\widehat{\lambda}_i$ described in Section 2.3. The order-of-magnitude gaps between the ranges provide confidence in the dominance of the active subspace. Figure 2.5 shows a plot of the components of the first (left) and second (right) eigenvectors of $\widehat{\boldsymbol{C}}$ with $M = 10^5$ samples of the gradient. The shaded region shows the 95% bootstrap region for the eigenvector components. The narrow shaded regions suggest that there is low variability in the estimator $\widehat{\boldsymbol{C}}$. These values can be used as measures of sensitivity of the output of $\Phi(\boldsymbol{\theta})$ with respect to the input variables $\boldsymbol{\theta}$, whose names are

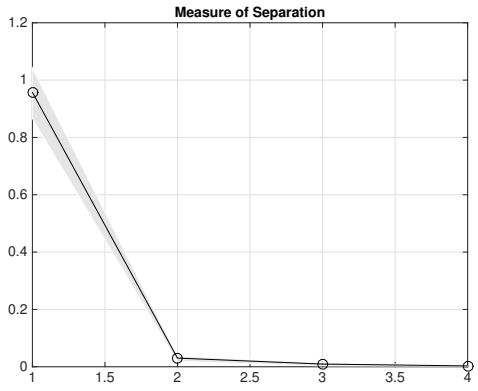


Figure 2.3: A plot of a measure separation between eigenvalues. Along the horizontal axis, '1' corresponds to the gap between the first and second eigenvalue, '2' corresponds to a gap between the second and third eigenvalue, and so on.

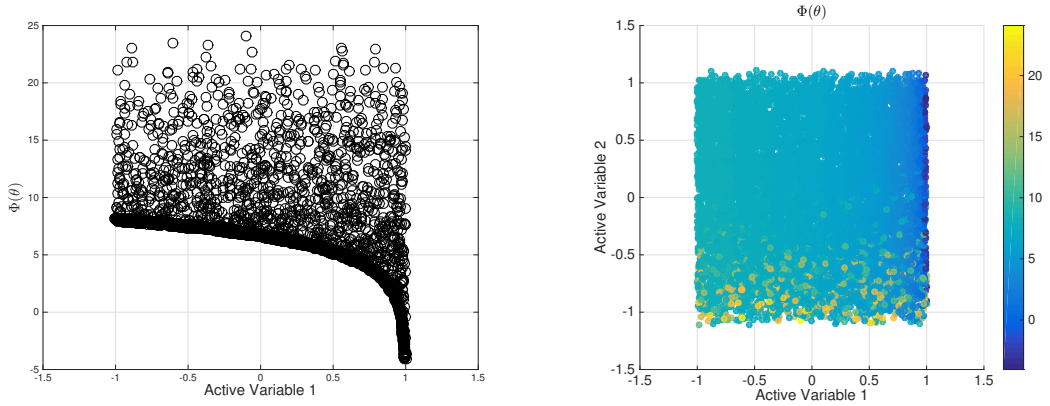


Figure 2.4: The plot on the left is a visualization of the $M = 10^5$ realizations of $\Phi(\boldsymbol{\theta})$ with inputs projected using the one-dimensional active subspace. The plot on the right is a visualization of the $M = 10^5$ realizations of $\Phi(\boldsymbol{\theta})$ with inputs projected using the two-dimensional active subspace. In both cases, the active subspace analysis was conducted using the **wide** ranges in Table 2.1.

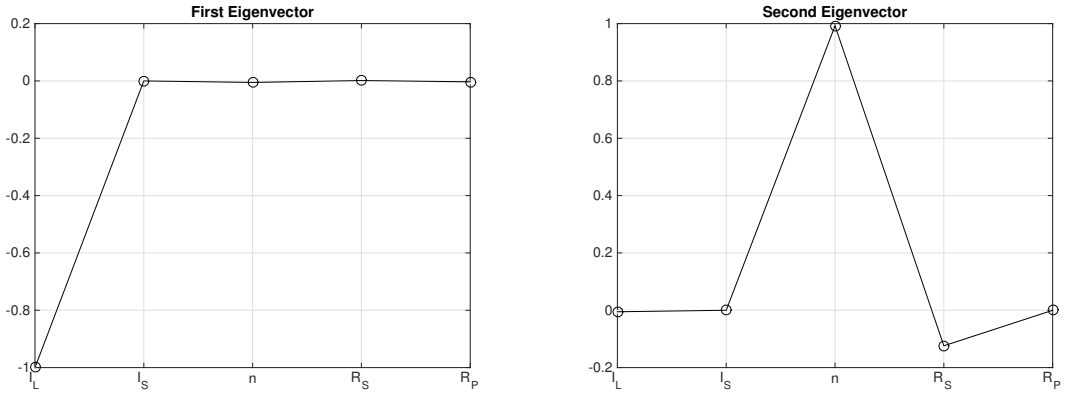


Figure 2.5: Here, the active subspace analysis was conducted using the **wide** ranges in Table 2.1. The plot on the left shows the components of the first eigenvector of $\widehat{\boldsymbol{C}}$. The plot on the right shows the components of the second eigenvector of $\widehat{\boldsymbol{C}}$. Each were computed with $M = 10^5$ samples of the gradient. In both cases, the variation in the estimators for the components of the eigenvectors is negligible.

given along the horizontal axis of the figures.

Table 2.4: 95% bootstrap confidence intervals for $\widehat{\lambda}_i$ calculated using $\boldsymbol{\Theta}_{\text{wide}}$.

Eigenvalue	Lower 95% limit	Upper 95% limit
$\widehat{\lambda}_1$	310.00	475.81
$\widehat{\lambda}_2$	35.90	40.33
$\widehat{\lambda}_3$	0.13	1.09
$\widehat{\lambda}_4$	1.21×10^{-3}	0.14
$\widehat{\lambda}_5$	0	0

2.6 Discussion and Conclusions

In this paper, we use two sensitivity measures to show that a popular objective function constructed from the PV single-diode mode has low-dimensional structure that can be exploited when performing calibration. In particular, evidence for the existence of an active subspace in the objective function $\Phi(\boldsymbol{\theta})$ can be exploited in a number of ways, many of which are current topics of active subspace research. Here, we outline one possible way that one could exploit the active subspace for more efficient optimization. We focus on the analysis using $\boldsymbol{\Theta}_{\text{narrow}}$ but note that the obvious extension applies to $\boldsymbol{\Theta}_{\text{wide}}$.

Recall that the plot on the right in Figure 2.1 suggests the existence of a two-dimensional active subspace. One could use this fact to find a reasonable starting value for the nonlinear optimization routine used in calibrating the PV single-diode model. Such a starting value can be important because bad starting values often mislead practitioners about the true global minimum of their objective function—i.e., a bad starting value can cause an optimization routine to settle into a local minimum. To use active subspaces to find a starting value of an optimization routine, we consider the following prodecure:

1. Find an approximation of $\Phi(\boldsymbol{\theta})$ using the active subspace; perhaps by using some nicely behaved surface:

$$\Phi(\boldsymbol{\theta}) \approx g(\underbrace{\boldsymbol{W}_1^T\boldsymbol{\theta}}_{\boldsymbol{y}}) = g(\boldsymbol{y}). \tag{2.16}$$

2. Use g to find the value $\boldsymbol{y}^*$ that (approximately) minimizes $\Phi(\boldsymbol{\theta})$.

3. Find a value of $\boldsymbol{\theta}$ that corresponds to $\boldsymbol{y}^*$. That is, solve

$$\boldsymbol{w}_{1_1}\widehat{\theta}_1 + \boldsymbol{w}_{1_2}\widehat{\theta}_2 + \boldsymbol{w}_{1_3}\widehat{\theta}_3 + \boldsymbol{w}_{1_4}\widehat{\theta}_4 + \boldsymbol{w}_{1_5}\widehat{\theta}_5 = \boldsymbol{y}^*, \tag{2.17}$$
$$\boldsymbol{w}_{2_1}\widehat{\theta}_1 + \boldsymbol{w}_{2_2}\widehat{\theta}_2 + \boldsymbol{w}_{2_3}\widehat{\theta}_3 + \boldsymbol{w}_{2_4}\widehat{\theta}_4 + \boldsymbol{w}_{2_5}\widehat{\theta}_5 = \boldsymbol{y}^*, \tag{2.18}$$

   such that $\widehat{\theta}_1, \widehat{\theta}_2 ..., \widehat{\theta}_5 \in [-1, 1]$, for $\boldsymbol{\theta}_0$.

4. Use $\boldsymbol{\theta}_0$ as a starting value in the optimization routine using $\Phi(\boldsymbol{\theta})$ as the objective function.

CHAPTER 3

ON THE ESTIMABILITY OF THE PV SINGLE-DIODE MODEL PARAMETERS

Sections 3.1 through 3.8 comprise of a paper to be published in the Conference on Data Analysis special edition of *Statistical Analysis and Data Mining*

Brian Zaharatos[9], Mark Campanelli[10], Luis Tenorio[11]

Abstract

The single-diode model is a widely used representation of the electrical performance of a photovoltaic (PV) device. This model relates the PV device's terminal current and voltage at a given irradiance and temperature. In order to obtain reasonable estimates of single-diode model parameters given noisy data, one ought to be able to characterize the estimability of the model parameters. Here, we look to an established method called data cloning to check for evidence of inestimability.

3.1 Introduction

Characterizing the performance of photovoltaic (PV) devices is important for the reduction of greenhouse gas emissions and for the growth of the PV industry. The performance of a PV device (a cell, module, or array) is determined by measuring its current-voltage (I-V) characteristics and using these measurements to estimate *key performance parameters*: short circuit current (I_{SC}), open circuit voltage (V_{OC}), and maximum power output (P_{max}). One method used to characterize PV performance using I-V data involves the development of a mathematical model of the PV device. In this case, one distinguishes between the *model parameters* of the underlying mathematical model and the resulting key performance parameters that are computed using the model.

[9]Primary Researcher and Author
[10]National Renewable Energy Lab
[11]Department of Applied Mathematics and Statistics, Colorado School of Mines

The choice of a PV performance model is nontrivial. A performance model should (1) be an accurate representation of a wide range of PV devices; (2) be simple enough to allow for tractable estimation of the model and performance parameters; (3) take into account some fluctuations in measurement conditions (e.g., irradiance and temperature); (4) allow for statistical modeling of data measured with noise; and, finally, (5) the model should have desirable statistical properties, such as parameter *identifiability*—i.e., two distinct values in the parameter space should not yield the same likelihood—and parameter *estimability*—i.e., estimation techniques, such as maximum likelihood estimation, should yield unique answers.

In this paper, we develop a model that is believed to satisfy (1)—(4) and assess whether it meets the estimability condition in (5). In particular, in Section 3.2 we focus on a formulation of the single-diode model that allows for modeling irradiance values that deviate slightly around the nominal irradiance value of 1000W/m^2, with temperature fixed at 25°C. In Section 3.4, we describe a statistical formulation of the single-diode model that quantifies uncertainty in the model parameters due to noise in the irradiance data. In Section 3.5 we explore the identifiability and estimability of the model and key performance parameters. Finally, in cases where there is no evidence of non-identifiability or inestimability, we estimate all parameters and quantify the uncertainty using maximum likelihood methods.

3.2 The Single Diode Model

The PV literature describes several models for PV performance, including double-and triple-diode models. However, for a wide range of PV devices, the single-diode model allows for a good compromise between accuracy and simplicity [27]. The following single-diode lumped-parameter equivalent-circuit model—similar to models used in [18] and [17]—adequately describes the I-V performance of many series-wired photovoltaic (PV) devices:

$$I = I_{\text{L}} - I_{\text{S}}\left(e^{\frac{V+IR_{\text{S}}}{N_{\text{S}}nV_{\text{th}}(T)}} - 1\right) - \frac{V + IR_{\text{S}}}{R_{\text{P}}}, \tag{3.1}$$

where the thermal voltage as a function of temperature is $V_{\text{th}}(T) := k_{\text{B}}\,T/q$, and

I device terminal current (A)

V device terminal voltage (V)

$I_{\mathbf{L}}$ photocurrent (A)

$I_{\mathbf{S}}$ diode reverse saturation current of device (A)

$N_{\mathbf{S}}$ number of cells connected in series (unit-less)

n ideality factor of device (unit-less)

$R_{\mathbf{S}}$ series resistance of device (Ω)

$R_{\mathbf{P}}$ parallel (shunt) resistance of device (Ω)

$k_{\mathbf{B}}$ Boltzmann constant ($1.3806504 \times 10^{-23}$ J/K)

T junction temperature of device (K)

q electron charge ($1.602176487 \times 10^{19}$ C).

In general, I_{L}, I_{S}, n, R_{S}, and R_{P}, can depend on temperature, irradiance, or terminal voltage, but here, we assume that these parameters depend on irradiance and temperature only. Typically in the PV literature, auxiliary equations are given that describe how each device parameter varies with temperature and irradiance from the reference value defined at standard reporting conditions (SRC) (e.g., [28]). Measurements in a calibration laboratory can usually be made with a fixed temperature, T, which is kept as close as possible to the SRC temperature of $T_0 = 25°\mathrm{C}$. Furthermore, solar simulators are set to have a nominal irradiance as close as possible to the SRC irradiance of 1000 W/m^2 under the standard spectrum defined in ASTM G173-03 [3]. The variability in the irradiance is typically less than a few percent. Consequently, we assume that the dependence of $I_{\mathrm{S}}, n, R_{\mathrm{S}}$, and R_{P} on temperature or irradiance can be appropriately neglected. Thus, the auxiliary equations for these parameters are given by $I_{\mathrm{S}} = I_{\mathrm{S}_0}$, $n = n_0$, $R_{\mathrm{S}} = R_{\mathrm{S}_0}$, and $R_{\mathrm{P}} = R_{\mathrm{P}_0}$ where the $-_0$ notation refers to the value at SRC. However, the dependence of I_{L} on irradiance is not negligible. Thus, I_{L} is [29]

$$I_{\mathrm{L}} := E\, I_{\mathrm{SC}_0} + I_{\mathrm{S}_0}\left(e^{\frac{E\, I_{\mathrm{SC}_0} R_{\mathrm{S}_0}}{N_{\mathrm{S}} n_0 V_{\mathrm{th}}(T_0)}} - 1\right) + \frac{E\, I_{\mathrm{SC}_0} R_{\mathrm{S}_0}}{R_{\mathrm{P}_0}}. \tag{3.2}$$

In equation (5.10), the *effective irradiance ratio* E is a unitless measure of the "effective" irradiance illuminating the device relative to the SRC irradiance. With $T = T_0$, E is defined as the ratio of the device's short-circuit current at the prevailing irradiance, I_{SC}, to the device's short-circuit current at SRC, I_{SC_0}:

$$E = \frac{I_{\mathrm{SC}}}{I_{\mathrm{SC}_0}}. \tag{3.3}$$

Because of temporal variability in many illumination sources, E is typically measured simultaneously with each measured I-V point using a separate calibrated reference device. From [30], it follows that

$$\frac{I_{\mathrm{SC}}}{I_{\mathrm{SC}_0}} = \frac{M I_{\mathrm{SC,R}}}{I_{\mathrm{SC,R}_0}}, \tag{3.4}$$

where $I_{\mathrm{SC,R}}$ is the short-circuit current of a reference device at the prevailing irradiance and $I_{\mathrm{SC,R}_0}$ is the short-circuit current of the reference device at SRC (typically a calibration value). In this work, we do not consider additional uncertainty due to the spectral correction factor M[12] or $I_{\mathrm{SC,R}_0}$. Measurements of E have random noise—which arises from the term $I_{\mathrm{SC,R}}$—whose characteristic is particular to the placement of the reference device and to the spatio-temporal uniformity of irradiance of the measurement testbed.

Auxiliary equation (5.10) describes I_{L} as a function of E and the model parameters. This equation is derived by setting $V = 0$ and $I = I_{\mathrm{SC}}$ in (5.9) and using the definition of E to solve for I_{L}. The model parameters, which correspond to reference values at SRC, are denoted by the vector $\boldsymbol{\theta} = (I_{\mathrm{SC}_0}, I_{\mathrm{S}_0}, n_0, R_{\mathrm{S}_0}, R_{\mathrm{P}_0})$.

3.3 Estimating the Parameters of the Single-Diode Model

There are many possible approaches to estimating the parameters of the single-diode model. In [6] and [7], the authors provide an overview of estimation techniques found in the PV literature. Broadly, these techniques fall into two categories: approaches that use limited

[12]The computation of M requires additional information about the quantum efficiencies of the devices and the spectral irradiance of the light source.

data, for example, approximations of I_{SC} and V_{OC} at SRC (such approximations are often used because they are readily given by the PV device's manufacturer) and approaches that use data from a full range I-V curve. In order to have a sufficiently large set of equations from which to estimate the parameters, approaches in the former category (e.g., [8], [9], [10]) use information contained on a PV manufacturer's datasheet. Approaches in the latter category (e.g., [11], [12], [13]) have a system of nonlinear equations as large as the sample size of the I-V data. Both approaches use either deterministic iterative root finding or optimization methods—e.g., using Newton's method [9], genetic algorithms [14], particle swarm optimization [15], or sequential techniques [6]—to solve for the parameters.

In general, there is no standard parameter estimation method. The authors of [6] write that "without a standard method for estimating model parameters, different modelers may obtain different model parameter values even when the parameter values are derived from a common set of measurements." In one sense, for a particular data set, the non-uniqueness of a parameter estimate to be expected; for example, for a fixed sample size, well-defined maximum likelihood and method of moments estimators can yield different estimates. However, it is desirable to know whether any particular estimator (e.g., the maximum likelihood estimator (MLE)) yields a unique answer. Non-uniqueness of an estimator may be due to the underlying structure of the model (what we call *non-identifiability* below) or may be an artifact of the estimation method; for example, MLEs are not guaranteed to be unique.

In the remaining sections, we develop a statistical model for the I-V-E data and explore potential sources of non-uniqueness. Specifically, we study the uniqueness of the MLE for $\boldsymbol{\theta}$ and the key performance parameters, P_{max_0} and V_{OC_0}; global uniqueness of the MLE is desirable because it (a) implies that the model is free from structural issues and (b) it may circumvent the worries in [6] by providing a standard method for single-diode model parameter estimation within the PV community. Further, using a statistical model allows for a natural way to quantify the uncertainty in the model parameters due to measurement error.

3.4 Statistical Model

For $i \in \{1, ..., 5\}$, we define the parameter space of $\boldsymbol{\theta}$ as $\boldsymbol{\Theta} = \{(\theta_1, \theta_2, ..., \theta_5) |\ \theta_i \in \boldsymbol{\Theta}_i\}$, where $\boldsymbol{\Theta}_1 = [1/12, 12]$, $\boldsymbol{\Theta}_2 = [10\text{e}-13, 10\text{e}-6]$, $\boldsymbol{\Theta}_3 = [1, 2]$, $\boldsymbol{\Theta}_4 = [0, 100]$, $\boldsymbol{\Theta}_5 = [1, 5\text{e}4]$. We chose $\boldsymbol{\Theta}$ to include all devices in the California Energy Commission's database [17]. Given $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^5$, (I, V, E) triplets are thought to be a solution surface to the single-diode model forward problem defined by (5.9)-(5.10). This surface implicitly defines three functions: $I = g_I(V, E; \boldsymbol{\theta})$, $V = g_V(I, E; \boldsymbol{\theta})$, and $E = g_E(I, V; \boldsymbol{\theta})$. The Lambert W-function can be employed for the efficient computation of any of these three functions [31]. To solve the inverse problem—estimating $\boldsymbol{\theta}$ given discrete points on the (I, V, E) surface—we first note that in many measurement contexts, the measurement noise for I and V is small as compared to the measurement noise in E. Consequently, here we specify a noise model for E and assume that measurements of I and V are exact. Because E is always positive, we assume that measurements E_i are log-normally distributed with homoscedastic noise. That is, for N points along an I-V curve, and for $i \in \{1, ..., N\}$, the observable E_i is modeled by the random variable[13]

$$\mathcal{E}_i \sim \text{lognormal}\Big(\log(g_E(I_i, V_i; \boldsymbol{\theta})), \sigma_E^2\Big).$$

For convenience, we work in the log space of $\boldsymbol{E}$. The likelihood function for $\log(\boldsymbol{E})$ is

$$L(\boldsymbol{\theta} |\ \log(\boldsymbol{E})) = \left(2\,\pi\,\sigma_E^2\right)^{-N/2} \exp\left(-\frac{\sum_{i=1}^{N} \left(\log(E_i) - \log(g_E(I_i, V_i; \boldsymbol{\theta}))\right)^2}{2\sigma_E^2}\right). \tag{3.5}$$

We denote an MLE of $\boldsymbol{\theta}$ as $\widehat{\boldsymbol{\theta}}_{\text{ML}}$ and observe that the MLE depends on the assumed statistical model. As noted in Section 3.3, the goal of this work is to study whether $\boldsymbol{\theta}$ is estimable and identifiable.

[13] Note that, since the variance encountered in the data is known to be small, the mean and median of the distribution will be very close.

3.5 Estimability and Identifiability

In general, given data $\boldsymbol{X}$ with a likelihood $L(\boldsymbol{\theta}|\boldsymbol{x})$, the model parameter $\boldsymbol{\theta}$ is said to be *identifiable* if, for all $\boldsymbol{\theta}_2, \boldsymbol{\theta}_1 \in \boldsymbol{\Theta}$, $L(\boldsymbol{\theta}_2|\boldsymbol{x}) = L(\boldsymbol{\theta}_1|\boldsymbol{x})$ for almost all $\boldsymbol{x}$ implies that $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_1$. Otherwise, $\boldsymbol{\theta}$ is *non-identifiable*. Consider the following examples. For each example, we let $i \in \{1, 2, ..., n\}$.

1. Let $X_i \sim \mathcal{N}(\theta, 1)$. The unknown parameter θ is identifiable.

2. Let $X_i \sim \mathcal{U}(\theta - 1, \theta + 1)$. The unknown parameter θ is identifiable.

3. Let $Y_i \sim \mathcal{N}(f(x_i; \theta_1, \theta_2), 1)$ where $f(x_i; \theta_1, \theta_2) = \exp(\theta_1 \, \theta_2 \, x_i)$. Neither θ_1 nor θ_2 is identifiable; however, the product $\theta_1 \, \theta_2$ is identifiable.

4. Let $X_i \sim \mathcal{N}(|\theta|, 1)$. The unknown parameter θ is not identifiable because for $\theta_1 = -\theta_2$, $L(\theta_2, \sigma^2|\boldsymbol{x}) = L(\theta_1, \sigma^2|\boldsymbol{x})$.

Parameter identifiability is important in practice because non-identifiable models admit multiple parameter sets that explain the data equally well; in the absence of independent a priori knowledge, there is no way of knowing which parameter set generated the data.

Model parameters $\boldsymbol{\theta}$ are said to be *estimable* if an MLE exists and

$$\{\boldsymbol{\theta} \in \boldsymbol{\Theta} : L(\boldsymbol{\theta} \mid \boldsymbol{x}) = L(\widehat{\boldsymbol{\theta}}_{\mathrm{ML}} \mid \boldsymbol{x})\} = \{\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}\}.$$

That is, $\boldsymbol{\theta}$ is estimable if the MLE is unique. Otherwise, $\boldsymbol{\theta}$ is said to be *inestimable*. We consider, again, Examples 1–4:

1. Let $X_i \sim \mathcal{N}(\theta, 1)$. The unknown parameter θ is estimable.

2. Let $X_i \sim \mathcal{U}(\theta - 1, \theta + 1)$. The unknown parameter θ is inestimable because any value in the interval $[\max\{\boldsymbol{x}\} - 1, \min\{\boldsymbol{x}\} + 1]$ maximizes the model's likelihood function.

3. Let $Y_i \sim \mathcal{N}(f(x_i; \theta_1, \theta_2), 1)$ where $f(x_i; \theta_1, \theta_2) = \exp(\theta_1 \, \theta_2 \, x_i)$. Neither θ_1 nor θ_2 is estimable; however, the product $\theta = \theta_1 \, \theta_2$ is estimable.

4. Let $X_i \sim \mathcal{N}(|\theta|, 1)$. The unknown parameter θ is inestimable because both $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$ and $-\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$ maximize the likelihood function.

Parameter estimability is important because inestimable models may yield disagreements between maximum likelihood estimates computed by different optimization methods using the same data. The method of data cloning—developed in [32], described in Section 3.6, and implemented in Section 3.7—has been used to diagnose inestimability.

The examples above suggest a relationship between identifiability and estimability. The model in Example 1 is an instance of the general claim that estimability implies local identifiability [33].[14] This claim is true because identifiability is typically assumed in the definition of the maximum likelihood estimator. If a parameter is estimable, then it has a (unique) MLE, and thus must be locally identifiable by assumption. The model in Example 2 shows that a model may be identifiable but inestimable. However, evidence of inestimability could be an indication of underlying (local) non-identifiability.

In the remainder of this paper, we study the estimability of $\boldsymbol{\theta}$ in equation (4.4). Evidence that $\boldsymbol{\theta}$ is estimable is also evidence that $\boldsymbol{\theta}$ is locally identifiable; evidence that $\boldsymbol{\theta}$ is inestimable may suggest that $\boldsymbol{\theta}$ is non-identifiable, but we note that some methods may be used to study identifiability more directly (e.g., [34]).

3.6 Data Cloning—Theory

Data cloning was first developed in [16] in the context of maximum likelihood estimation for hierarchical models. Then, it was shown in [32] that data cloning can be used to diagnose inestimability. Data cloning uses tools from the Bayesian framework, such as Markov chain Monte Carlo (MCMC) methods, but it does not require any adherence to the assumptions of Bayesian inference (e.g., a Bayesian interpretation of probability or the modeling of parameters as random variables). To implement data cloning, one develops a Bayesian model for the problem at hand and uses MCMC to compute MLEs or assess estimability. Specifically,

[14] A parameter $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$ is *locally* identifiable if there exists a neighborhood $\mathbf{N}$ around $\boldsymbol{\theta}^*$ such that, for all $\boldsymbol{\theta} \in \mathbf{N}$, $L(\boldsymbol{\theta}^*|\boldsymbol{x}) = L(\boldsymbol{\theta}|\boldsymbol{x})$ for almost all $\boldsymbol{x}$ implies that $\boldsymbol{\theta}^* = \boldsymbol{\theta}$ [34].

the motivation for data cloning is a thought experiment: suppose that, by coincidence, one observed k independent identical samples of size n:

$$\boldsymbol{x}_{n_k} = \underbrace{(x_1, ..., x_n, ..., x_1, ..., x_n)}_{k \text{ independent repeats}}.$$

The resulting likelihood function is $L_k(\boldsymbol{\theta}|\boldsymbol{x}_{n_k}) = [L(\boldsymbol{\theta}|\boldsymbol{x})]^k$. For $\pi(\boldsymbol{\theta})$, any proper prior density for $\boldsymbol{\theta}$, the k^{th} *posterior distribution* is defined as

$$\pi_k(\boldsymbol{\theta}|\boldsymbol{x}) \propto L_k(\boldsymbol{\theta}|\boldsymbol{x}_{n_k})\pi(\boldsymbol{\theta}).$$

Properties of the k^{th} posterior distribution are used to diagnose estimability. If $\boldsymbol{\theta}$ is estimable then for sufficiently large k, $\pi_k(\boldsymbol{\theta}|\boldsymbol{x})$ is nearly degenerate around $\widehat{\boldsymbol{\theta}}_{\text{ML}}$ with covariance matrix approximately equal to $(1/k)\,\mathcal{I}(\widehat{\boldsymbol{\theta}}_{\text{ML}})^{-1}$ [32], where $\mathcal{I}(\widehat{\boldsymbol{\theta}}_{\text{ML}})$ is the *Fisher information matrix* evaluated at $\widehat{\boldsymbol{\theta}}_{\text{ML}}$, and is defined as

$$\mathcal{I}(\widehat{\boldsymbol{\theta}}_{\text{ML}}) = \mathbb{E}\left[-\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log(L(\boldsymbol{\theta}|\boldsymbol{x})) \right] \Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{\text{ML}}}. \tag{3.6}$$

So, as $k \to \infty$, if the variance of the k^{th} marginal posterior $\pi_k(\theta_i|\boldsymbol{x})$ does not converge to zero, then θ_i is inestimable, which, in turn, may be and indication of non-identifiability.

3.7 Numerical Experiments

In this section we use ten synthetic (I,V,E) data sets and one real data set to gain insight into local estimability of the statistical model described in Section 3.4. First, in Section 3.7.1, we consider the synthetic data sets, each of which has $N = 50$ points. Then, in Section 3.7.2, we repeat the analysis on a real data set from a 4cm^2 mono-crystalline silicon (mono-Si) PV cell. Note that for computational reasons, when running simulations, we work in log space for θ_2 and use the domains in Section 3.4 to normalize $\boldsymbol{\theta}$ to the hypercube $[-1, 1]^5$. The normalized parameters, $\boldsymbol{\theta}^*$, are defined as:

$$\theta_i^* = 2\left(\frac{\theta_i - A_i}{B_i - A_i}\right) - 1, \quad i \in \{1, 3, 4, 5\},$$
$$\theta_2^* = 2\left(\frac{\log(\theta_2) - \log(A_2)}{\log(B_2) - \log(A_2)}\right) - 1,$$

where A_i and B_i are the lower and upper bounds, respectively, of the domain of θ_i, $i \in \{1, ..., 5\}$.

Table 3.1: A table showing the true values of the parameters for each of the ten synthetic data sets used in Section 3.7.1.

	$\theta_1 = I_{\mathrm{SC}_0}$,	$\theta_2 = I_{\mathrm{S}_0}$,	$\theta_3 = n_0$,	$\theta_4 = R_{\mathrm{S}_0}$,	$\theta_5 = R_{\mathrm{P}_0}$
$\boldsymbol{\theta}_1$	0.11979	2.2e -8	1.50	0.3336	187.5
$\boldsymbol{\theta}_2$	0.15000	1.00e -10	1.30	0.5000	300.0
$\boldsymbol{\theta}_3$	0.50000	1.94e-09	1.023	0.09400	15.7
$\boldsymbol{\theta}_4$	1.22000	1.43e -12	1.24	0.1000	1.0e4
$\boldsymbol{\theta}_5$	9.00000	1.93e -10	1.00	0.01000	430.0
$\boldsymbol{\theta}_6$	0.087000	1e -7	1.60	0.4000	96.0
$\boldsymbol{\theta}_7$	2.60760	3.33e-7	1.80	0.08470	1.7e4
$\boldsymbol{\theta}_8$	4.17600	1.16e-09	1.19	0.07160	250.3
$\boldsymbol{\theta}_9$	0.68340	1.61e-10	1.40	0.2130	83.3
$\boldsymbol{\theta}_{10}$	3.22100	3.42e-7	1.75	0.1000	1.0e4

3.7.1 Synthetic Data

Table 3.2 shows the parameters used to generate the synthetic data sets. These parameters are chosen to be typical of a variety of cells measured at a PV characterization laboratory. The synthetic data sets were generated in the following way:

1. Fix $\boldsymbol{\theta} = \boldsymbol{\theta}_i$.

2. Find appropriate (I, V) points corresponding to $\boldsymbol{\theta}$ at a light level 1000 W/m^2, by:

   a. Setting $\boldsymbol{E}_{\mathrm{true}} = \mathbf{1}_{50\times 1}$;

   b. Calculating V_{OC} (V when $I = 0$) at E_{true}: $V_{\mathrm{OC}} = g_V(I = 0, E_{\mathrm{true}} = 1; \boldsymbol{\theta})$;

   c. Creating an evenly spaced grid of $N = 50$ voltages from 0 to V_{OC};

d. Using the grid of voltages, $\boldsymbol{V}_{\text{syn}}$, the light level $\boldsymbol{E}_{\text{true}}$ to solve for $\boldsymbol{I}_{\text{syn}}$:

$$\boldsymbol{I}_{\text{syn}} \;=\; g_I(\boldsymbol{V}_{\text{syn}}, \boldsymbol{E}_{\text{true}}; \boldsymbol{\theta});$$

3. Model the measurement noise using a log-normal distribution:

$$\log\big(E_{\text{syn}_i}\big) \;=\; \log\left(g_E(I_{\text{syn}_i}, V_{\text{syn}_i}; \boldsymbol{\theta})\right) \;+\; \varepsilon_i, \text{ where } \varepsilon_i \sim \mathcal{N}(0, 10^{-6}).$$

We note that this procedure leads to a well-defined forward model, i.e., $E_{\text{true}_i} = g_E(I_{\text{syn}_i}, V_{\text{syn}_i}; \boldsymbol{\theta})$ for $i \in \{1, ..., N\}$. Figure 3.1 shows plots of these synthetic data sets.

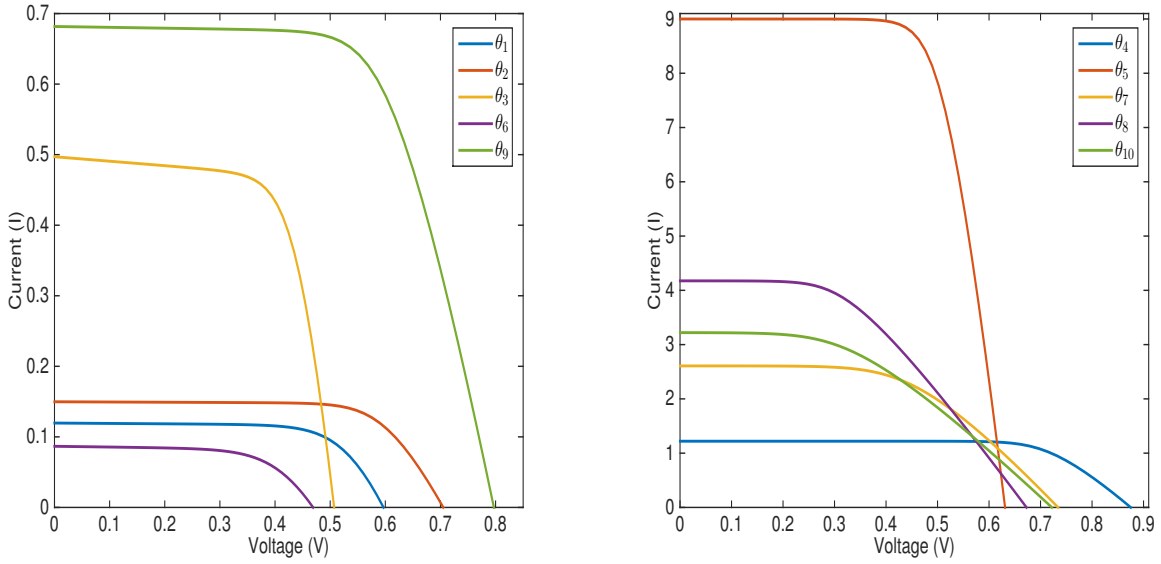


Figure 3.1: Synthetic data sets generated from the single-diode model with noise in $\boldsymbol{E} \;=\; \mathbf{1}_{50\times 1}$. The parameters used to generate these curves are given in Table 3.1. For each data set, $\sigma_E^2 = 10^{-6}$.

For each synthetic data set, we perform data cloning for $k \in \{1, 10, 20, 30, ..., 70\}$ to test for inestimability. For the k^{th} iteration, we use the likelihood function defined in equation (4.4) with k repeats of $(I_{\text{syn}_i}, V_{\text{syn}_i}, E_{\text{syn}_i})$. Note that equation (4.4) takes σ_E^2 as an input. In most realistic settings, practitioners won't know σ_E^2. Thus, in MCMC evaluations of the likelihood, we use the estimator

$$\widehat{\sigma}_E^2 = \frac{1}{N-5}\sum_{i=1}^{N}\left(\log(E_i) - \log(g_E(I_i, V_i; \widehat{\boldsymbol{\theta}})\right)^2,$$

where $\widehat{\boldsymbol{\theta}}$ is calculated in `MATLAB` using `lsqnonlin`'s trust-region-reflective algorithm. We repeat the data cloning diagnostic for three distinct prior distributions for normalized parameters θ_i^*: a uniform distribution on $[-1, 1]$; a normal distribution centered at zero and truncated over $[-1, 1]$; and a normal distribution centered at $\widehat{\boldsymbol{\theta}}$ and truncated over $[-1, 1]$. Choosing different priors strengthens the evidence found in the data cloning diagnostics [35]. All Markov chains used in data cloning were constructed using the Delayed Rejection Adaptive Metropolis (DRAM) algorithm in the MCMC toolbox for MATLAB [36].

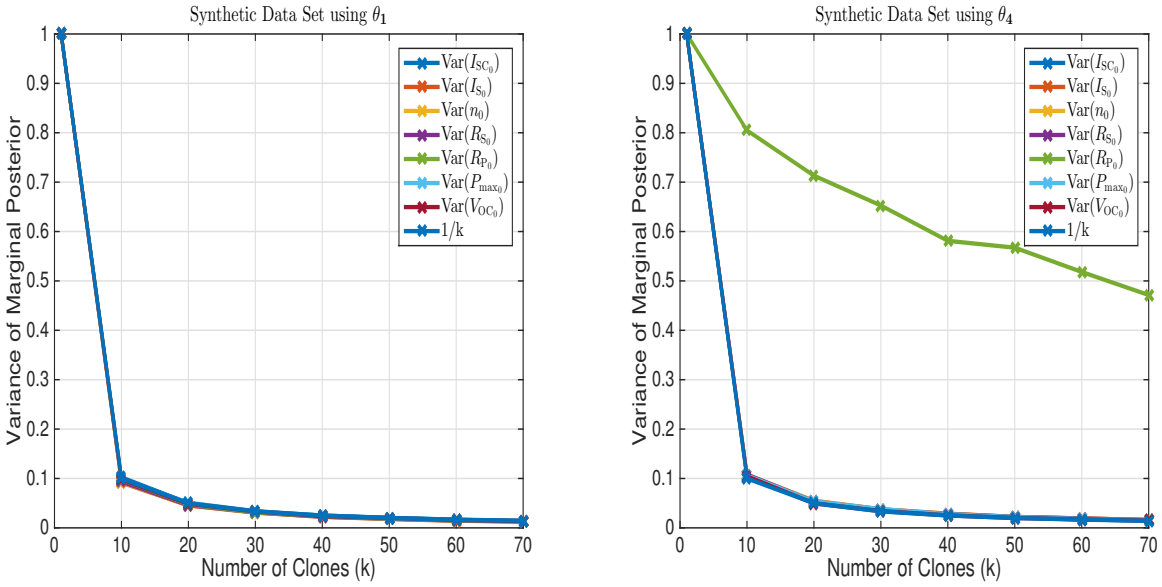


Figure 3.2: The plot on the left shows shows the variance of the marginal posterior distribution for $\boldsymbol{\theta}_1$ and the corresponding key performance parameters $P_{\mathrm{MAX}_0} = P_{\mathrm{MAX}_0}(\boldsymbol{\theta})$ and $V_{\mathrm{OC}_0} = V_{\mathrm{OC}_0}(\boldsymbol{\theta})$. The variances decay at a rate of $1/k$ giving no evidence of inestimability. The plot on the right shows the variance of the marginal posterior distribution for $\boldsymbol{\theta}_4$ and the corresponding key performance parameters $P_{\mathrm{MAX}_0} = P_{\mathrm{MAX}_0}(\boldsymbol{\theta})$ and $V_{\mathrm{OC}_0} = V_{\mathrm{OC}_0}(\boldsymbol{\theta})$. We see that the variance of $\theta_5 = R_{\mathrm{P}_0}$ does not decay at a rate of $1/k$, giving evidence of inestimability.

Estimability and Identifiability Results. The results of the data cloning diagnostics can be categorized into two classes. The first class consists of synthetic data sets for which there is no evidence of inestimability for any of the model parameters. For data sets in this class—those generated by $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, $\boldsymbol{\theta}_3$, $\boldsymbol{\theta}_6$, $\boldsymbol{\theta}_9$—the variances of the marginal posterior for each

model parameter decreases at a rate of $1/k$. Figure 4.1 (left) shows a standardized variance plot for $\boldsymbol{\theta}_1$. The standardized variance is defined as the variance of the k^{th} marginal posterior divided by the variance of the 1^{st} marginal posterior (the standardized variance is used for visualization purposes). Figure 3.3 shows an estimate of the k^{th} posterior distribution for the model parameter R_{P_0}, $\pi_k(R_{\text{P}_0}|\boldsymbol{E})$, for $k = 1$ and $k = 70$. Here, we see that as k grows, the variance decreases and the distribution centers on $\widehat{\boldsymbol{\theta}}$. This analysis provides strong evidence that the trust region algorithm converged to the MLE, i.e., $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\text{ML}}$.

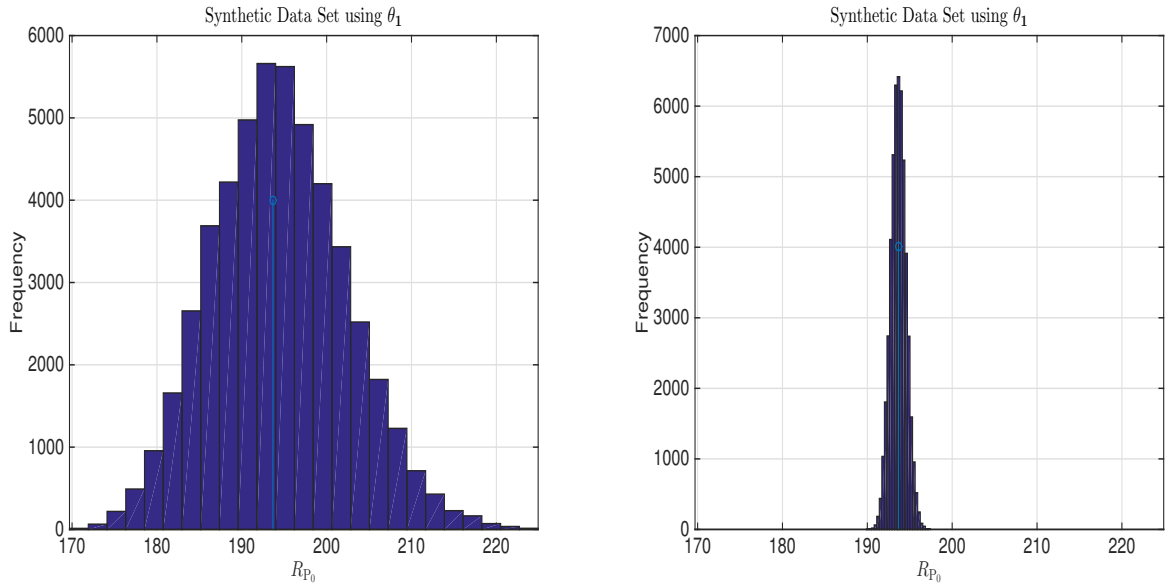


Figure 3.3: Histograms of $\pi_k(R_{\text{P}_0}|\boldsymbol{E})$ for $k = 1$ and $k = 70$ using the synthetic data generated by $\boldsymbol{\theta}_1$. After cloning, the variance of the distribution becomes small and the mean of the distributions is the MLE (shown with a vertical line).

The second class consists of synthetic data sets for which there is evidence of inestimability for R_{P_0}. For data sets in this class—data sets generated by $\boldsymbol{\theta}_4$, $\boldsymbol{\theta}_5$, $\boldsymbol{\theta}_7$, $\boldsymbol{\theta}_8$, and $\boldsymbol{\theta}_{10}$—as k increases, the variance of $\pi_k(R_{\text{P}_0}|\boldsymbol{E})$ does not decay at a rate of $1/k$. Figure 4.1 (right) provides a visualization of this fact using a standardized variance plot for $\boldsymbol{\theta}_4$. Figure 3.4 shows an estimate of $\pi_k(R_{\text{P}_0}|\boldsymbol{E})$ for $k = 1$ and $k = 70$. Here, we do not see a decaying variance as k grows. Figure 3.5 shows a plot of values of R_{P_0} in the support of $\pi_{70}(R_{\text{P}_0}|\boldsymbol{E})$ against the $\log(L(\boldsymbol{\theta}|\boldsymbol{E}))$ with all other model parameters fixed. We see that for these values of R_{P_0}, there is very little difference in $\log(L(\boldsymbol{\theta}|\boldsymbol{E}))$. This plot indicates that a maximum is attained

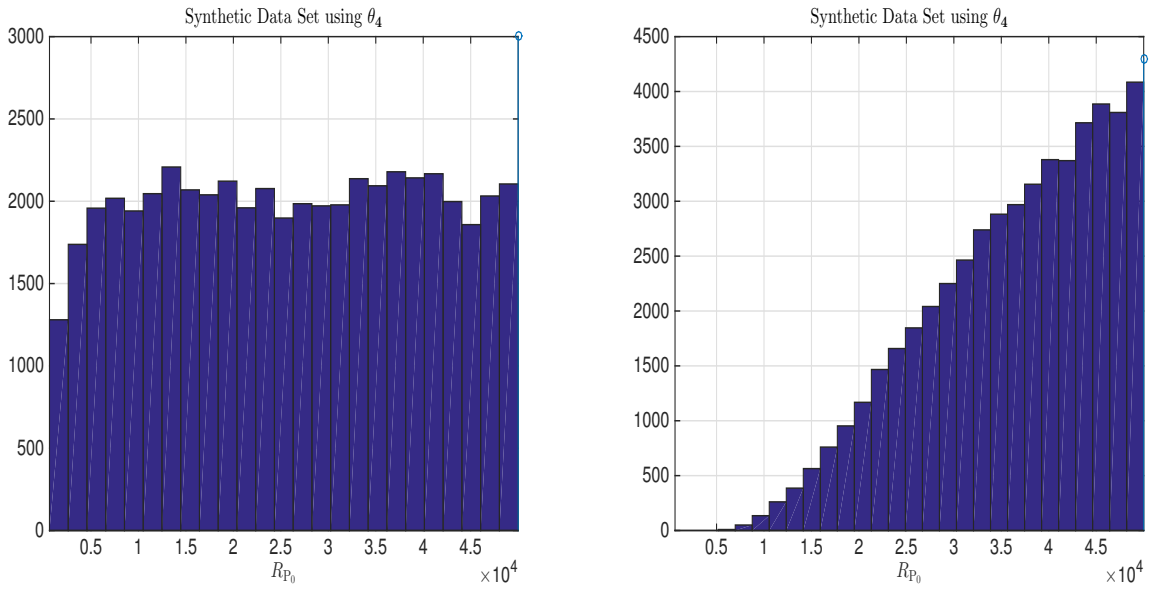


Figure 3.4: Histograms of $\pi_k(R_{\mathrm{P}_0}|\boldsymbol{E})$ for $k = 1$ and $k = 70$ using the synthetic data generated by $\boldsymbol{\theta}_4$. After cloning, the variance of the distribution does not become small. The vertical line denotes the MLE found by MATLAB's `lsqnonlin` function.

near the boundary and that many values in the domain nearly maximize $\log(L(\boldsymbol{\theta}|\boldsymbol{E}))$.

For all synthetic data sets, Markov chains for the key performance parameters $P_{\mathrm{MAX}_0} = P_{\mathrm{MAX}_0}(\boldsymbol{\theta})$ and $V_{\mathrm{OC}_0} = V_{\mathrm{OC}_0}(\boldsymbol{\theta})$ were calculated.[15] The variances of the marginal posterior distributions of these parameters decay at a rate of $1/k$ in both classes. That is, even though there is evidence that, in some regions of the parameter space, R_{P_0} is inestimable, there is no evidence that $P_{\mathrm{MAX}_0} = P_{\mathrm{MAX}_0}(\boldsymbol{\theta})$ or $V_{\mathrm{OC}_0} = V_{\mathrm{OC}_0}(\boldsymbol{\theta})$ is inestimable.

3.7.2 PVM 101 Data

PVM 101 is a 4 cm^2 mono-Si PV cell manufactured by PV Measurements, Inc. We chose the PVM 101 cell to check for evidence of inestimability because the key performance parameters of this cell are thought to be well known. The data used in this analysis consist

[15]Note that for $E = 1$ and for a fixed $\boldsymbol{\theta}$,

$$P_{\mathrm{MAX}_0}(\boldsymbol{\theta}) = \max_{V,I} V\, I = \max_{V} V\, g_I(V, E; \boldsymbol{\theta}),$$

where $g_I(V, E; \boldsymbol{\theta})$ is equation (5.9) solved for I.

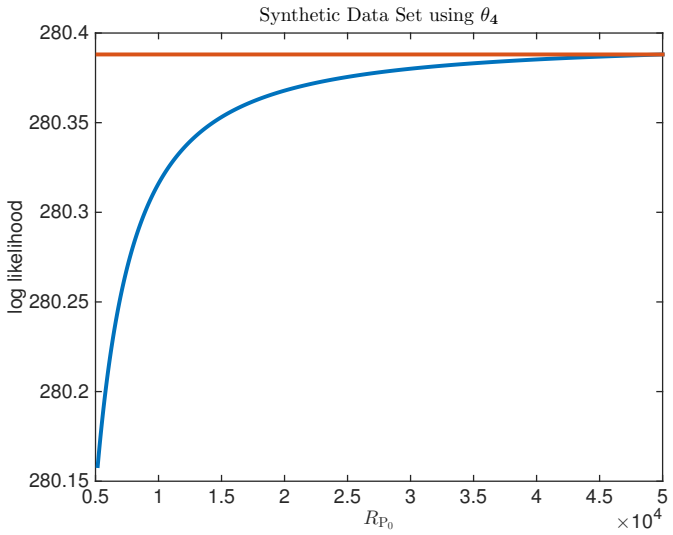


Figure 3.5: A plot of values of R_{P_0} in the support of $\pi_{70}(R_{P_0}|\boldsymbol{E})$ against the $\log(L(\boldsymbol{\theta}|\boldsymbol{E}))$ with all other model parameters fixed.

of 97 I-V-E points measured at NREL's Measurements and Characterization Group. We repeat the analysis from Section 3.7.1 with these data for $k \in \{1, 10, 20, 30, ..., 70\}$, using the same likelihood and prior distributions described in Section 3.7.1.[16] Figure 3.6 shows plots of the standardized variances of the marginal posterior distributions for the model and key performance parameters against the number of clones k. We see that the variances decay to zero at the rate of $1/k$. This analysis provides no evidence of inestimability and thus, no evidence of non-identifiability. Further, for each of the three prior distributions used in the data cloning diagnostic and for $i \in \{1, ..., 5\}$, the mean of $\pi_k(\theta_i|\boldsymbol{E})$ converges to the same value (shown in Table 3.3). Thus, given the modeling assumptions—e.g., that the single-diode model is a reasonable representation of the performance of the PVM 101, that the noise in E is truly iid log-normal—we believe that this analysis shows that maximum likelihood estimation performs well as a method for estimating model and key performance parameters of PVM 101.

[16]Note that, as with the synthetic data, we use $\widehat{\sigma}_E^2$ in MCMC evaluations of the likelihood.

Table 3.2: A table showing the true values of the parameters for each of the ten synthetic data sets used in Section 3.7.1.

$\boldsymbol{\theta}'$	$\boldsymbol{\theta}'_1$	$\widehat{\boldsymbol{\theta}'}_{\text{ML}}$	95% Confidence Interval	% error
$\theta_1 = I_{\text{SC}_0}$	0.11979	0.11974	(0.11966, 0.11982)	0.03930
$\theta_2 = I_{\text{S}_0}$	2.20×10^{-8}	2.21×10^{-8}	$(1.82 \times 10^{-8}, 2.60 \times 10^{-8})$	0.16
$\theta_3 = n_0$	1.500	1.500	(1.483, 1.517)	0.0146
$\theta_4 = R_{\text{S}_0}$	0.3336	0.3337	(0.3279, 0.3395)	0.3660
$\theta_5 = R_{\text{P}_0}$	187.5	194.0	(180.0 208.0)	3.4
P_{max_0}	0.050060	0.050059	(0.050029, 0.050090)	0.00399
V_{OC_0}	0.59678	0.59682	(0.59676, 0.59688)	0.0073

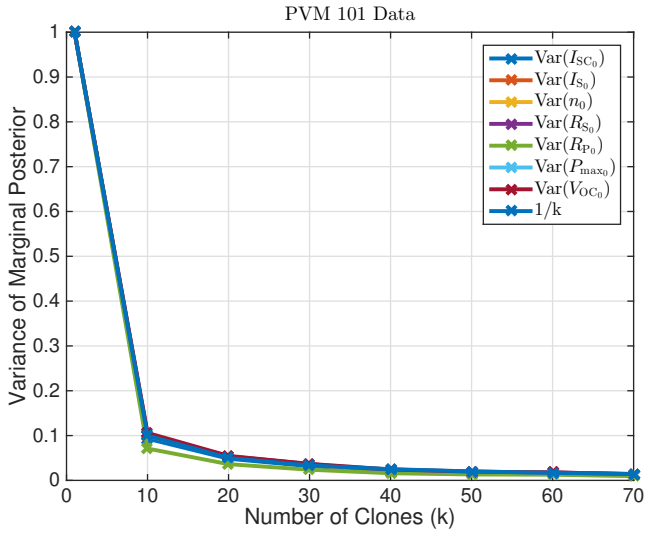


Figure 3.6: A plot of the standardized variance of the marginal posterior distribution for the PVM101 model and key performance parameters. We see that the variance decay at a rate of $1/k$ giving no evidence of inestimability.

Table 3.3: A table comparing $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$ with values found in a controlled laboratory setting. $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$ was obtained using MATLAB's `lsqnonlin` function and verified using data cloning ($k = 70$). Also shown are the 95% confidence intervals and the percent error from $\boldsymbol{\theta}'_{\mathrm{lab}}$ (when applicable).

$\boldsymbol{\theta}'$	$\boldsymbol{\theta}'_{\mathrm{lab}}$	$\widehat{\boldsymbol{\theta}'}_{\mathrm{ML}}$	95% Confidence Interval	% error
$\theta_1 = I_{\mathrm{SC}_0}$	0.113800	0.1138800	(0.113836, 0.113930)	0.07
$\theta_2 = I_{\mathrm{S}_0}$	—	5.27e-10	(4.72e-10, 5.81e-10)	—
$\theta_3 = n_0$	—	1.087	(1.081, 1.093)	—
$\theta_4 = R_{\mathrm{S}_0}$	—	0.03026	(0.02878, 0.03174)	—
$\theta_5 = R_{\mathrm{P}_0}$	—	905.4	(732.6, 1078.20	—
P_{max_0}	0.048310	0.048380	(0.048343, 0.048413)	0.15
V_{OC_0}	0.5366	0.5357	(0.5356, 0.5357)	0.17

3.8 Conclusions and Future Work

The approach to single-diode parameter estimation adopted here models measurement noise in the data using a statistical framework. This framework is desirable for at least two reasons. First, through data cloning, the statistical framework provides a way to check for evidence that the model and key performance parameters are inestimable or non-identifiable. In regions of the parameter space where there is no evidence of inestimability or non-identifiability, the statistical framework may help overcome worries in the PV literature about the non-uniqueness of single-diode model parameter estimates; if the MLE is unique, then maximum likelihood estimation may work well as a PV industry standard for measuring device performance. In regions of the parameter space where there is evidence of inestimability in the model parameters, the statistical framework indicates that the key performance parameters are, nevertheless, estimable. Second, if the parameters are estimable, the statistical approach outlined here can easily quantify the uncertainty in the parameter estimates—tracing it back to the noise in the data—using confidence intervals.

The present work can be improved upon and expanded in several ways. For one, the assumed log-normal noise model for E should be validated. Second, the given modeling procedure can be generalized to a double-diode or triple diode-model and can be expanded

to account for variations in temperature and larger variations in the irradiance. These expansions are nontrivial because they would introduce, and thus require the estimation of, several new parameters. Third, the data cloning diagnostic can be used to evaluate whether formulating the single-diode model in terms of conductance ($1/R_{\mathrm{P}}$), rather than R_{P}, would eliminate estimability issues. Such a reformulation may be more numerically stable than the widely used single-diode model formulation that has been investigated in this paper. Fourth, the statistical model can be generalized to account for measurement noise in I and V. Finally, it would be useful to consider uncertainty caused by model discrepancy in the current and future models.

3.9 Methodological Investigations

In section 3.5, we defined identifiability and estimability, and gave examples of models that exhibit these properties. In this section, we expand on that analysis to show how data cloning works on a few simple problems. Such simple problems give insight into the cloning method. In Section 3.9.1, we look at estimating the mean of a normal distribution. Then, in Sections 3.9.2 and 3.9.3, we look at two non-identifiable models—a normal model with $\mu = |\theta|$ and a nonlinear model that includes the product of two parameters—to show how inestimability can be diagnosed with data cloning; these models also provide an opportunity to show how, in some cases, data cloning may fail to diagnose inestimability.

3.9.1 Normal Mean

Consider the data set $\boldsymbol{X} = (X_1, ..., X_N)$ where $X_i \overset{iid}{\sim} \mathcal{N}(\theta, 1)$. We know analytically that θ is both identifiable and estimable. It is identifiable because any change in θ will result in a change in the center of the probability distribution; the model is estimable because

$$\widehat{\theta}_{\mathrm{ML}} = \bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

yields a unique answer. The data cloning diagnostic provides results consistent with these facts. Figure 3.7 shows the variance of the marginal posterior distribution for θ, $\pi_k(\theta|\boldsymbol{x})$, with $X_i \overset{iid}{\sim} \mathcal{N}(4,1)$, $N = 30$, and the prior on θ defined as $\theta \sim \mathcal{N}(3,1)$. We see that the standardized variance decreases at a rate of $1/k$ giving no evidence of inestimability. Figure 3.8 shows a histogram of $\pi_1(\theta|\boldsymbol{x})$ and $\pi_{40}(\theta|\boldsymbol{x})$. After cloning, the mass of the distribution accumulates around $\widehat{\theta}_{\text{ML}} \approx 4.375$.

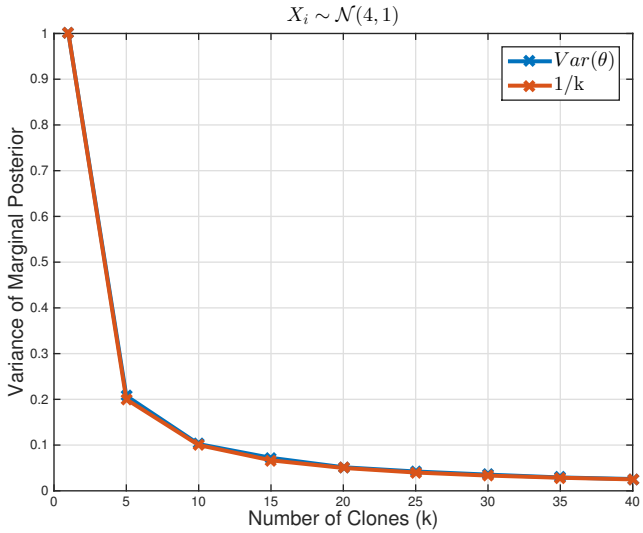


Figure 3.7: A plot of the standardized variance of the marginal posterior distribution for θ, $\pi_k(\theta|\boldsymbol{x})$. The variance decreases at a rate of $1/k$ giving no evidence of inestimability.

3.9.2 Absolute Value of Normal Mean

Consider the data set $\boldsymbol{X} = (X_1, ..., X_N)$ where $X_i \overset{iid}{\sim} \mathcal{N}(|\theta|, \sigma^2)$. We noted above that the model parameter θ is non-identifiable because for $\theta_2 = a \neq 0$ and $\theta_1 = -a$, $L(\theta_2, \sigma^2|\boldsymbol{x}) = L(\theta_1, \sigma^2|\boldsymbol{x})$. The parameter θ is also inestimable because, if $\widehat{\theta}$ maximizes the likelihood then so does $-\widehat{\theta}$. Thus, in this case, data cloning ought provide evidence of inestimability, (and non-identifiability). However, in some cases, we see that data cloning may fail to provide such evidence. Below, we look at two such cases. But first, we look at a case where data cloning correctly diagnoses inestimability.

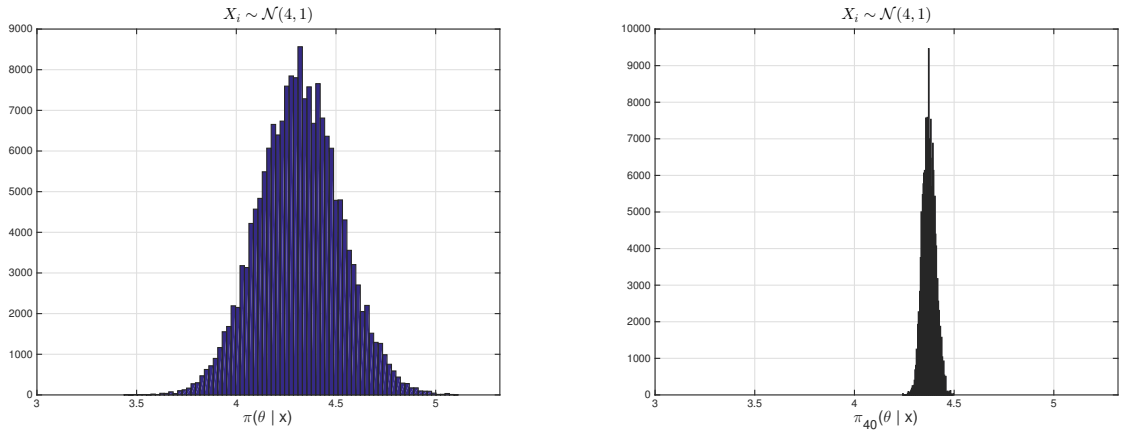


Figure 3.8: Histograms of $\pi_1(\theta|\boldsymbol{x})$ and $\pi_{40}(\theta|\boldsymbol{x})$ for $X_i \overset{iid}{\sim} \mathcal{N}(4,1)$, $N = 30$, and $\theta \sim \mathcal{N}(3,1)$. After cloning, the mass of the distribution accumulates around $\widehat{\theta}_{\mathrm{ML}} = 4.375$.

Correct Diagnosis. Consider $X_i \overset{iid}{\sim} \mathcal{N}(|-4|,1)$, $N = 30$ and $\theta \sim \mathcal{N}(0,2)$. Figure 3.9 shows a plot of the variance of the posterior distribution for θ and the curve $1/k$. Using the Metropolis-Hastings MCMC algorithm with 3×10^5 samples and $k = 40$ clones, we see that the variance of the posterior of θ does not decrease at the rate of $1/k$. Figure 3.10 shows histograms for of $\pi_1(\theta|\boldsymbol{x})$ and $\pi_{40}(\theta|\boldsymbol{x})$; we note that the mass of $\pi_{40}(\theta|\boldsymbol{x})$ accumulates around $\theta \approx \pm 4$. These figures provide evidence of inestimability in θ.

Choice of Prior. The first case of the failure of the data cloning diagnostic is related to the choice of the prior distribution. With respect to using data cloning as a method for calculating MLEs, the authors of [32] write that "as long as the prior distribution is not degenerate and the model satisfies some regularity conditions, the results do not depend on the choice of the prior distribution. Nevertheless, a prior that has large probability mass near the true MLE requires fewer clones to achieve the desired accuracy." Although in theory, any proper prior distribution ought to suffice in the data cloning algorithm, in practice, when using data cloning to diagnose inestimability, certain priors will lead practitioners astray. For example, consider again $X_i \overset{iid}{\sim} \mathcal{N}(|-4|,1)$, $N = 30$ and $\theta \sim \mathcal{N}(5,1)$. Because the prior distribution has mass concentrated near one of the two possible values for θ but quite far from the other, most reasonable attempts at data cloning will conclude that there is no evidence of ines-

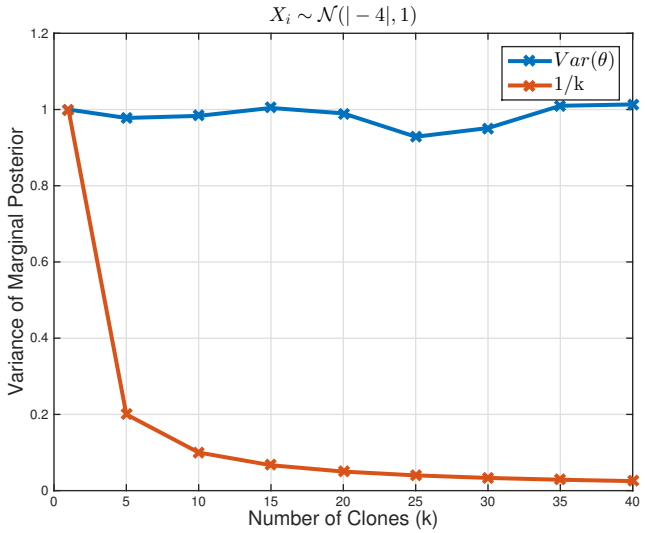


Figure 3.9: A plot of the standardized variance of the marginal posterior distribution for θ, $\pi_k(\theta|\boldsymbol{X})$ with $X_i \overset{iid}{\sim} \mathcal{N}(|-4|, 1)$ and $\theta \sim \mathcal{N}(0, 2)$. We see that the variance does not decrease at a rate of $1/k$ giving evidence of inestimability.

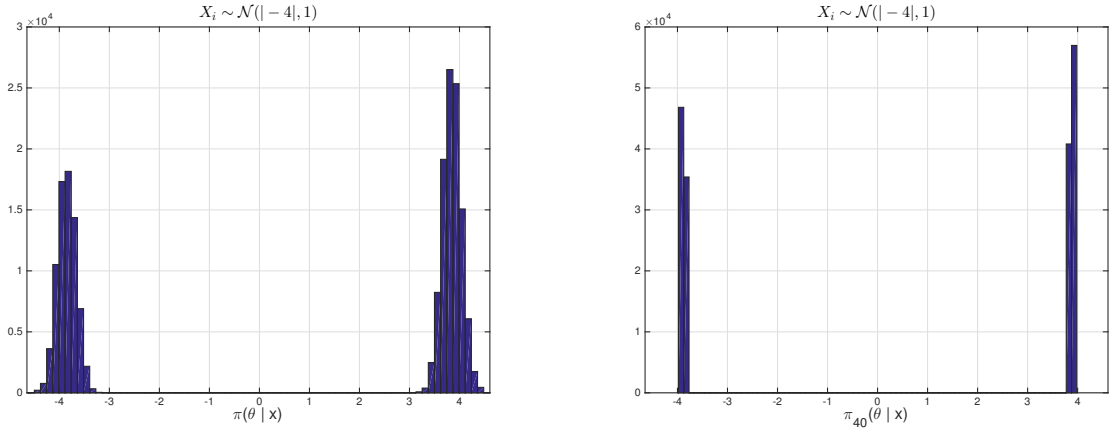


Figure 3.10: Histograms of $\pi_1(\theta|\boldsymbol{x})$ and $\pi_{40}(\theta|\boldsymbol{x})$ for $X_i \overset{iid}{\sim} \mathcal{N}(|-4|, 1)$, $N = 30$, and $\theta \sim \mathcal{N}(0, 2)$. After cloning, the mass of the distribution accumulates around $\theta \approx \pm 4$.

timability in θ. Using Metropolis-Hastings MCMC with 3×10^5 samples and $k = 40$ clones, we see that, in Figure 3.11, the variance decreases at a rate of $1/k$ giving no evidence of inestimability. Figure 3.12 shows histograms of $\pi_1(\theta|\boldsymbol{x})$ and $\pi_{40}(\theta|\boldsymbol{x})$ for this model. We see that the mass of the distribution concentrates around $\theta \approx 3.8$. These results are misleading because θ is inestimable and non-identifiable. An example such as this one serves as a warning about the use of data cloning. In particular, one should be careful in selecting a prior distribution because, in practice, the wrong prior may affect the reliability of the results.

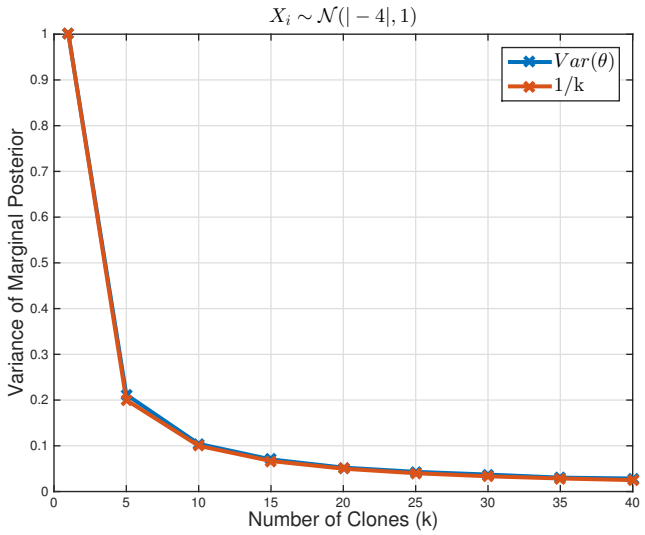


Figure 3.11: A plot of the standardized variance of the marginal posterior distribution for θ, $\pi_k(\theta|\boldsymbol{X})$ with $X_i \overset{iid}{\sim} \mathcal{N}(|-4|, 1)$ and $\theta \sim \mathcal{N}(5, 1)$. We see that the variance decreases at a rate of $1/k$ giving no evidence of inestimability even though θ is, in fact, inestimable (and non-identifiable).

Choice of MCMC Algorithm. We now look at a second case, which shows that the choice of MCMC algorithm affects the data cloning results. Note above that, when data cloning correctly diagnosed inestimability for the model parameter θ in $X_i \overset{iid}{\sim} \mathcal{N}(|-4|, 1)$, Metropolis-Hastistings was used to estimate the posterior distributions $\pi_k(\theta|\boldsymbol{x})$. When repeating data cloning on the same model using the DRAM algorithm, we see that data cloning fails to diagnose inestimability. Figure 3.13 shows a plot of the standardized vari-

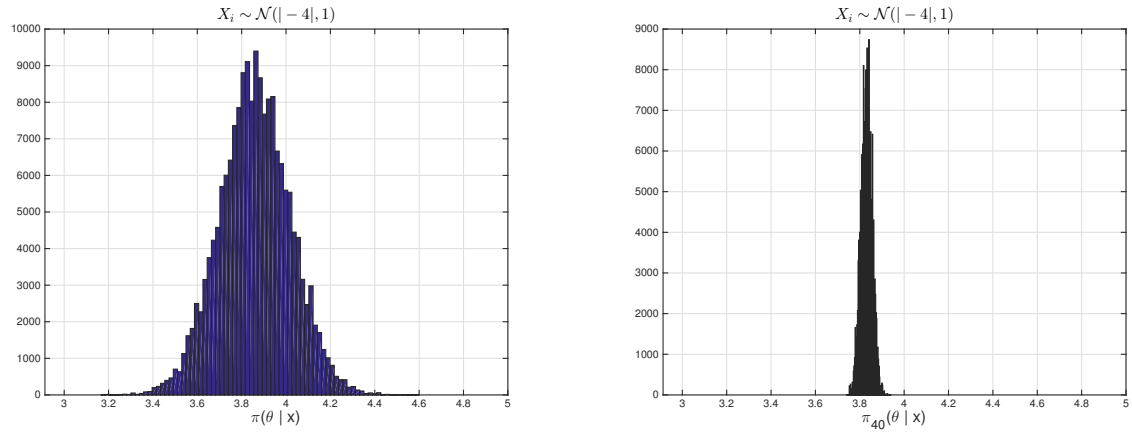


Figure 3.12: Histograms of $\pi_1(\theta|\boldsymbol{x})$ and $\pi_{40}(\theta|\boldsymbol{x})$ for $X_i \overset{iid}{\sim} \mathcal{N}(|-4|, 1)$, $N = 30$, and $\theta \sim \mathcal{N}(5, 1)$. After cloning, the mass of the distribution accumulates around $\theta \approx 3.8$ rather than $\theta \approx \pm 3.8$.

ance of the marginal posterior distribution. Unlike the case where $\pi_k(\theta|\boldsymbol{X})$ was estimated using Metropolis-Hastings, here, the variance decreases at a rate of $1/k$ giving no evidence of inestimability in θ. However, the model is inestimable (and non-identifiable). Figure 3.14 shows the corresponding histogram. A reason for the failure may be due to the fact that the DRAM algorithm adapts the proposal covariance after every M samples. Thus, DRAM is likely exploring the peak closest to the prior, adapting the proposal covariance matrix to be smaller, and then never (or very rarely) jumping far enough to get out of that peak.

3.9.3 Multiplication of Parameters

In this section, we consider the model

$$Y_i = e^{\theta_1 \theta_2 x_i} + \varepsilon_i, \tag{3.7}$$

where $\varepsilon_i \sim \mathcal{N}(0, 1)$ and $i \in \{1, ..., N\}$, i.e., $Y_i \sim \mathcal{N}(e^{\theta_1 \theta_2 x_i}, 1)$. The parameters θ_1 and θ_2 in this model are non-identifiable and inestimable. To see this fact, we note that there are infinitely many values such that $\theta_1 \theta_2 = b$, $b \in \mathbb{R}$. However, we note that the product $\theta_1 \theta_2$ is identifiable and estimable. In some cases, data cloning will diagnose the inestimability in θ_1 and θ_2, but in other cases, it will not. We consider each case in turn.

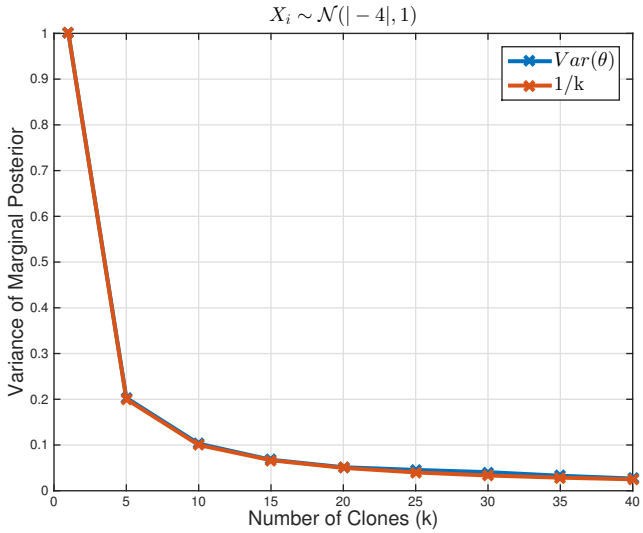


Figure 3.13: A plot of the standardized variance of the marginal posterior distribution for θ, $\pi_k(\theta|\boldsymbol{X})$ with $X_i \overset{iid}{\sim} \mathcal{N}(|-4|, 1)$ and $\theta \sim \mathcal{N}(0, 2)$. In this case, $\pi_k(\theta|\boldsymbol{X})$ was estimated using the DRAM algorithm. We see that, unlike when $\pi_k(\theta|\boldsymbol{X})$ was estimated using Metropolis-Hastings, the variance decreases at a rate of $1/k$ giving no evidence of inestimability in θ. However, θ is inestimable (and non-identifiable).

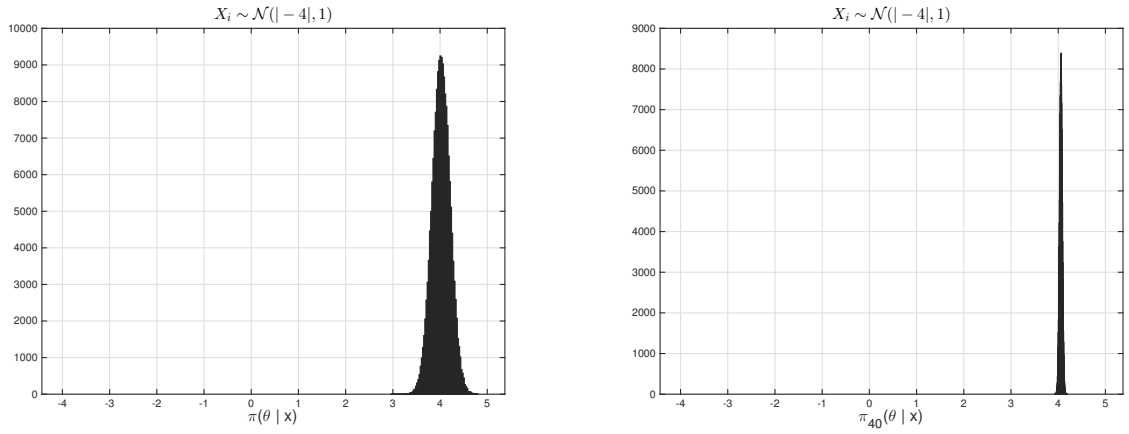


Figure 3.14: Histograms of $\pi_1(\theta|\boldsymbol{x})$ and $\pi_{40}(\theta|\boldsymbol{x})$ for $X_i \overset{iid}{\sim} \mathcal{N}(|-4|, 1)$, $N = 30$, and $\theta \sim \mathcal{N}(2, 1)$ calculated using the DRAM algorithm. After cloning, the mass of the distribution accumulates around $\theta \approx 4$ rather than $\theta \approx \pm 4$.

Correct Diagnosis To show that data cloning can work to diagnose inestimability, we let $Y_i \sim \mathcal{N}(e^{\theta_1\theta_2 x_i}, 1)$, where $\theta_1 = 2$, $\theta_2 = 3$, and $i \in \{1, ..., 30\}$, $\theta_1 \sim \mathcal{N}(2, 1)$, and $\theta_2 \sim \mathcal{N}(3, 1)$. We find that the standardized variance of the marginal posterior distribution for θ_1 and θ_2, $\pi_k(\theta,\theta_2|\boldsymbol{Y})$, does not decrease at a rate of $1/k$, which yields the desired result, that θ_1 and θ_2 are inestimable. However, the standardized variance of the marginal posterior distribution for the product $\theta_1\theta_2$ does decrease at rate $1/k$. Figure 3.15 and Figure 3.16 show these results.

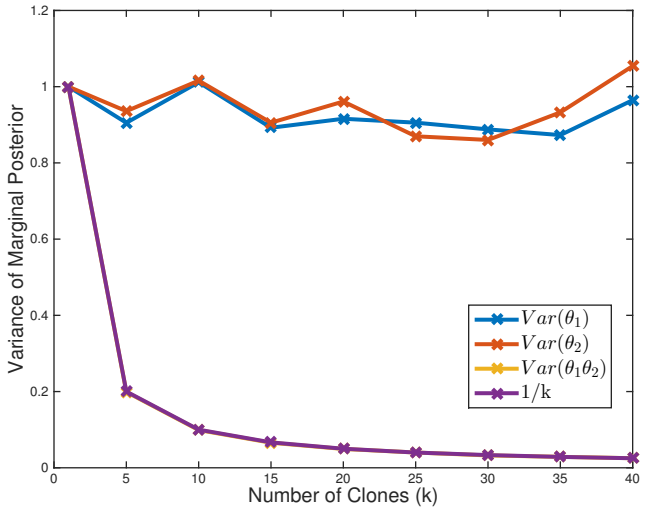


Figure 3.15: A plot of the standardized variance of the marginal posterior distribution for θ_1, θ_2, and the product $\theta_1\theta_2$. The data cloning diagnostic finds evidence of inestimability for θ_1 and θ_2 but not for the product $\theta_1\theta_2$ (the standardized variance for the product follows the $1/k$ curve). These results are in accord with what we know to be true of the parameters.

Incorrect Diagnosis As with the model in Section 3.9.2, it is possible to choose prior distributions such that data cloning yields undesirable results. For such an example, suppose that $Y_i \sim \mathcal{N}(e^{\theta_1\theta_2 x_i}, 1)$, where $\theta_1 = 2$, $\theta_2 = 3$, $N = 30$, and $\theta_1 \sim \mathcal{N}(2, 10)$ and $\theta_2 \sim \mathcal{N}(3, 0.001)$. In this case, we see that the standardized variance of the marginal posterior distribution for θ_1 decreases at rate $1/k$ giving no evidence of inestimability when, in fact, θ_1 is inestimable (Figure 3.17). Figure 3.18 shows the corresponding histograms.

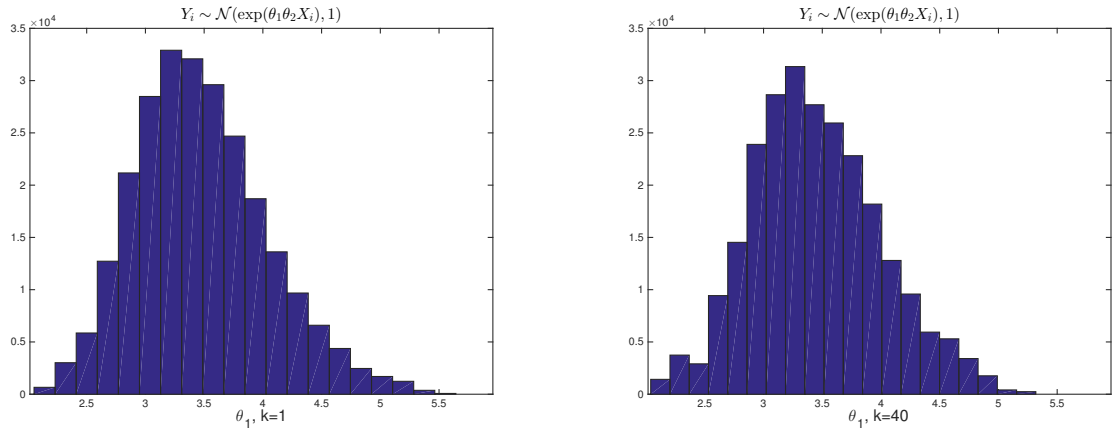


Figure 3.16: Histograms of $\pi_k(\theta_1|\boldsymbol{x})$ for $k \in \{1, 40\}$ and $Y_i \sim \mathcal{N}(e^{\theta_1\theta_2 x_i}, 1)$, where $\theta_1 = 2$, $\theta_2 = 3$, and $N = 30$. Here, $\theta_1 \sim \mathcal{N}(2, 1)$ and $\theta_2 \sim \mathcal{N}(3, 1)$.

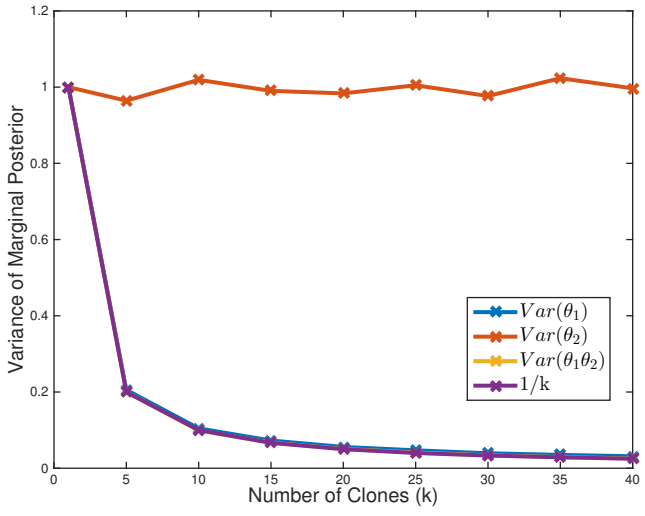


Figure 3.17: A plot of the standardized variance of the marginal posterior distribution for θ_1, θ_2, and the product $\theta_1\theta_2$ with priors $\theta_1 \sim \mathcal{N}(2, 10)$ and $\theta_2 \sim \mathcal{N}(3, 0.001)$. The data cloning diagnostic finds no evidence of inestimability for θ_1 even though θ_1 is inestimable.

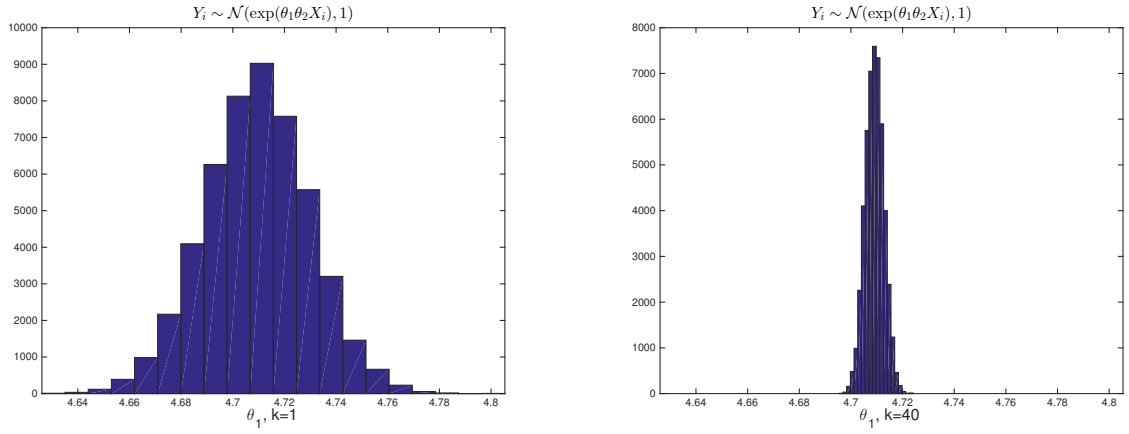


Figure 3.18: Histograms of $\pi_k(\theta_1|\boldsymbol{x})$ for $k \in \{1, 40\}$ and $Y_i \sim \mathcal{N}(e^{\theta_1\theta_2 x_i}, 1)$, where $\theta_1 = 2$, $\theta_2 = 3$, and $N = 30$. Here, $\theta_1 \sim \mathcal{N}(2, 10)$ and $\theta_2 \sim \mathcal{N}(3, 0.001)$. We see that, in using these prior distributions, data cloning finds no evidence of inestimability for θ_1 even though θ_1 is inestimable.

CHAPTER 4

LIKELIHOOD METHODS FOR SINGLE DIODE MODEL PARAMETER ESTIMATION FROM NOISY I-V CURVE DATA

Sections 4.1 through 4.7 comprise a paper featured in the proceedings of the *40th IEEE Photovoltaics Specialists Conference (PVSC)*

Brian Zaharatos[17], Mark Campanelli[18], Clifford Hansen[19], Keith Emery[20], Luis Tenorio[21]

Abstract

Characterizing photovoltaic (PV) device performance is important for the growth of the PV industry. Performance is often characterized by a set of key parameters for the PV device in question: open-circuit voltage, short-circuit current, and maximum power. For a wide range of devices, the key performance parameters are a function of the parameters of a single diode circuit model. In this paper, we developed a statistical model for current-voltage-irradiance data of a PV device using a five-parameter single diode model. The goal was to estimate the single diode model parameters and key performance parameters with quantified uncertainty. Specifically, we found maximum likelihood estimates, quantified uncertainty via confidence intervals for the model and key performance parameters, and explored two important statistical properties of this model—identifiability and estimability.

4.1 Introduction

Decreasing the amount of risk taken on by investors, producers, and consumers encourages the growth of the photovoltaics (PV) industry. To decrease risk, PV manufacturers measure the performance of PV cells and modules, and use these performance characterizations to categorize devices, improve processes, and calibrate predictive models. Performance

[17]Primary researcher and author
[18]National Renewable Energy Lab
[19]Sandia National Laboratories
[20]National Renewable Energy Lab
[21]Department of Applied Mathematics and Statistics, Colorado School of Mines

is often characterized by a set of *key performance parameters*: open-circuit voltage (V_{OC}), short-circuit current (I_{SC}), and maximum power (P_{max}). While PV manufacturers are often able to estimate these values with quantified uncertainty at standard reporting conditions (abbreviated SRC and defined at a standard spectral irradiance, 1000 W/m^2, and 25°C), characterizing performance away from SRC is more challenging. Yet, such characterizations are important because PV devices operate at a wide range of conditions.

One method used to characterize performance away from SRC involves the development of a mathematical model of the PV device. Such models require calibration through parameter estimation from data such as current-voltage (I-V) curve measurements. These parameter estimations can be difficult for numerous reasons. In some cases only limited data are available, and computational algorithms for parameter estimation may have convergence issues, produce non-physical values, and/or be highly sensitive to the limited data (for example, see [9]). Consequently, addressing and quantifying the uncertainty in estimated parameters is becoming more widely recognized [6].

In this paper, we present a single diode model that describes the PV performance of a wide range of devices at irradiance values that deviate slightly around the nominal SRC irradiance value, with the temperature fixed at 25°C (Section 4.2). Although these conditions are typical of a calibration laboratory, the methodology is extendable to diode models parametrized for a wide operating range of irradiance and temperature. Obtaining estimates of the single diode model parameters given I-V curve data is desirable because from these estimates one can estimate the key performance parameters. Our formulation of the single diode model includes a statistical model that accounts for measurement noise in irradiance and thus is amenable to model parameter estimation using maximum likelihood estimation (MLE) techniques. To that end, the likelihood function for the single diode model is defined, incorporating a noise model for the points in one or more measured I-V curves (Section 4.4). Then, assuming that the model is estimable—i.e., assuming that there is a unique maximum of the likelihood function—we calculated MLEs of the single diode model parameters and the key

performance parameters I_{SC}, $P_{\max}$, and V_{OC} for a synthetic data set (Section 4.5). Finally, we use a statistical method called *data cloning* to explore whether the model is in fact estimable (Section 4.6).

4.2 Single Diode Model

For non-negative terminal voltages, the following single-diode lumped-parameter equivalent-circuit model adequately describes the I-V performance of many series-wired photovoltaic (PV) devices near a nominal irradiance and at a fixed temperature:

$$I = I_{\mathrm{L}} - I_{\mathrm{S}}\left(e^{\frac{V+IR_{\mathrm{S}}}{N_{\mathrm{S}}nV_{\mathrm{th}}}} - 1\right) - \frac{V + IR_{\mathrm{S}}}{R_{\mathrm{P}}}, \tag{4.1}$$

where the thermal voltage is $V_{\mathrm{th}} := k_{\mathrm{B}}\,T/q$, and

I device terminal current (A)

V device terminal voltage (V)

$I_{\mathbf{L}}$ photocurrent (A)

$I_{\mathbf{S}}$ diode reverse saturation current of device (A)

$N_{\mathbf{S}}$ number of cells connected in series (unit-less)

n ideality factor of device (unit-less)

$R_{\mathbf{S}}$ series resistance of device (Ω)

$R_{\mathbf{P}}$ parallel (shunt) resistance of device (Ω)

$k_{\mathbf{B}}$ Boltzmann constant ($1.3806504 \times 10^{-23}$ J/K)

T junction temperature of device (K)

q electron charge ($1.602176487 \times 10^{19}$ C).

In general, the five parameters, I_{L}, I_{S}, n, R_{S}, and R_{P}, can depend on temperature, irradiance, or terminal voltage. Here, we will assume that there is no dependence upon terminal

voltage. Thus, auxiliary equations are given that define how each device parameter varies with temperature and irradiance from a reference value defined at SRC. Since measurements in a calibration laboratory are made with a fixed temperature, $T = T_0 = 25°\text{C}$ and an irradiance near 1 sun, we assume that the dependence of n, R_S, and R_P on temperature or irradiance is negligible. Thus, the auxiliary equations for these parameters are given by

$$\begin{aligned} n &= n_0, \\ R_\text{S} &= R_{\text{S}_0}, \\ R_\text{P} &= R_{\text{P}_0}, \end{aligned}$$

where

n_0 diode ideality factor at SRC

$R_{\mathbf{S}_0}$ series resistance at SRC

$R_{\mathbf{P}_0}$ parallel resistance at SRC.

For $T = T_0 = 25°\text{C} = 298.15$ K, the photocurrent I_L is given by the auxiliary equation

$$I_\text{L} := E\, I_{\text{SC}_0} + I_{\text{S}_0}\left(e^{\frac{E\, I_{\text{SC}_0} R_{\text{S}_0}}{N_\text{S} n_0 V_{\text{th}_0}}} - 1\right) + \frac{E\, I_{\text{SC}_0} R_{\text{S}_0}}{R_{\text{P}_0}}. \tag{4.2}$$

The *effective irradiance ratio* E is a unit-less measure of the "effective" irradiance illuminating the device relative to 1 sun at SRC. E is defined as the ratio of the device's short-circuit current at the prevailing irradiance, I_SC, to the device's short-circuit current at SRC, I_{SC_0}. This ratio is readily derived by measuring the short-circuit current of a calibrated reference device that is spectrally matched to the device and having the same angular response. E's measurement has random noise whose characteristic is particular to the measurement testbed.

If the reference device is not spectrally matched, then from the definition of the unit-less spectral correction factor M [30], which compensates for spectral response differences between the device and the reference device, it follows that

$$E := \frac{I_{\mathrm{SC}}}{I_{\mathrm{SC}_0}} = \frac{M I_{\mathrm{SC,R}}}{I_{\mathrm{SC,R}_0}} \tag{4.3}$$

where $I_{\mathrm{SC,R}}$ is the reference device's short-circuit current at the prevailing irradiance and $I_{\mathrm{SC,R}_0}$ is the reference device's short-circuit current at SRC (typically a calibration value). The computation of M requires additional information about the quantum efficiencies of the devices and the spectral irradiance of the light source. In this work, we do not consider additional uncertainty due to M and $I_{\mathrm{SC,R}_0}$.

Auxiliary equation (5.10) describes the device's photocurrent generation as a function of E and the model parameters. This equation is derived by setting $V = 0$ and $I = I_{\mathrm{SC}}$ in (5.9), using definition (4.3) to substitute $I_{\mathrm{SC}} = E\, I_{\mathrm{SC}_0}$, and then solving for I_{L}. We choose to infer the short-circuit current at SRC, I_{SC_0}, instead of the photocurrent at SRC, I_{L_0}, because I_{SC_0} is typically of primary interest for performance evaluation. I_{L_0} can be derived directly from I_{SC_0} using (5.10) with $E = 1$. Altogether, the model parameters, which correspond to reference values at SRC, are denoted by the vector $\boldsymbol{\theta} = (I_{\mathrm{SC}_0}, I_{\mathrm{S}_0}, n_0, R_{\mathrm{S}_0}, R_{\mathrm{P}_0})$.

Our initial investigation into likelihood methods for inferring the five single diode model parameters concerned I-V curve measurements taken very close to "1-sun equivalent" irradiance in well-controlled laboratory conditions, i.e., $E \approx 1$ and $T = T_0$. Thus, we did not consider the effects that large variations in temperature or irradiance have on several of the model parameters. For example, I_{S_0} or R_{P_0} would typically change if SRC were defined at a different temperature or irradiance, respectively. Including these dependencies typically requires additional equations auxiliary to (5.9) that capture the deviation of the diode model parameters from their SRC values as temperature and irradiance change [28]. Extra auxiliary equations typically introduce additional model parameters that likely depend upon the PV material system. Such model extensions and comparisons are a subject of current investigation.

4.3 Solving the Single Diode Model

Given $\boldsymbol{\theta}$ in the parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^5$, solutions to the single diode model (5.9) define a *performance surface* in (V, I, E)-space. This surface implicitly defines three functions: $I = g_I(V, E; \boldsymbol{\theta})$, $V = g_V(I, E; \boldsymbol{\theta})$, and $E = g_E(V, I; \boldsymbol{\theta})$. The Lambert W function can be employed for the efficient computation of any of these three functions [31]. The PV_LIB Toolbox [37] for MATLAB [38] provides vectorized computation of Lambert W via the function `wapr_vec`. Computation of I-V curves in the first quadrant is verified against an explicit computational scheme using junction voltages (detailed in [39]), over a very large parameter space for $\boldsymbol{\theta}$ and for a range of practical values for N_{S}. This verification ensures that the computational algorithms operate correctly for a wide variety of PV devices.

Table 4.1: A table comparing the true values of $\boldsymbol{\theta}$ with $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$. $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ was obtained using MATLAB's `mle` function. Also shown for $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ are the 95% confidence intervals and the percent error from $\boldsymbol{\theta}'_{\mathrm{true}}$.

$\boldsymbol{\theta}'$	$\boldsymbol{\theta}'_{\mathrm{true}}$	parameter domain	$\hat{\boldsymbol{\theta}}'_{\mathrm{ML}}$	95% Confidence Interval	% error
$\theta_1 = I_{\mathrm{SC}_0}$	0.11979	(1/12, 12)	0.11974	(0.11966, 0.11982)	0.0393
$\theta_2 = I_{\mathrm{S}_0}$	2.2×10^{-8}	$(10^{-12}, 10^{-5})$	2.2061×10^{-8}	$(1.815 \times 10^{-8}, 2.595 \times 10^{-8})$	0.1602
$\theta_3 = n_0$	1.5	(1, 2)	1.5003	(1.483, 1.517)	0.0146
$\theta_4 = R_{\mathrm{S}_0}$	0.3336	(0.1, 10)	0.33370	(0.3279, 0.3395)	0.3660
$\theta_5 = R_{\mathrm{P}_0}$	187.5	(50, 5000)	194.02	(180.01 207.97)	3.416
V_{OC_0}	0.59678	—	0.59682	(0.59676, 0.59688)	0.0073
P_{max_0}	0.05006	—	0.050059	(0.050029, 0.050090)	0.00399
σ_E^2	1×10^{-6}	—	8.3903×10^{-7}	$(5.101 \times 10^{-7}, 1.168 \times 10^{-6})$	16.10

4.4 The Likelihood Function

In defining a likelihood function, we first make our assumptions explicit. First, we assume that the single diode model (5.9)–(5.10) is a sufficiently accurate representation of the PV device's performance at fixed temperature near 1 sun.[22] Second, we must specify a noise model for the observables. We assume that the noise in the measurement of V and I is

[22] Uncertainty caused by so-called model discrepancy is not considered here but will be assessed in future work.

negligible as compared to the measurement noise in E (this is a reasonable assumption in many measurement contexts). Consequently, given a value for $\boldsymbol{\theta}$, we consider an observed effective irradiance value to be generated by the single diode model (5.9)–(5.10) through the corresponding function $g_E(V, I; \boldsymbol{\theta})$ but measured with noise. We assume the noise in observed E values is normally distributed. That is, for a sample of size N data points measured along an I-V curve, and for $n = 1, ..., N$, the statistical model for the value of the n^{th} observable E_n is given by the normal random variable

$$\mathcal{E}_n = g_E(V_n, I_n; \boldsymbol{\theta}) + \varepsilon_n,$$

where the additive noise ε_n is assumed to be independent and identically distributed (i.i.d.) normally with zero mean and variance σ_E^2, i.e., $\varepsilon_n \sim \mathcal{N}(0, \sigma_E^2)$. From these assumptions, we can define the likelihood function, explore the estimability and identifiability of the model, and obtain the MLE.

In defining the likelihood, we note that, for $n = 1, ..., N$, the n^{th} measured effective irradiance value is determined by the given statistical model, and is thus a realization of the random variable, $\mathcal{E}_n$, which has a normal distribution

$$\mathcal{E}_n \sim \mathcal{N}\big(g_E(V_n, I_n; \boldsymbol{\theta}), \sigma_E^2\big)\,.$$

The likelihood function is defined by fixing the data to actual observed values, denoted $\mathcal{E} = \boldsymbol{E} = (E_1, ..., E_N)$, and interpreting the joint probability density function (pdf) as a function of $\boldsymbol{\theta}$ and σ_E^2. Since the data are assumed i.i.d., the likelihood function for the data is

$$L(\boldsymbol{\theta}|\boldsymbol{E}) = \left(2\,\pi\,\sigma_E^2\right)^{-N/2} e^{-\frac{\sum_{n=1}^{N}(E_n - g_E(V_n, I_n; \boldsymbol{\theta}))^2}{2\sigma_E^2}}. \tag{4.4}$$

4.5 Maximum Likelihood Estimation

Given observed data $\boldsymbol{E}$, an MLE of the model parameters $\boldsymbol{\theta}$, denoted $\hat{\boldsymbol{\theta}}_{\text{ML}}$, is a value in the parameter space $\boldsymbol{\Theta}$ that maximizes the likelihood function $L(\boldsymbol{\theta}|\boldsymbol{E})$:

$$\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}}\, L(\boldsymbol{\theta}|\boldsymbol{E}).$$

Along with finding an MLE for $\boldsymbol{\theta}$, we are also interested in finding an MLE for the key performance parameters at SRC: V_{OC_0}, I_{SC_0} and P_{max_0}. For convenience, we define $\boldsymbol{\theta'} := (I_{\mathrm{SC}_0}, I_{\mathrm{S}_0}, n_0, R_{\mathrm{S}_0}, R_{\mathrm{P}_0}, V_{\mathrm{OC}_0}, P_{\mathrm{max}_0}, \sigma_E^2)$—which differs from $\boldsymbol{\theta}$ by the inclusion of V_{OC_0}, P_{max_0}, and σ_E^2—and denote the MLE of $\boldsymbol{\theta'}$ as $\hat{\boldsymbol{\theta}'}_{\mathrm{ML}}$. Because we are assuming normal i.i.d. noise in our data, we note that the MLE of $\boldsymbol{\theta}$ is equivalent to the nonlinear least squares estimate of $\boldsymbol{\theta}$. This is true because maximizing (4.4) with respect to $\boldsymbol{\theta}$ is equivalent to minimizing $\sum_{i=1}^{N} \left(E_n - g_E(V_n, I_n; \boldsymbol{\theta})\right)^2$ with respect to $\boldsymbol{\theta}$. In addition, maximum likelihood estimation provides a method for quantifying our uncertainty in both $\boldsymbol{\theta}$ and $\boldsymbol{\theta'}$.

An MLE is not guaranteed to exist nor is an MLE guaranteed to be unique if it does exist. For a more thorough understanding of the statistical model, the likelihood function, and the MLE, two properties of the given model must be satisfied: parameter identifiability and parameter estimability. The model parameters are *identifiable* if, for all $\boldsymbol{\theta}_2, \boldsymbol{\theta}_1 \in \boldsymbol{\Theta}$, $L(\boldsymbol{\theta}_2|\boldsymbol{E}) = L(\boldsymbol{\theta}_1|\boldsymbol{E})$ almost everywhere implies that $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_1$. The model parameters are *estimable* if

$$\{\boldsymbol{\theta} \in \boldsymbol{\Theta} : L(\boldsymbol{\theta}|\boldsymbol{E}) = L(\hat{\boldsymbol{\theta}}_{\mathrm{ML}}|\boldsymbol{E})\} = \{\hat{\boldsymbol{\theta}}_{\mathrm{ML}}\}.$$

That is, the model parameters are estimable if an MLE exists and is unique. Often, estimates of parameters are calculated under the assumption that the model is estimable. In this section, we assume that the model is estimable (and thus, identifiable) and calculate MLEs for the parameters. Then, in Section 4.6 we use an exploratory method called *data cloning* [32] to gather evidence about the estimability of $\boldsymbol{\theta'}$.

The basic procedure for finding MLEs of $\boldsymbol{\theta}$ is the same as for finding MLEs of simpler problems: one uses an optimization technique to maximize the likelihood function. In simple cases such as estimating the mean and variance of data known to be from a normal distri-

bution or in linear regression, one can find the MLE of the parameters by taking partial derivatives of the log likelihood function with respect to the parameters, setting those partial derivatives equal to zero, and solving the resulting system of equations over a particular domain. Since equation (4.4) is nonlinear and of higher dimension, we use MATLAB's `mle` function to estimate $\boldsymbol{\theta}$ and then use the corresponding functions of $\hat{\boldsymbol{\theta}}_{\text{ML}}$ to estimate the key performance parameters.

To test this method, we simulated $N = 50$ data points, equally spaced in voltage, according to $E_n \sim \mathcal{N}(g_E(V_n, I_n; \boldsymbol{\theta}), \sigma_E^2)$, with

$$\boldsymbol{\theta}_{\text{true}} = (0.119788, 2.200 \times 10^{-8}, 1.5, 0.3325, 187.5)$$

and $\sigma_E^2 = 10^{-6}$. Table 4.1 shows the MLEs for $\boldsymbol{\theta'}$ along with the domains over which the likelihood function was maximized. These domains are thought to be reasonable in many measurement scenarios.

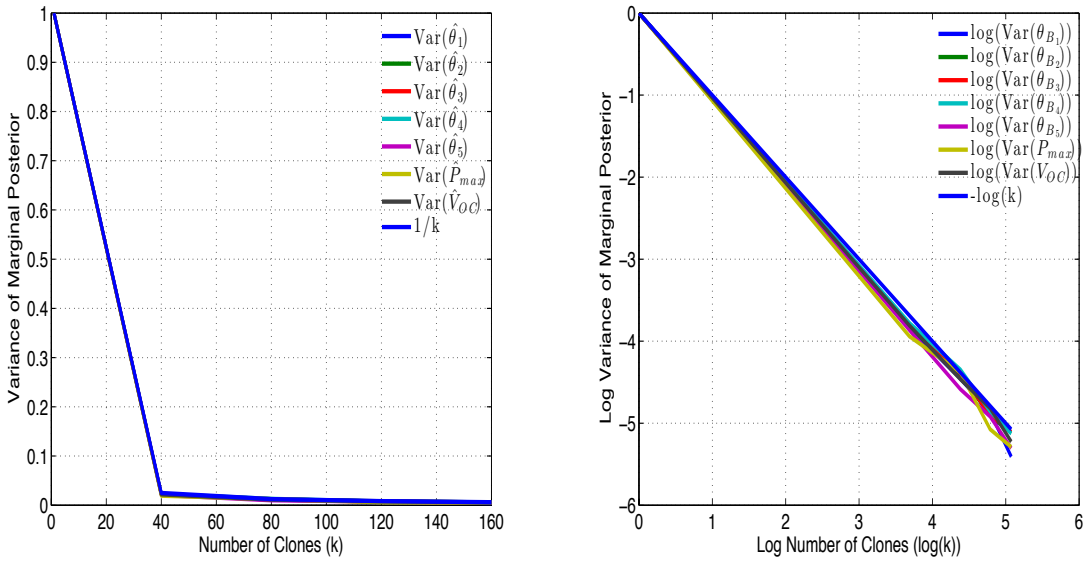


Figure 4.1: Plots of the variance of the k^{th} marginal distributions, $\pi_k(\theta_i|\boldsymbol{E})$, against the number of clones ($k = 1, 40, 80, 120, 160$) are shown to diagnose estimability (standard space on the left and log space on the right). We notice that the marginal variances for the model and performance parameters decay to zero at a rate of $1/k$, giving evidence of estimability and thus, identifiability.

Interval estimates quantify the uncertainty in point estimates by reporting a range of values that is likely to cover the true parameter $\boldsymbol{\theta}'$. Confidence intervals are typically used with the MLE. In this case we use the $(1-\alpha)\times 100\%$ confidence interval for the i^{th} parameter estimate, derived from the asymptotic distribution of $\hat{\boldsymbol{\theta}'}_{\text{ML}}$ [40] using

$$\hat{\theta'}_{\text{ML}_i} \pm z_{\alpha/2}\sqrt{[\mathcal{I}(\hat{\boldsymbol{\theta}'}_{\text{ML}})^{-1}]_{ii}}, \tag{4.5}$$

where $z_{\alpha/2}$ is the $\alpha/2$ quantile for the normal distribution and $[\mathcal{I}(\boldsymbol{\theta}')^{-1}]_{ii}$ is the i^{th} diagonal entry of the inverse of the *Fisher information matrix*, which is defined as

$$\mathcal{I}(\hat{\boldsymbol{\theta}}_{\text{ML}}) = \mathbb{E}\bigg[-\frac{\partial^2}{\partial \boldsymbol{\theta}^2}\log(L(\boldsymbol{\theta}|\boldsymbol{E}))\bigg]\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{ML}}}. \tag{4.6}$$

Informally, the Fisher information matrix evaluated at the MLE measures the curvature of the likelihood function around the MLE. We estimate the inverse Fisher information matrix using MATLAB's `nlinfit` function. For the synthetic data described in this section, equation (4.5) was used to calculate the 95% confidence intervals in Table 4.1 ($z_{\alpha/2} = 1.96$).

4.6 Estimability and Identifiability: An Exploratory Analysis

Recall that, in Section 4.5, we assumed that the model was estimable to find MLEs of the parameters. In this section, we use an exploratory technique called data cloning to assess the estimability and identifiability of the model. These properties are important in practice because unidentifiable models admit multiple parameter sets that explain the data equally well. This situation could lead to confusion from (potentially large) disagreements between parameter estimates computed by different methods using the same data. We note that estimability is a stronger condition than identifiablity: if model parameters are estimable then they are identifiable. Thus, evidence of estimability is evidence of identifiability. In general, assessing the identifiability and estimability of a statistical model with several parameters is difficult. Data cloning provides a way to gather evidence about the estimability of a set of model parameters.

4.6.1 Data Cloning: Theory

Data cloning utilizes some tools in the Bayesian statistical framework, such as Markov chain Monte Carlo (MCMC) methods, to assess estimability.[23] The motivation for data cloning is a thought experiment: suppose that, by coincidence, one observed k independent **identical** samples of size N: $\boldsymbol{E}_{N_k} = \underbrace{(E_1, ..., E_N, ..., E_1, ..., E_N)}_{k \text{ independent repeats}}$. The resulting likelihood function would be

$$L_k(\boldsymbol{\theta}|\boldsymbol{E}_{N_k}) = [L(\boldsymbol{\theta}|\boldsymbol{E})]^k.$$

Bayes' theorem states that the *posterior probability (belief) distribution for* $\boldsymbol{\theta}$ conditioned upon $\boldsymbol{E}$ is proportional to the likelihood function times a *prior probability (belief) distribution for* $\boldsymbol{\theta}$. That is, $\pi(\boldsymbol{\theta}|\boldsymbol{E}) \propto L(\boldsymbol{\theta}|\boldsymbol{E})\pi(\boldsymbol{\theta})$. For $k \in \mathbb{N}$, applying Bayes' theorem to $L_k(\boldsymbol{\theta}|\boldsymbol{E}_{N_k})$ yields the "k^{th} posterior" distribution

$$\pi_k(\boldsymbol{\theta}|\boldsymbol{E}) \propto L_k(\boldsymbol{\theta}|\boldsymbol{E})\pi(\boldsymbol{\theta}), \tag{4.7}$$

where $\pi(\boldsymbol{\theta})$ is any reasonable prior on $\boldsymbol{\theta}$. The study of this posterior distribution, as $k \to \infty$, yields the estimability and MLE results. If $\boldsymbol{\theta}$ is estimable, then for sufficiently large k, $\pi_k(\boldsymbol{\theta}|\boldsymbol{E})$ is nearly degenerate around $\hat{\boldsymbol{\theta}}_{\text{ML}}$ with covariance matrix approximately equal to $\frac{1}{k}\mathcal{I}(\hat{\boldsymbol{\theta}}_{\text{ML}})^{-1}$ [32]. So, as $k \to \infty$, if the variance of the marginal posterior $\pi_k(\theta_i'|\boldsymbol{E})$ does not converge to zero, then θ_i' is inestimable, and thus, nonidentifiable. Further, if θ_i' is inestimable, then, as k increases, $\pi_k(\theta_i'|\boldsymbol{E})$ converges to a truncated distribution over the space of inestimable parameter values.

4.6.2 Data Cloning: Results Using Synthetic Data

Numerical experiments were conducted to check the identifiability and estimability of $\boldsymbol{\theta}$. We used the simulated data described in Section 4.5 for these experiments. Data cloning

[23] Note that there are important philosophical differences in the frequentist statistical paradigm, under which MLE falls, and the Bayesian statistical paradigm, where Bayesian inference is performed. Data cloning itself does not require that one adhere to Bayesian assumptions (e.g., the Bayesian interpretation of probability).

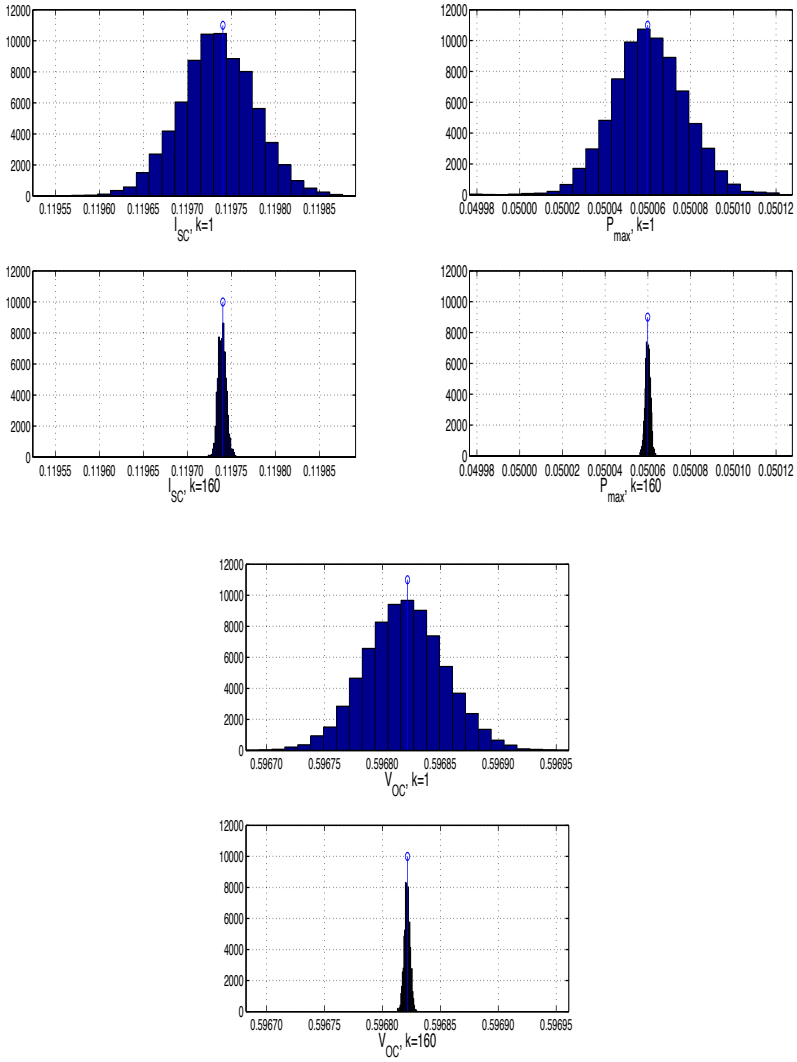


Figure 4.2: . Histograms of the k^{th} marginal posteriors $\pi_k(I_{\text{SC}_0}|\boldsymbol{E})$ (left), $\pi_k(P_{\text{max}_0}|\boldsymbol{E})$ (center), and $\pi_k(V_{\text{OC}_0}|\boldsymbol{E})$ (right) for $k = 1$ and $k = 160$. We notice that after cloning, the variance of the distributions becomes small and the mean of the distributions is the MLE (shown with a vertical line).

was performed using the DRAM algorithm in the MCMC toolbox [36] for MATLAB. For the prior, $\pi(\boldsymbol{\theta})$, we chose a uniform distribution over physically reasonable values (these values correspond to the MLE domains for $\boldsymbol{\theta}$ and are given in Table 4.1).Figure 4.1 shows plots (in standard and log space) of the standardized variance of the marginal posterior distributions of the model and performance parameters against the number of clones ($k = 1, 40, 80, 120, 160$), along with the curve $1/k$—the expected decay rate if the parameters were estimable.[24] We see that the model and performance parameters decay to zero at a rate of $1/k$, giving evidence of estimability and thus, identifiability.Figure 4.2 and Figure 4.3 show histograms of the key performance parameters and model parameters for $k = 1$ and $k = 160$. We see that in each case the variance has been greatly reduced and that the distributions become centered at the MLEs found in Section 4.5, Table 4.1.[25] It follows from this analysis that the estimates of all parameters are meaningful because we have strong evidence of identifiability. Finally, we note that the $k = 1$ case gives a Bayesian uncertainty from the variance of the posterior distribution for $\boldsymbol{\theta}'$.

4.7 Conclusions and Future Work

The likelihood approach to single diode parameter estimation adopted here uses statistical modeling to account for measurement noise in the data. This approach is more useful than some other deterministic single diode model parameter estimation techniques, such as nonlinear least squares, because the MLE approach easily quantifies the uncertainty in the parameter estimates—tracing it back to the noise in the data—using confidence intervals. Using data cloning, we see that the key performance parameters are indeed estimable, and MLEs with confidence intervals can be calculated with relative ease.

There are several ways that the present work can be improved upon and expanded. For one, the assumed normal noise model for E should be validated. Second, the coverage of the

[24] The standardized variance is equal to the variance of the k^{th} posterior divided by the variance of the 1^{st} posterior.

[25] We note that, as data cloning theory suggests, for $k = 160$, k times the covariance matrix of $\pi_k(\boldsymbol{\theta}|\boldsymbol{E})$ is approximately equal to the estimate of the inverse of the Fisher information matrix obtained from MATLAB's `nlinfit` function in Section 4.5.

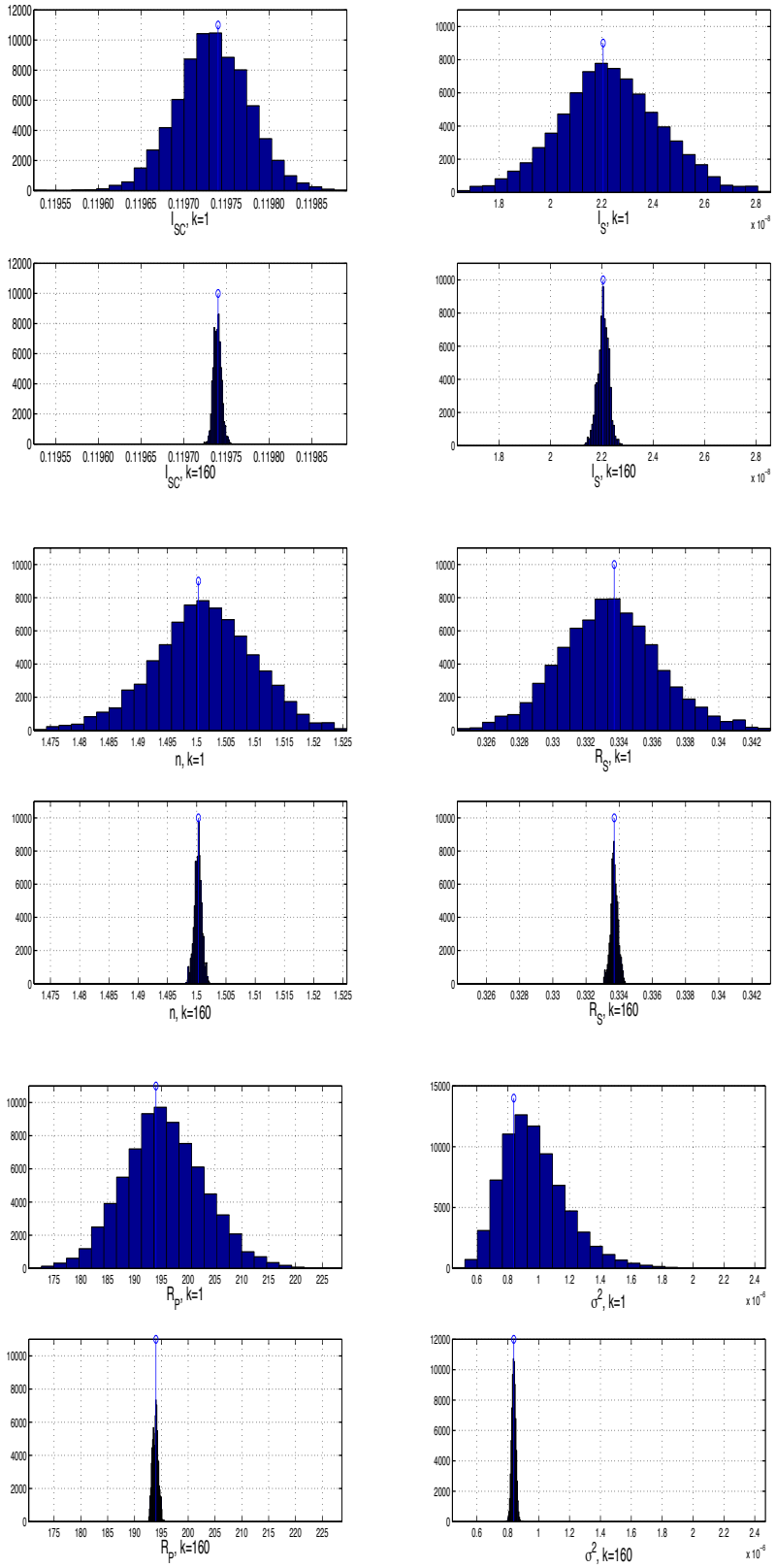


Figure 4.3: Histograms of the k^{th} marginal posterior distributions of $\boldsymbol{\theta} = (I_{\text{SC}_0}, I_{\text{S}_0}, n_0, R_{\text{S}_0}, R_{\text{P}_0})$ and σ_E^2 for $k = 1$ (first and third row) and $k = 160$ (second and fourth row). We notice that after cloning, the variance of the distributions becomes small and the mean of the distributions is the MLE (shown with a vertical line).

confidence interval in (4.5) should be assessed through simulations to see how it compares to the claimed coverage of 95% (significant discrepancies would be due to a small sample size). Third, our formulation of the single diode model can be expanded to account for variations in temperature and larger variations in the irradiance. This expansion is nontrivial because it would introduce, and thus require the estimation of, several new parameters, and because measurements of the junction temperature may have significant systematic and random error. Fourth, the statistical model can be generalized to account for measurement noise in I and V. Finally, it would be useful to consider uncertainty caused by model discrepancy in the current model and future models.

4.8 Extensions of the PVSC Paper

Here, I will address some of the "future work" items outlined in the conclusion of this paper. In particular, I address the true coverage of the asymptotic confidence intervals described in equation (4.5), check the sensitivity of the MLE to the starting values of the optimization algorithm, and use Bayesian inference to estimate model and performance parameters.

4.8.1 True Coverage

In this section, we examine how the true coverage of the asymptotic confidence intervals given in equation (5) compares to the claimed 95% coverage. Specifically, we conduct a parametric bootstrap simulation study using the synthetic data described in the paper above. Recall that, for $n \in \{1, 2, ...N\}$ and for fixed (I_n, V_n, $\boldsymbol{\theta}$), E_n was given by $E_n = g_E(I_n, V_n; \boldsymbol{\theta}) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma_E^2)$. For the parametric bootstrap, we

- Sample $\hat{\epsilon} \sim \mathcal{N}(0, \hat{\sigma}_E^2)$, where $\hat{\sigma}_E^2$ is the MLE of σ_E^2.
- Compute $\boldsymbol{E}^\star = g_E(\boldsymbol{I}, \boldsymbol{V}; \hat{\boldsymbol{\theta}}_{\text{ML}}) + \hat{\epsilon}$.
- Perform data cloning, with $k = 70$. $\hat{\boldsymbol{\theta}}^\star$ denotes the model parameter vector at each step of the Markov chain.

- Compute $\hat{P}^{\star}_{\max}$, $\hat{I}^{\star}_{\mathrm{SC}}$, and $\hat{V}^{\star}_{\mathrm{OC}}$, the key performance parameters at each step of the Markov chain (creating a posterior distribution for each).
- Compute $\mathcal{I}(\hat{\boldsymbol{\theta}'}^{\star})^{-1} = k \times \mathrm{Cov}(\pi_k(\hat{\boldsymbol{\theta}'}^{\star}|\boldsymbol{E}^{\star}))$, where $\hat{\boldsymbol{\theta}'}^{\star}$ is the mean of the joint posterior distribution (including model and key performance parameters).
- Compute

$$\hat{I}^{\star}_i = \hat{\theta'}^{\star}_i \pm z_{\alpha/2}\sqrt{[\mathcal{I}(\hat{\boldsymbol{\theta}'}^{\star})^{-1}]_{ii}}, \quad i \in \{1, ...8\}. \tag{4.8}$$

- Repeat $B = 400$ times.
- Calculate the ratio of times that $\hat{I}^{\star}_i$ covers θ'_i.

Table 4.2: Estimates of the true coverage of each θ'_i

$\boldsymbol{\theta}'$	$\boldsymbol{\theta}'_{\mathrm{true}}$	coverage	mean length
$\theta'_1 = I_{\mathrm{SC}_0}$	0.11979	92.8 ± 0.026	1.4×10^{-4}
$\theta'_2 = I_{\mathrm{S}_0}$	2.2×10^{-8}	92.8 ± 0.026	6.8×10^{-9}
$\theta'_3 = n_0$	1.5	92.5 ± 0.026	0.026
$\theta'_4 = R_{\mathrm{S}_0}$	0.3336	91.5 ± 0.028	0.01
$\theta'_5 = R_{\mathrm{P}_0}$	187.5	92.5 ± 0.026	24.5
$\theta'_6 = V_{\mathrm{OC}_0}$	0.59678	92.5 ± 0.026	1.2×10^{-4}
$\theta'_7 = P_{\max_0}$	0.05006	93.5 ± 0.025	5.9×10^{-5}
$\theta'_8 = \sigma^2_E$	10^{-6}	100 ± 0.007	6.6×10^{-7}

Table 4.3 shows the results of this coverage analysis. We see that the estimated coverage for the model and key performance parameters is lower but close to the claimed coverage, which is evidence that the confidence interval formula is reasonable for this synthetic data set. The confidence interval used in Table 4.3 is the Agresti-Coull confidence intervals for proportions:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{b}}}, \tag{4.9}$$

where

$$\tilde{b} = b + z^2_{\alpha/2},$$

$$\tilde{p} = \frac{b}{\tilde{b}}\hat{p} + \frac{z^2_{\alpha/2}}{2\tilde{b}},$$

and where b the number of bootstrap simulations. This interval is thought to yield better estimates of the true coverage [41].

4.8.2 Bayesian Inversion

In this section, we use the mean and quantiles of $\pi(\theta_i|\boldsymbol{E})$ as point and interval estimates of θ_i, respectively. We note that using the data cloning diagnostic from above, the estimates here are identical to $\pi_k(\theta_i|\boldsymbol{E})$ where $k = 1$. We denote the mean of $\pi(\theta_i|\boldsymbol{E})$ as θ_{B_i}. Table 4.3 summarizes the results of the Bayesian inversion. Figure 4.4 and Figure 4.5 show histograms for $\pi(\theta_i|\boldsymbol{E})$. Figure 4.6 shows histograms of the posterior distributions of $P_{\max}$ and V_{OC}

Table 4.3: A table comparing the true values of $\boldsymbol{\theta}$ with $\boldsymbol{\theta}_{\mathrm{B}}$. $\boldsymbol{\theta}_{\mathrm{B}}$ was obtained using MATLAB's `mcmcrun` function. Also shown for $\boldsymbol{\theta}_{\mathrm{B}}$ are the 95% credible intervals and the percent error from $\boldsymbol{\theta}'_{\mathrm{true}}$.

$\boldsymbol{\theta}'$	$\boldsymbol{\theta}'_{\mathrm{true}}$	$\pi(\boldsymbol{\theta})$	$\boldsymbol{\theta}'_{\mathrm{B}}$	95% Credible Interval
$\theta_1 = I_{\mathrm{SC}_0}$	0.119788	U(1/12, 12)	0.119694	(0.119591, 0.119791)
$\theta_2 = I_{\mathrm{S}_0}$	2.2×10^{-8}	U(10^{-13}, 10^{-6})	2.098×10^{-8}	(1.626×10^{-8}, 2.616×10^{-8})
$\theta_3 = n_0$	1.5	U(1, 2)	1.495	(1.471, 1.516)
$\theta_4 = R_{\mathrm{S}_0}$	0.3336	U(0, 100)	0.33352	(0.3260, 0.3413)
$\theta_5 = R_{\mathrm{P}_0}$	187.5	U(1, 50000)	195.01	(178.17 215.18)
V_{OC_0}	0.59678	—	0.59673	(0.59665, 0.59682)
P_{max_0}	0.05006	—	0.050059	(0.050032, 0.050121)

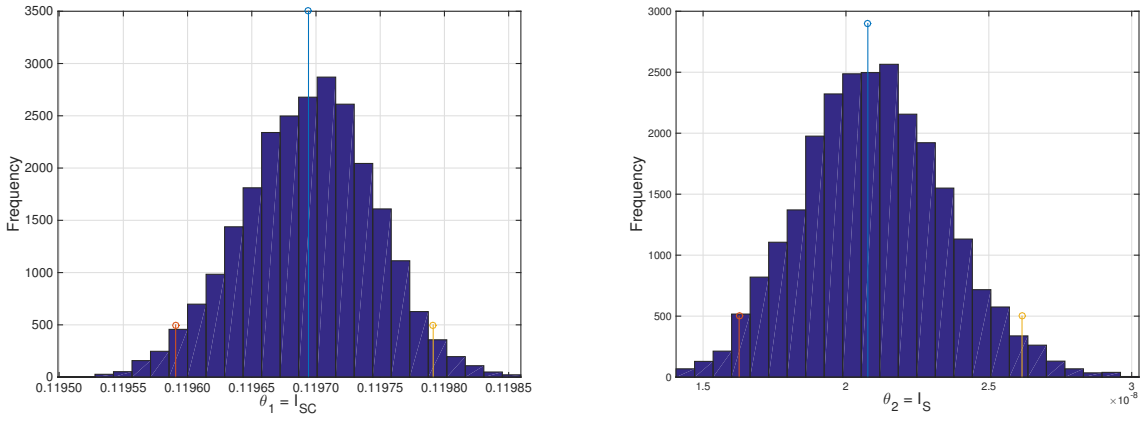


Figure 4.4: A histogram for $\pi(\theta_i|\boldsymbol{E})$ for $i \in \{1, 2\}$. The short vertical lines represent the lower and upper bounds of the 95% credible interval. The tall vertical lines represents the mean of $\pi(\theta_i|\boldsymbol{E})$.

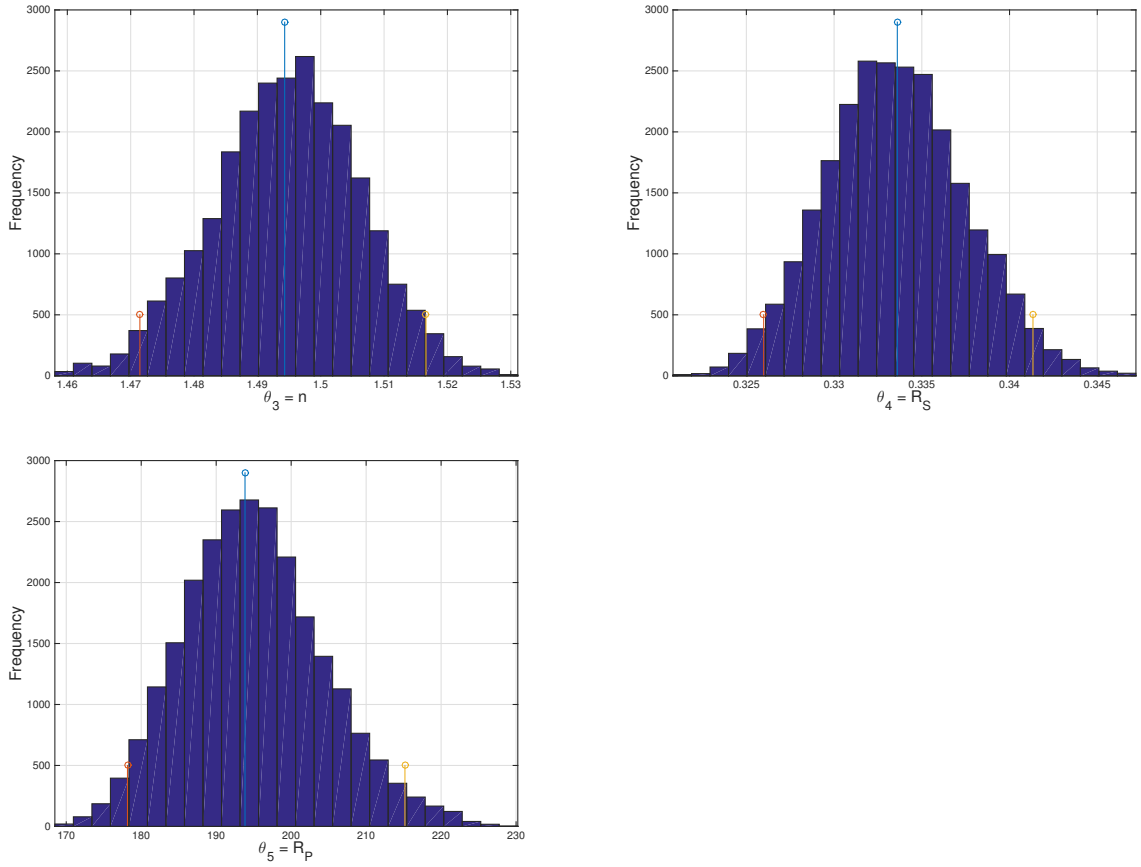


Figure 4.5: A histogram for $\pi(\theta_i|\boldsymbol{E})$ for $i \in \{3, 4, 5\}$. The short vertical lines represent the lower and upper bounds of the 95% credible interval. The tall vertical lines represents the mean of $\pi(\theta_i|\boldsymbol{E})$.

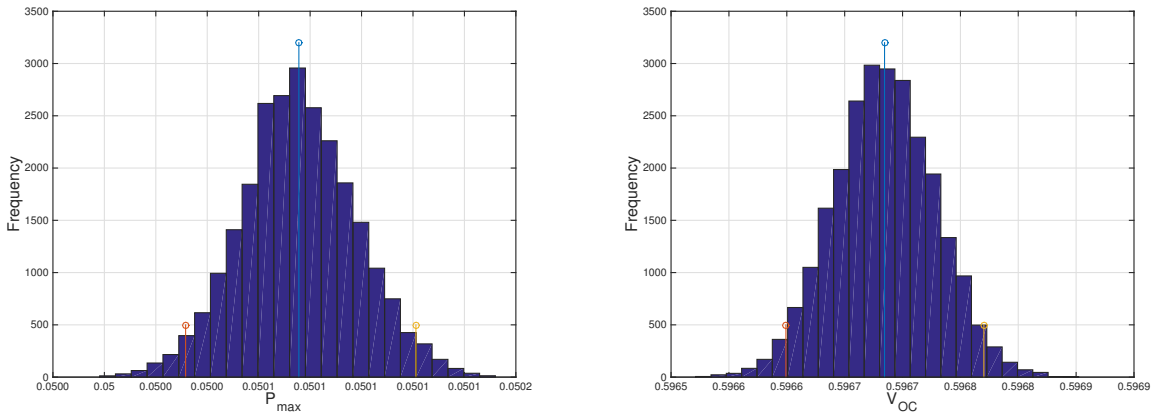


Figure 4.6: A histogram of the posterior for $P_{\max}$ and V_{OC}. The short vertical lines represent the lower and upper bounds of the 95% credible interval. The tall vertical lines represents the mean.

CHAPTER 5

DISCOVERING AN ACTIVE SUBSPACE IN A SINGLE-DIODE SOLAR CELL MODEL

A paper to be published in the Conference on Data Analysis special edition of *Statistical Analysis and Data Mining*

Paul Constantine[26], Brian Zaharatos[27], Mark Campanelli[28]

Abstract

Predictions from science and engineering models depend on the values of the model's input parameters. As the number of parameters increases, algorithmic parameter studies like optimization or uncertainty quantification require many more model evaluations. One way to combat this curse of dimensionality is to seek an alternative parameterization with fewer variables that produces comparable predictions. The *active subspace* is a low-dimensional linear subspace defined by important directions in the model's input space; input perturbations along these directions change the model's prediction more, on average, than perturbations orthogonal to the important directions. We describe a method for checking if a model admits an exploitable active subspace, and we apply this method to a single-diode solar cell model with five input parameters. We find that the maximum power of the solar cell has a dominant one-dimensional active subspace, which enables us to perform thorough parameter studies in one dimension instead of five.

5.1 Introduction

Science and engineering simulations often contain several input parameters—e.g., physical constants, boundary conditions, or geometry descriptions. When presented with such

[26]Development of active subspace methods. Department of Applied Mathematics and Statistics, Colorado School of Mines

[27]Development of PV application.

[28]Development of PV application. National Renewable Energy Lab

parameterized simulations, the scientist naturally wonders how the simulated predictions depend on the input parameters. Which parameters, when perturbed, create the largest change in predictions? How precisely must the parameters be specified to ensure accurate predictions? And what is the effect of imprecisely prescribed input parameters on the predictions? If the goal is to maximize or minimize a predicted quantity, which combinations of parameters correspond to larger or smaller values of the prediction? In a similar vein, one may ask which parameter values correspond to predictions outside a safe region of operation—or which parameter values yield predictions that are consistent with a set of observations.

Many scientists rely on intuition about the physical system to answer these questions. But intuition becomes less trustworthy as the simulations become more complex, e.g., when they include several interacting physical components. Algorithms for optimization, uncertainty quantification, and model calibration become more attractive as model complexity increases. If an algorithm can easily interface with the simulation—e.g., by automatically evaluating predictions given values for the inputs—then applying the algorithm to the simulation becomes relatively easy.

The number of times the algorithm needs to evaluate a prediction increases—sometimes extremely rapidly—as the number of inputs increases. The situation is worse if each evaluation requires significant computational resources. For example, finding the global minimum of a complicated prediction depending on 100 inputs is not tractable if available resources permit only ten model evaluations. In practice, the scientist may choose only the most important parameters to vary so the study fits within the computational budget. Alternatively, one may seek a low-dimensional description of the prediction as a function of the input parameters. If such a description is sufficiently accurate, then studies like optimization or calibration can work in the space of fewer variables—potentially allowing the desired parameter studies within the given budget.

The *active subspace* is a low-dimensional linear subspace defined by important directions in the model's input space; input perturbations along these directions change the model's

prediction more, on average, than perturbations orthogonal to the important directions. Not all models have an active subspace. Some model's predictions respond significantly to input perturbations along all directions. However, if a model does admit an active subspace, then one can exploit it to perform parameter studies in the coordinates of the subspace—i.e., the *active variables*—which are linear combinations of normalized versions of the model's input parameters. Therefore, it can be very advantageous to discover that a model admits an active subspace.

Active subspaces have been studied in a variety of contexts under different names. Cook's excellent text *Regression Graphics* [42] reviews and develops statistical methods for dimension reduction in the context of regression surfaces, and it contains references to the major works in the statistics literature. What we call the *active subspace* is a type of *dimension-reduction subspace* in Cook's parlance [42, Chapter 6]—though we are working with noiseless computer simulations in contrast to general regression surfaces. Russi's 2010 Ph.D. dissertation uses the phrase *active subspace* in a way comparable to our use [25]. He exploits the active subspace to construct quadratic surrogate models for uncertainty quantification in chemical kinetics models. Our prior work develops a theoretical framework for active subspaces including applications to kriging response surfaces [19]. We have applied these methods to several models in aerospace engineering [43, 44, 45]. Abdel-Khalik has applied similar methods in nuclear engineering applications [46].

In this paper, we describe how to test a model for an active subspace, and we apply this test to the maximum power from a single-diode model of a photovoltaic solar cell with five input parameters. In Section 5.2, we generically describe the active subspace and how to search for it. We then describe the single-diode model, its input parameters, and its predicted *performance parameters* (i.e., model outputs) in Section 5.3. We apply the tests for the active subspace to the model's maximum power in Section 5.4 and show that a dominant one-dimensional active subspace is present in the five-dimensional space of input parameters. We conclude in Sections 5.5 and 5.6 with a discussion of how one may exploit

the low-dimension of the active subpace to further study the parameter dependence in the single-diode model's maximum power.

5.2 Active subspaces

We consider a generic multivariate function $f = f(\mathbf{x})$, where $\mathbf{x}$ represents the inputs of the model, and f represents a specific scalar performance parameter that the model predicts. Let $\mathcal{X} = [-1, 1]^m$ be the domain with $\mathbf{x} \in \mathcal{X}$, and let $\rho : \mathcal{X} \to \mathbb{R}_+$ be a bounded and continuous weight function on $\mathcal{X}$; we assume ρ is normalized to integrate to 1. Assume f is diffentiable and absolutely continuous, and denote the gradient $\nabla_{\mathbf{x}} f(\mathbf{x}) = [\partial f/\partial x_1, \ldots, \partial f/\partial x_m]^T$ oriented as a column vector.

Consider the following matrix $\boldsymbol{C}$ defined as

$$\boldsymbol{C} \;=\; \int_{\mathcal{X}} (\nabla_{\mathbf{x}} f)(\nabla_{\mathbf{x}} f)^T \, \rho \, d\mathbf{x}. \tag{5.1}$$

In the context of dimension reduction for regression functions, Samarov calls this matrix an *average derivative functional* [47]. Note that we are not studying regression functions, per se. Instead, f represents a noiseless, parameterized computer simulation. The matrix $\boldsymbol{C}$ is symmetric and positive semi-definite, so it admits a real eigenvalue decomposition

$$\boldsymbol{C} \;=\; \boldsymbol{W} \Lambda \boldsymbol{W}^T, \qquad \Lambda \;=\; \operatorname{diag}(\lambda_1, \ldots, \lambda_m), \qquad \lambda_1 \geq \cdots \geq \lambda_m \geq 0. \tag{5.2}$$

We can partition the eigenvectors,

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{W}_1 & \boldsymbol{W}_2 \end{bmatrix}, \qquad \Lambda = \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix}, \tag{5.3}$$

where $\boldsymbol{W}_1$ contains the first n eigenvectors, and Λ_1 contains the n largest eigenvalues. We use the two sets of eigenvectors to create new sets of variables $\mathbf{y} = \boldsymbol{W}_1^T \mathbf{x}$ and $\mathbf{z} = \boldsymbol{W}_2^T \mathbf{x}$. We call the subspace defined by $\boldsymbol{W}_1$ the *active subspace*, and we call the variables $\mathbf{y}$ the *active variables*; the term *active subspace methods* for this type of analysis was first used in Russi's 2010 Ph.D. dissertation [25]. The following two lemmas justify these labels.

Lemma 5.1 *The mean-squared directional derivative of f with respect to the eigenvector $\mathbf{w}_i$ is equal to the corresponding eigenvalue,*

$$\int_{\mathcal{X}} \left((\nabla_{\mathbf{x}} f)^T \mathbf{w}_i\right)^2 \rho \, d\mathbf{x} \;=\; \lambda_i. \tag{5.4}$$

Lemma 5.2 *The mean-squared gradients of f with respect to the coordinates $\mathbf{y}$ and $\mathbf{z}$ satisfy*

$$\begin{aligned} \int_{\mathcal{X}} (\nabla_{\mathbf{y}} f)^T (\nabla_{\mathbf{y}} f) \, \rho \, d\mathbf{x} &= \lambda_1 + \cdots + \lambda_n, \\ \int_{\mathcal{X}} (\nabla_{\mathbf{z}} f)^T (\nabla_{\mathbf{z}} f) \, \rho \, d\mathbf{x} &= \lambda_{n+1} + \cdots + \lambda_m. \end{aligned} \tag{5.5}$$

The proofs of these lemmas can be found in our prior work [19]. In words, they mean that f changes more, on average, when its inputs are perturbed along the directions $\boldsymbol{W}_1$ than along the directions $\boldsymbol{W}_2$; the eigenvalues quantify precisely how much more. If an eigenvalue is exactly zero, then $f(\mathbf{x})$ is constant along the direction defined by the corresponding eigenvector over all of $\mathcal{X}$.

To gain some intuition, consider the extreme case where all eigenvalues are precisely zero except λ_1. Then $f(\mathbf{x}) = g(\mathbf{w}_1^T \mathbf{x})$, where $\mathbf{w}_1$ is the first column of $\boldsymbol{W}$, and g is a function of one variable. In many applications, the smaller eigenvalues are not precisely zero, but they may be much (e.g., orders of magnitude) smaller so that $f(\mathbf{x})$ may be reasonably approximated by a function of $n < m$ linear combinations of $\mathbf{x}$.

If a given model f admits such structure, then certain operations—e.g., response surface modeling or optimization—become much less expensive. In particular, these operations can be performed in the n-dimensional space of the active variables $\mathbf{y}$ instead of the full m-dimensional space. It is therefore extremely valuable to determine if f admits an active subspace. This analysis assumes that the weight function ρ on the domain $\mathcal{X}$ is given. The computed quantities (like $\boldsymbol{W}$ and Λ) will change if a different ρ is given.

5.2.1 Identifying an active subspace

To identify the active subspace, we must approximate the matrix $\boldsymbol{C}$ from (5.1). Since $\boldsymbol{C}$ is the mean of the outer product of the gradient, we can approximate it with simple Monte Carlo. One could achieve a more accurate approximation of $\boldsymbol{C}$ and its eigendecomposition with an integration rule that is more accurate than simple Monte Carlo. However, if m is

greater than two or three, then tensor product constructions of accurate univariate numerical integration rules (e.g., Gaussian quadrature) require too many evaluations of the gradient to be practical—especially if the gradient is expensive to compute. The simple Monte Carlo proceeds as follows. Draw $\mathbf{x}_i$ independently according to ρ with $i = 1, \ldots, M$. Then

$$\boldsymbol{C} \approx \hat{\boldsymbol{C}} = \frac{1}{M} \sum_{i=1}^{M} \nabla_{\mathbf{x}} f(\mathbf{x}_i) \nabla_{\mathbf{x}} f(\mathbf{x}_i)^T = \hat{\boldsymbol{W}} \hat{\Lambda} \hat{\boldsymbol{W}}^T. \tag{5.6}$$

We use the eigenvalues of the approximation $\hat{\boldsymbol{C}}$ as evidence of an active subspace. In particular, a large gap in the eigenvalues indicates a separation between the corresponding active and inactive subspaces defined by $\hat{\boldsymbol{W}}_1$ and $\hat{\boldsymbol{W}}_2$, respectively. The qualification *large* depends on the application. For example, a good low-dimensional approximation of f over its entire domain may need a much larger separation than a good approximation of the bounds or the average of f.

Other eigenvalue-based dimension reduction schemes (e.g., principal component analysis) use heuristics based on the magnitude of the eigenvalues, such as choosing the dimension of the subspace such that the ratio $\sum_{i=1}^{n} \lambda_i / \sum_{i=1}^{m} \lambda_i$ is larger than some threshold. For active subspaces, such heuristics are not well-justified. We are not interested in capturing the gradient's variance; we want to identify directions along which the function changes the most. In our recent work, we show that the distance between the n-dimensional subspace defined by $\boldsymbol{W}_1$ and its approximation defined by $\hat{\boldsymbol{W}}_1$ is inversely proportional to $\lambda_{n+1} - \lambda_n$ [24]. Therefore, the quality of the finite-sample subspace approximation depends more on the gap between the eigenvalues than their magnitude.

5.2.2 Bootstrap to estimate variability

If a gap is present in the approximate eigenvalues $\hat{\Lambda}$, one may naturally ask if a comparable gap is present in the true eigenvalues Λ. To address this question, we use nonparametric bootstrap. The bootstrap is most appropriate when the simulation is expensive, and the number M of gradient evaluations is constrained by a computational budget. For $j = 1, \ldots, M'$, let $\boldsymbol{\pi}_j = [\pi_1^j, \ldots, \pi_M^j]$ be an M-vector of integers drawn uniformly at random

between 1 and M. The jth *bootstrap replicate* of $\hat{\boldsymbol{C}}$ is computed as

$$\hat{\boldsymbol{C}}^j \;=\; \frac{1}{M}\sum_{i=1}^{M} \nabla_{\mathbf{x}} f(\mathbf{x}_{\pi_i^j})\, \nabla_{\mathbf{x}} f(\mathbf{x}_{\pi_i^j})^T \;=\; \hat{\boldsymbol{W}}^j \hat{\Lambda}^j \left(\hat{\boldsymbol{W}}^j\right)^T. \tag{5.7}$$

The collection of eigenvectors $\hat{\Lambda}_j$ yields a bootstrap distribution for the eigenvalues, which can be used to estimate boostrap intervals. Note that these intervals are not true confidence intervals, since the estimates are biased. However, this bias decreases as M increases; see [48, Section 7.2] and [49, Chapter 3] for related discussions of the bias in bootstrap estimates for principal components. Nevertheless, we use the 99% bootstrap intervals as evidence of gaps in the estimated eigenvalues.

We can also use the bootstrap replicates to estimate the error in the estimated subspace. Partition the eigenvectors from the bootstrap replicate as in (5.3). Define

$$e_j \;=\; \|\hat{\boldsymbol{W}}_1 \hat{\boldsymbol{W}}_1^T - (\hat{\boldsymbol{W}}_1^j)(\hat{\boldsymbol{W}}_1^j)^T\|_2 \;=\; \|\hat{\boldsymbol{W}}_1^T \hat{\boldsymbol{W}}_2^j\|, \tag{5.8}$$

which is the distance between the subspace defined by $\hat{\boldsymbol{W}}_1$ and bootstrap replicate $\hat{\boldsymbol{W}}_1^j$ [50]. The mean and 99% bootstrap intervals of the set $\{e_j\}$ quantify the varibility in the estimated active subspace.

Our current research efforts are focused on rigorously justified, computable metrics for determining the relationship between the approximate eigenpairs $\hat{\Lambda}$, $\hat{\boldsymbol{W}}$ and the true eigenpairs Λ, $\boldsymbol{W}$. Our recent paper applies non-asymptotic random matrix theory to characterize the error in these approximations [24].

5.2.3 Visualization with the active subspace

Scatter plots are a common way to visualize data sets in search of a trend. Unfortunately, visualization tools can only display scalar responses f as a function of at most two variables. When the response depends on more than two variables, one can plot responses versus each variable or each pair of variables.

Using the vectors defining the active subspace, we can create scatter plots based on the active variables, which often show discernible trends that can be exploited in further studies

(e.g., building response surfaces or optimization). This idea is described in detail in the texts on *regression graphics* [42], where such plots are called *sufficient summary plots*. For example, suppose we have noticed a large gap between the first and second eigenvalue. We can confirm the presence of the active subspace by first sampling $\mathbf{x}_i$ from ρ as in (5.6), computing $f_i = f(\mathbf{x}_i)$ (often computed along with the gradients in (5.6)), and plotting the pairs $(\mathbf{w}_1^T\mathbf{x}_i, f_i)$—where $\mathbf{w}_1$ is the first column of $\boldsymbol{W}$. If a tight, univariate trend is visibly present, then this verifies the active subspace. Figure 5.4 in Section 5.4 shows these scatter plots for the single-diode model.

These sufficient summary plots can also be used to examine the variability in the vectors defining the active subspace. In particular, we can use the bootstrap replicates $\hat{\boldsymbol{W}}^j$ from (5.7). The first eigenvector from each can be used to create pairs for the scatter plot. Plotting all pairs together creates clusters in the scatter plot that can indicate the spread around a trend. We create these visualizations for the predicted performance parameter from the single-diode model in Section 5.4.

5.2.4 Gradient approximation

Often the gradient $\nabla_{\mathbf{x}} f$ is not available or is too complicated to compute. To approximate partial derivatives, one builds a model of $f(\mathbf{x})$ that is easily differentiable, e.g., a polynomial model. A first-order finite difference approximation of the partial derivative with respect to x_i at $\mathbf{x}$ computes the slope of the plane that interpolates f at $\mathbf{x}$ and $\mathbf{x} + \varepsilon\mathbf{e}_i$, where $\mathbf{e}_i$ is a vector of zeros with a one in the ith component. When the variables are properly normalized, a reasonable choice for ε to approximate the derivative and avoid round-off issues is the square root of machine precision [51, Section 8.1].

If the simulation output is well-behaved as a function of the variables $\mathbf{x}$, then finite difference approximations of gradients can be used as a substitute for the true gradient. Each finite difference approximation requires $m + 1$ evaluations of f, so the number of evaluations to compute $\hat{\boldsymbol{C}}$ is $M(m + 1)$. Therefore, finite differences are most appropriate when the following are satisfied:

- the simulation output behaves well as a function of the parameters, i.e., there is no noise due to limited iterations of a nonlinear solver or large changes of the function on scales smaller than the finite difference step size,

- the simulation completes in a short enough time on available computing resources to permit $M(m+1)$ evaluations.

Both of these conditions are satisfied for the performance parameter of interest from the single-diode model as a function of its input parameters. Therefore, we use finite differences in place of the true gradient to search for an active subspace.

If these conditions are not satisfied, then one must pursue other options for estimating the gradients. One idea is to use local linear models within a budget-constrained set of runs. In short, assume one has B pairs $(\mathbf{x}_j, f(\mathbf{x}_j))$. Given a point $\mathbf{x}$ in the parameter space, select M points in $\{\mathbf{x}_j\}$ near $\mathbf{x}$, fit a linear model to the selected pairs, and estimate the gradient as the gradient of the linear model. M must be large enough and the points must be chosen to enable fitting the linear model. We have used this approach in practice, but we are still analyzing its properties. The advantage over finite differences is that it is robust to noise in $\{f_j\}$ when the number of points used to fit the linear model is greater than the number of variables ($M > m+1$). However, if the points $\{\mathbf{x}_j\}$ are not sufficiently dense in the parameter space, then the gradient estimates can be very poor, which lead to poor estimates of the eigenvalues and eigenvectors. We continue to analyze this approach.

5.3 A single-diode solar cell model

In this section we introduce the single-diode solar cell model, the input parameters, and the predicted quantity of interest. Over the past decade, a growing demand for clean energy has caused rapid growth in the photovoltaic (PV) industry. The overall health of the industry depends on proper characterization of the risks associated with a PV system. One can characterize such risks using a mathematical model to describe the performance of the PV device, such as a single-diode lumped-parameter equivalent-circuit model defined in Section

5.3.1. For many series-wired PV devices, the single-diode model accurately describes the device's current-voltage (I-V) characteristics at given irradiance and temperature conditions. The single-diode model contains several input parameters that must be estimated—e.g., with methods in [11] or [8]—to properly pose the model. With the inputs fixed, practitioners can estimate the *key performance parameters*, such as the maximum power output P_{max} and energy conversion efficiency η.

Estimating the model's input parameters presents challenges. For example, many single-diode model parameter estimation methods have been proposed in the PV literature, and it is not always clear if any particular estimation yields a unique result [6]. Further, since each key performance parameter is a function of the model parameters, errors in estimating the latter will cause issues in estimating the former.

Given these issues, it is desirable to know how the variability or uncertainty in each of the input parameters affects variability in the outputs. In particular, it would be useful to know whether the model admits an active subspace. If an active subspace is present, then estimation of the performance parameters is greatly simplified. For example, rather than conducting current-voltage measurements in a way that yields accurate estimates of all model parameters, practitioners could instead focus on measurements that accurately estimate parameters in the active subspace.

5.3.1 Calculating the key performance parameters

The single-diode model describes a relationship between the current (I) and voltage (V) in the single diode. There are multiple variants of the single-diode model, each of which has accompanying auxiliary equations that describe how components depend on irradiance and temperature. In this work, we use the auxiliary equations derived in [29]; see [28] for an example of others. These relationships are defined in the following equations.

$$I \;=\; I_{\text{L}} - I_{\text{S}} \left(\exp\left(\frac{V + I\,R_{\text{S}}}{N_S\, n\, V_{\text{th}}} \right) - 1 \right) - \frac{V + I\,R_{\text{S}}}{R_{\text{P}}}. \tag{5.9}$$

N_S is the number of cells connected in series, which we set to 1. The thermal voltage $V_{\rm th}$ is a given, fixed constant for fixed temperature $T = 25°$C. An auxiliary equation defines the photocurrent $I_{\rm L}$ as

$$I_{\rm L} \;=\; I_{\rm SC} + I_{\rm L}\left(\exp\left(\frac{I_{\rm SC} R_{\rm S}}{N_S\, n\, V_{\rm th}}\right) - 1\right) + \frac{I_{\rm SC}\, R_{\rm S}}{R_{\rm P}}. \tag{5.10}$$

The remaining terms are input parameters to the single-diode model; their names, units, and ranges are given in Table 5.1. The ranges correspond to typical 2cm^2 crystalline silicon PV cells.

Table 5.1: The ranges of the input parameters for the single-diode model in (5.9) and (5.10). These ranges correspond to typical 2cm^2 crystalline silicon PV cells.

Parameter	Name	Lower bound	Upper bound	Units
$I_{\rm SC}$	short-circuit current	0.05989	0.23958	amps
$I_{\rm S}$	diode reverse saturation current	2.2e-11	2.2e-7	amps
n	ideality factor	1	2	unit-less
$R_{\rm S}$	series resistance	0.16625	0.66500	ohms
$R_{\rm P}$	parallel (shunt) resistance	93.75	375.00	ohms

Given a device and I-V data, one can estimate precise values for the input parameters $(I_{\rm SC}, I_{\rm S}, n, R_{\rm S}, R_{\rm P})$; see [11, 8, 29]. With these values fixed, one can compute the maximum power of the device as

$$P_{\rm max} \;=\; \max_{I,V} \; I\, V, \tag{5.11}$$

where current I and voltage V are constrained by (5.9) and (5.10). Changing the inputs $(I_{\rm SC}, I_{\rm S}, n, R_{\rm S}, R_{\rm P})$ changes the nonlinear constraints in the optimization (5.11), thus changing $P_{\rm max}$. Therefore, we can write

$$P_{\rm max} \;=\; P_{\rm max}(I_{\rm SC}, I_{\rm S}, n, R_{\rm S}, R_{\rm P}). \tag{5.12}$$

We use MATLAB to solve the optimization (5.11) subject to the constraints (5.9) and (5.10). Figure 5.1 shows ten I-V curves with the single-diode model parameters chosen uniformly at random from the ranges in Table 5.1. The black dots show the values of $I_{\rm max}$ and $V_{\rm max}$ that produce $P_{\rm max}$. A wrapper function computes $P_{\rm max}$ as a function of the input parameters.

Requests for the MATLAB code may be sent to the third author Campanelli.

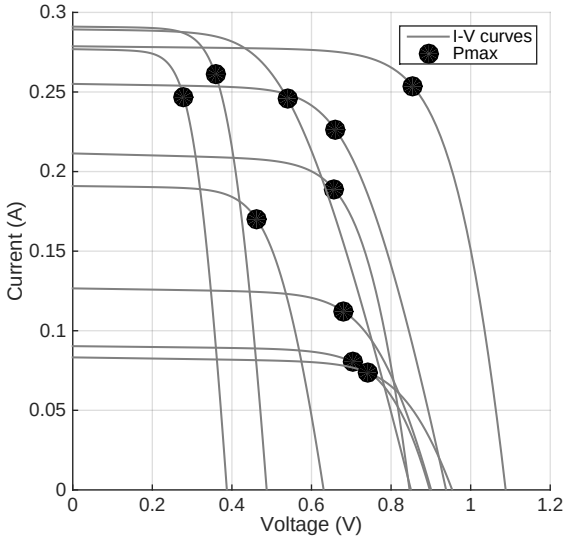


Figure 5.1: Ten current-voltage curves with parameters chosen uniformly at random from the ranges in Table 5.1. The black circles plot the location of $P_{\max} = I_{\max} V_{\max}$.

5.4 An active subspace in the single-diode model

We apply the techniques from Section 5.2 to search for an active subspace of the $P_{\max}$ performance parameter as a function of the parameters defined in Table 5.1. Note that a different performance parameter (e.g., efficiency) would require an independent and identical analysis. The relationship between the respective active subspaces depends on the relationship between the performance parameters and their inputs. In some cases, the active subspaces might be similar; in other cases, they may differ substantially.

The MATLAB scripts for generating the figures in this section can be found at `https://bitbucket.org/paulcon/active-subspaces-in-a-single-diode-model`. To generate the figures and results, we evaluated the model $(m+1)M = 6000$ times, which took approximately three minutes on a MacBook Air. A more expensive model would permit fewer evaluations, and we would choose M differently—or perhaps choose a different method.

5.4.1 Normalizing the input parameters

We first normalize the domain of $P_{\max}$ to the hypercube; denote the normalized variables by $\mathbf{x} = [x_1, x_2, x_3, x_4, x_5]^T$. The variable I_{S} varies over several orders of magnitude, and preliminary tests show that $P_{\max}$ changes rapidly near smaller values of I_{S}. To address this, we work with $\log(I_{\mathrm{S}})$. The range of $\log(I_{\mathrm{S}})$ is bounded below by -24.54 and above by -15.32.

Let $\mathcal{L}(\cdot)$ and $\mathcal{U}(\cdot)$ return the upper and lower bounds, respectively, of the argument. We define the normalized input parameters as

$$\begin{aligned}
x_1 &= 2\left(\frac{I_{\mathrm{SC}} - \mathcal{L}(I_{\mathrm{SC}})}{\mathcal{U}(I_{\mathrm{SC}}) - \mathcal{L}(I_{\mathrm{SC}})}\right) - 1 \\
x_2 &= 2\left(\frac{\log(I_{\mathrm{S}}) - \mathcal{L}(\log(I_{\mathrm{S}}))}{\mathcal{U}(\log(I_{\mathrm{S}})) - \mathcal{L}(\log(I_{\mathrm{S}}))}\right) - 1 \\
x_3 &= 2\left(\frac{n - \mathcal{L}(n)}{\mathcal{U}(n) - \mathcal{L}(n)}\right) - 1 \\
x_4 &= 2\left(\frac{R_{\mathrm{S}} - \mathcal{L}(R_{\mathrm{S}})}{\mathcal{U}(R_{\mathrm{S}}) - \mathcal{L}(R_{\mathrm{S}})}\right) - 1 \\
x_5 &= 2\left(\frac{R_{\mathrm{P}} - \mathcal{L}(R_{\mathrm{P}})}{\mathcal{U}(R_{\mathrm{P}}) - \mathcal{L}(R_{\mathrm{P}})}\right) - 1
\end{aligned} \tag{5.13}$$

Then $\mathbf{x} \in [-1, 1]^5$ and has no units. Note that this map is invertible. In other words, given a value $\mathbf{x} \in [-1, 1]^5$, one can shift and scale the components of $\mathbf{x}$—and transform the second component with the exponential—to produce a valid input for the single-diode model. We take the weight function ρ in (5.1) to be a constant 2^{-5} inside $[-1, 1]$ and zero elsewhere. One can interpret this weight function as a uniform probability density on the space of input parameters.

Proper scaling is important. The results of the active subspaces analysis are not scale invariant, so the scientist must carefully choose input parameter bounds that are appropriate for the application. Changing the bounds can change the results—sometimes dramatically. For example, the quantity of interest may be irregular and badly behaved over a wide range of inputs. But reducing the range of interest might focus on a small region in the parameter space where f is smooth—even nearly linear. A linear function has a one-dimensional active subspace. Thus, loosely speaking, if the range of parameters is sufficiently small, then the

active subspace will be nearly one-dimensional for continuous and differentiable quantities of interest.

5.4.2 Estimating the eigenvalues and eigenvectors

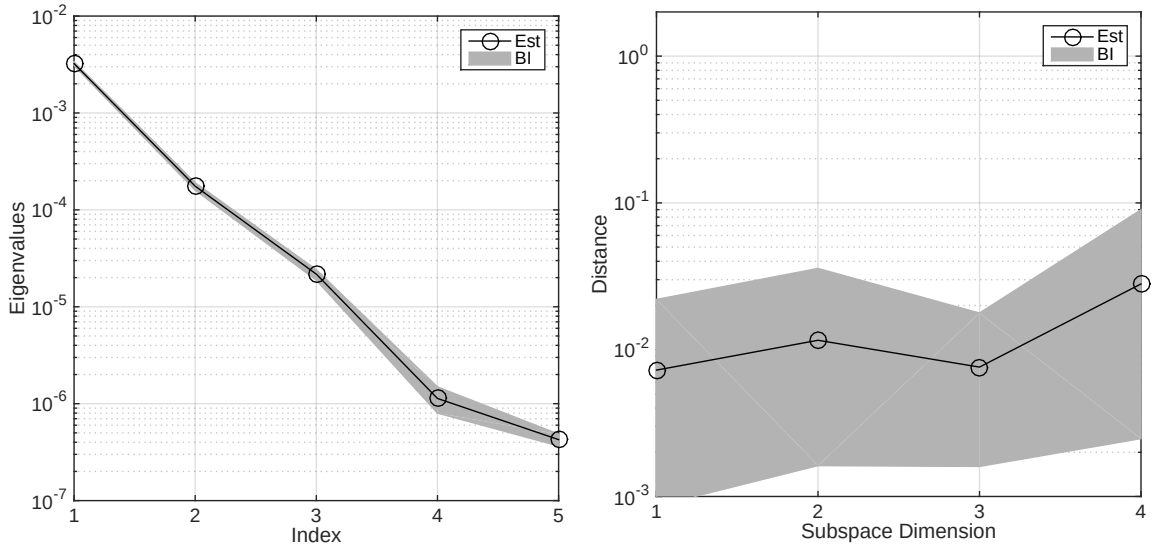
We choose $M = 1000$ points $\mathbf{x}_i$ uniformly at random from $[-1, 1]^5$, and for each $\mathbf{x}_i$ we compute both $P_{\max}$ and a first-order finite difference approximation of the gradient. We use a finite difference step size of 10^{-6} in the normalized domain, which was small enough to ensure that gradient approximation errors were substantially smaller than the variance due to random sampling.

To study variability in the computed components of $\hat{\boldsymbol{C}}$ from (5.6), we use the bootstrap described in Section 5.2.2 with $M' = 1000$ bootstrap replicates. The 99% boostrap intervals for the eigenvalue estimates are shown in Figure 5.2(a). The tight bootstrap ranges suggest there is little eigenvalue variability in the bootstrapped data set. There is a gap of nearly an order of magnitude between both the first and second and the second and third eigenvalues. This suggests a very dominant one-dimensional active subspace. Figure 5.2(b) shows the estimated error in the approximated subspace using (5.8). Note that this error is actually the cosine of the principal angle between the subspaces, so it is bounded above by 1.

Figure 5.3 displays the components of the first and second eigenvectors of $\hat{\boldsymbol{C}}$ computed with $M = 1000$ samples—along with the $M' = 1000$ bootstrap replicates. They are normalized so that the first component is positive. The small ranges suggest that the one- and two-dimensional active subspaces are stable within the gradient samples.

5.4.3 Identifying the active subspace

The order-of-magnitude gap between the first and second eigenvalues suggests a dominant one-dimensional active subspace. Recall that gaps between eigenvalues are more important than the magnitudes of the eigenvalues. Figure 5.4(a) plots 100 values of $P_{\max}$ against a linear combination of the corresponding normalized inputs; the weights of the linear combination are the components of the first eigenvector $\hat{\mathbf{w}}_1$ from (5.6). Figure 5.4(a) is a one-dimensional



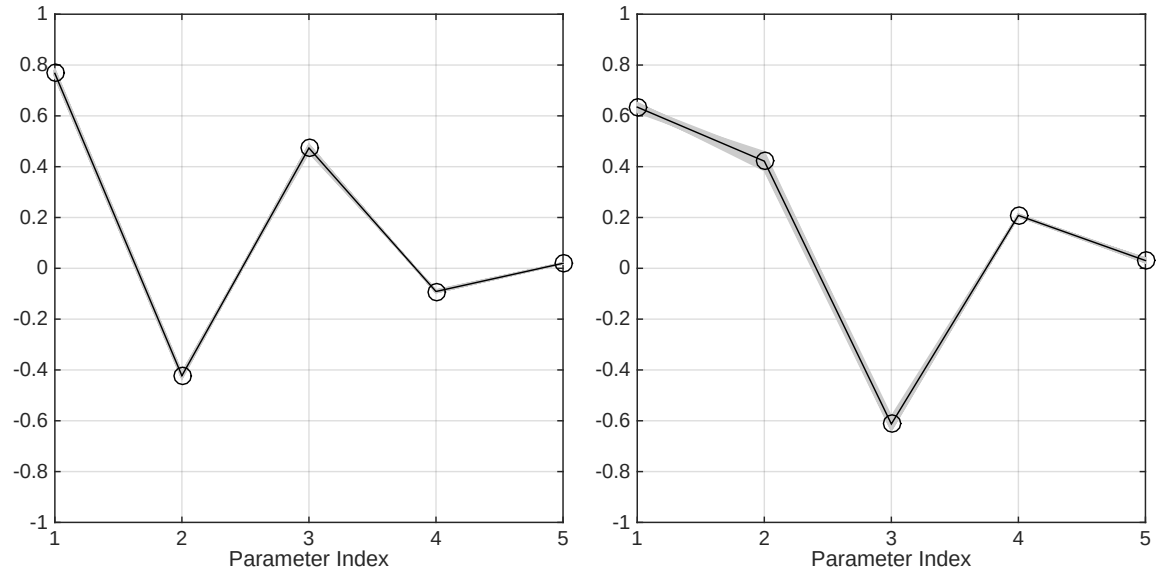
(a) Eigenvalues (b) Subspace error

Figure 5.2: The left figure shows the eigenvalues of the matrix $\hat{\boldsymbol{C}}$ from (5.6) with $M = 1000$ samples and their bootstrap ranges with $M' = 1000$ boostrap replicates. The right figure shows the mean and ranges of the bootstrap estimates of the subspace error (see (5.8)) with $M' = 1000$ bootstrap replicates.

sufficient summary plot, as described in [42]. The plot shows a potentially exploitable low-dimensional relationship between the input parameters and the quantity of interest, $P_{\max}$.

Figure 5.4(b) shows the two-dimensional sufficient summary plot. The grayscale corresponds to the value of $P_{\max}$. The black dot in the center of each cluster has a horizontal and vertical location determined by linear combinations of $P_{\max}$'s corresponding input parameters; the weights of the linear combination are the eigenvectors $\hat{\mathbf{w}}_1$ and $\hat{\mathbf{w}}_2$, respectively. Each cluster corresponds to the value of $P_{\max}$ with horizontal and vertical position determined by 20 bootstrap replicates of $\hat{\mathbf{w}}_1$ and $\hat{\mathbf{w}}_2$. The relatively small spread in the clusters indicates that the estimated subspace is stable.

In Figure 5.4(b), note that $P_{\max}$ increases primarily as a function of the first active variable. This provides more evidence of the dominance of the one-dimensional active subspace.



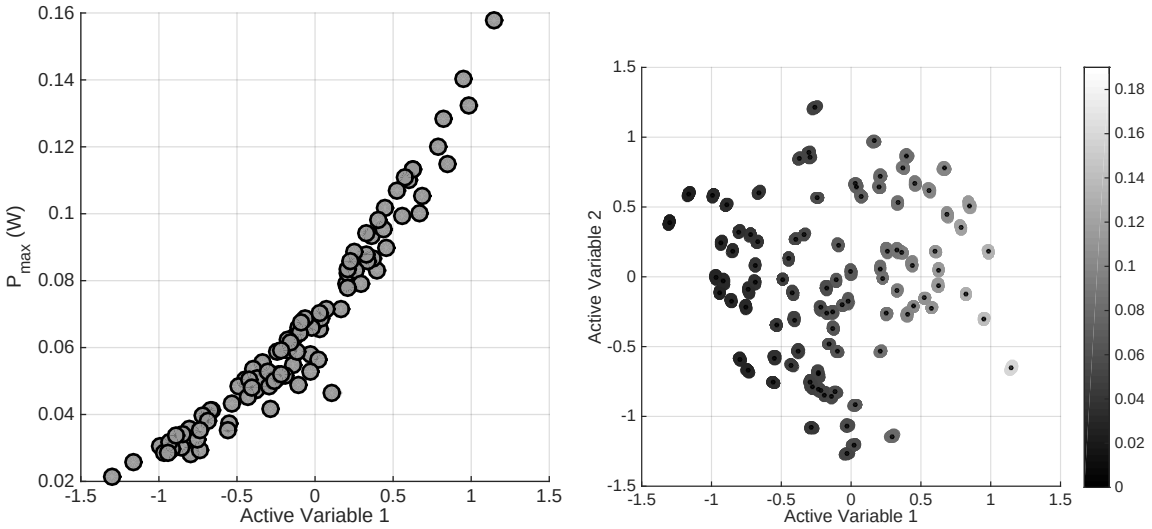
(a) First eigenvector (b) Second eigenvector

Figure 5.3: These figures show the components of the first (left) and second (right) eigenvectors of $\hat{\boldsymbol{C}}$ with $M = 1000$ gradient samples. The grey shaded region superimposes the $M' = 1000$ bootstrap replicates of the eigenvector components; the tight ranges suggests stability in the gradient samples. These values can be used as measures of sensitivity of $P_{\max}$ with respect to the input variables in the model, whose names are given along the horizontal axis of the figures. See Table 5.1 for a description of the variables.

5.4.4 Eigenvectors for sensitivity analysis

The eigenvector components connect the active subspace to the (normalized) variables in the model; Figure 5.3 includes the corresponding labels in the horizontal axis. The magnitudes of the eigenvector components can be used as measures of relative sensitivity for each of the parameters in the model. A large absolute value of the eigenvector component implies that this variable is important in defining the direction along which input perturbations produce the most change, on average, in $P_{\max}$. The components corresponding the parallel resistance R_{P} and series resistance R_{S} are close to zero, which implies that (normalized) changes in R_{P} and R_{S} do not change $P_{\max}$ as much as changes in the other parameters.

We compare the components of $\hat{\mathbf{w}}_1$, which defines the dominant one-dimensional active subspace, to the Sobol' first-order sensitivity indices and total sensitivity indices. We im-



(a) Projected with first eigenvector (b) Projected with first two eigenvectors

Figure 5.4: One-dimensional (left) and two-dimensional sufficient summary plots of 100 realizations of $P_{\max}$ using $\hat{\mathbf{w}}_1$ (left) and both $\hat{\mathbf{w}}_1$, $\hat{\mathbf{w}}_2$ (right). The clusters of points on the right correspond to 20 bootstrap replicates of $\hat{\mathbf{w}}_1$ and $\hat{\mathbf{w}}_2$. The strong trend in Figure 5.4(a) is verified by looking at the variation along *Active Variable 1* (the horizontal axis) in Figure 5.4(b), and the combination of these provide evidence of a one-dimensional active subspace.

plemented the method to estimate these indices from Sudret [52] using a tensor product Gaussian quadrature rule with 32768 points in 5 dimensions. Table 5.2 displays the eigenvector components and the Sobol' indices.

Table 5.2: The components of the eigenvector $\hat{\mathbf{w}}_1$, the Sobol' first-order indices, and the Sobol' total sensitivity indices for the single-diode model parameters. The metrics provide same ranking of importance for the parameters.

Parameter	$\hat{\mathbf{w}}_1$	Sobol' first-order index	Sobol' total sens. index
I_{SC}	0.77	0.56	0.61
I_{S}	-0.42	0.17	0.19
n	0.47	0.21	0.25
R_{S}	-0.09	0.01	0.01
R_{P}	0.02	0.00	0.00

There are several things to note in Table 5.2. First, the importance ranking from the Sobol' indices and the ranking from the eigenvector component magnitudes are the same. However, the signs of eigenvector components indicate whether $P_{\max}$ will increase or decrease,

on average, with changes in the corresponding parameter. (Eigenvectors are unique up to a sign, so these signs should be considered relative to one another.) There is no such interpretation in the Sobol' indices. In other words, the eigenvector components provide more information about the relationship between the corresponding parameters and $P_{\max}$.

The numeric values are difficult to compare, since they are normalized differently. The Sobol' indices are divided by the estimated total variance in $P_{\max}$, whereas the eigenvector components are normalized to have Euclidean norm equal to 1. In this case, they are the same order of magnitude.

We emphasize the different interpretations of these numbers. The eigenvector $\hat{\mathbf{w}}_1$ is a single direction in the input space. Input perturbations along this direction change $P_{\max}$ more, on average, than perturbations orthogonal to this direction. The Sobol' indices indicate the proportion of $P_{\max}$'s total variance attributable to the factors of a variance-based decomposition of $P_{\max}$. In terms of dimension reduction, one might (i) use the Sobol' indices to conclude that only three of the five parameters were important and (ii) approximate $P_{\max}$ as a function of those three parameters. In contrast, the active subspace is used to approximate $P_{\max}$ as with a univariate function of a linear combination of all five parameters. In other words, the variance-based approach reduces the dimension from 5 to 3, and the active subspace approach reduces the dimension from 5 to 1.

There are other challenges facing the active subspace style of dimension that do not affect the variance-based approach. The variance-based approach uses a subset of the model's parameters—as opposed to a subspace. Fixing a subset of parameters at nominal values and allowing the others to vary is relatively straighforward. Exploiting one important linear combination of the normalized parameters is not as straightforward. In the next section we discuss some possibilities for exploiting the active subspace.

5.4.5 Interpreting the sensitivity analyses

We can interpret the consistent importance rankings in the sensitivity metrics in terms of the physics of the single-diode model. First, $P_{\max} = I_{\max}V_{\max}$, where $(I_{\max}, V_{\max})$ is the point

on the I-V curve that maximizes (5.11). $I_{\max}$ typically scales linearly with I_{SC}. The voltage at $I = 0$—denoted V_{oc} for *voltage at open circuit*—typically scales logarithmically with I_{SC}. $V_{\max}$ behaves like V_{oc}. Thus, increases in I_{SC} affect $I_{\max}$ linearly and $V_{\max}$ logarithmically, which changes $P_{\max}$ as the product of a linear term and a logarithmic term. Since I_{SC} affects both $I_{\max}$ and $V_{\max}$, it is natural for it to be the driving parameter in $P_{\max}$.

Recall that these sensitivity analyses are not scale invariant; the input parameter ranges and the weight function affect the results. The ranges for n and I_{S} are at about their maximum extent for a crystalline silicon PV cell, while the ranges for R_{S} and R_{P} are somewhat restricted. All of these parameters affect the shape and/or onset of the *knee* of the I-V curve, which determines the location of $(V_{\max}, I_{\max})$ and, thus, $P_{\max}$. Here R_{S} and R_{P} are relatively unimportant because their ranges are relatively smaller than the ranges for n and I_{S}.

5.5 Using the active variables

The estimated eigenvalues in Figure 5.2(a) and the plots in Figure 5.4(a) and Figure 5.4(b) provide strong evidence for the presence of a dominant one-dimensional active subspace in the $P_{\max}$ performance parameter computed from the single-diode model as a function of the model's five input parameters. The natural question is how one can exploit the dimension reduction afforded by the active subspace. There are five types of studies that benefit greatly from fewer input parameters. To keep the scope of this paper limited, we do not address the details of any of these studies, and we prefer to reserve them for future exploration.

Visualization. The active subspace enables one to view the model output's dependence on its inputs with standard computer graphics tools when the active subspace is not more than two-dimensional. Sufficient summary plots, such as those in s Figure 5.4(a) and Figure 5.4(b) can provide insights to modelers seeking to improve their models.

Optimization. Suppose one wanted to maximize $P_{\max}$ over the input variables. The plot in Figure 5.4 shows a monotonic trend in the univariate function of the first active variable. For such functions, maximization is trivial; simply make the first active variable $\hat{\mathbf{w}}_1^T\mathbf{x}$ as

large as possible subject to the constraints $\mathbf{x} \in [-1, 1]^m$. In general, a five-dimensional global optimization where the input/output relationship is not well-understood is a difficult problem. For the $P_{\max}$ output from the single-diode model, the active subspace provides a way to discover the location of the global optimum with ease.

Response surfaces. If the single-diode model were expensive to evaluate, then one may wish to construct a response surface that approximates the map from inputs to outputs. A one-dimensional active subspace allows one to build a response surface on only the active variable instead of the five model input parameters. Constructing a response surface in one variable is certainly preferred to constructing one in five variables. Real-time control systems such as maximum power point trackers might benefit from the characterization of a simple, low-dimensional response surface. In addition, the visualization tools give one confidence that the response is sufficiently smooth with respect to the active variable to permit an accurate response surface.

Averages. One may wish to compute an average of $P_{\max}$ over all five input variables. Since the average of $P_{\max}$ over its inputs is the same as the average of the conditional expectation given the active variable, we need only to approximate a marginal density of the active variable to compute the average. We are working on the specifics of this idea, but the goal would be to transform a five-dimensional integral into a one-dimensional integral requiring many fewer evaluations of $P_{\max}$.

Design. Suppose a modeler wanted to design a solar cell with $P_{\max}$ in a specified range and sought an appropriate range of values for the model inputs. A one-dimensional active subspace makes this query much easier—especially if $P_{\max}$ is monotonic with respect to the active variable as in Figure 5.4(b).

5.6 Conclusions

We have discussed methods for discovering the directions in a model's input space that change the model's prediction the most, on average. The active subspace is the span of these directions. For global sensitivity analysis, the active subspace offers an alternative to

variance decomposition techniques such as Sobol' indices. We have applied this procedure to a single-diode solar cell model with key performance parameter $P_{\max}$, and we discovered a dominant one-dimensional active subspace. We offered several possible ways to exploit the knowledge of this low-dimensional parametric dependence to gain greater insight into the model.

CHAPTER 6

CONCLUSIONS

The research presented in this thesis has been driven by the belief that, in addition to being personally fulfilling, ones work ought to foster some ends that are important to the global community. PV performance modeling plays a modest but important role in furthering our ability to increase the proportion of solar energy in our energy portfolio. A higher proportion of solar energy in our energy portfolio is important if one is concerned with meeting future energy needs and reducing or reversing the effects of climate change. This thesis contributes to PV performance modeling in at least two ways. First, the work in Chapter 2 suggests that, in estimating the single-diode model parameters and other key performance parameters, one could rely on and exploit a low dimensional structure for more reliable estimates. That is, the five-dimensional optimization problem constructed from the single-diode model in Chapter 2 may be reducible to one (or perhaps two) dimensions. Such dimension reductions could be helpful for increasing accuracy in new or currently available PV modeling software packages, such as PVsyst [18].

Second, PV devices are typically rated by manufactures at SRC (1000 W/m^2 under the standard spectrum defined in ASTM G173-03 [3] and $T = 25^\circ$ C); but in the field, PV devices operate at a range of temperature, irradiance, and spectra. The framework provided in this thesis can be thought of as making some important initial steps in constructing estimation procedures that allow for ratings away from SRC. For example, the single-diode model auxiliary equations described in Chapter 3 and Chapter 4 model PV performance at operating conditions that deviate slightly from an irradiance of 1000 W/m^2. Further, through statistical techniques, we are able to model the noise in the data to quantify the uncertainty in parameter estimates. Such techniques can be extended to other models, such as the Desoto model [28], which describes PV performance at a wider range of operating

conditions.

There is quite a bit of potential for future work in the areas of dimension reduction and statistical modeling of PV device performance. First, further advancements in active subspace research will suggest more ways that the low dimensional structure in the constructed objective function can be exploited to further PV performance parameter estimation methods. Second, the active subspace analysis conducted in Chapter 2 could be repeated on different choices for objective functions. Such choices could come from different sets of single-diode model auxiliary equations or from different loss functions (e.g., the loss functions described in Chapter 3 and Chapter 4).

Third, we note that the statistical methods used in Chapter 3 and Chapter 4, and in particular, the assumed normal noise model for E, are in need of validation. Because the single-diode model is nonlinear in its parameters, validation of the noise model for E is nontrivial, but the methods in [53] may be of use. Fourth, the statistical model can be generalized to account for measurement noise in I and V. Finally, it would be useful to consider uncertainty caused by model discrepancy in the current model and future models.

REFERENCES CITED

[1] Life Cycle Greenhouse Gas Emissions from Solar Photovoltaics . `http://www.nrel.gov/docs/fy13osti/56487.pdf`, 012. [Online; accessed 25-August-2014]".

[2] David J. Hurlbut, Joyce McLaren, and Rachel Gelman. *Beyond Renewable Portfolio Standards: An Assessment of Regional Supply and Demand Conditions Affecting the Future of Renewable Energy in the West*. 2013.

[3] ASTM Standard G173–03. *Standard Tables for Reference Solar Spectral Irradiances: Direct Normal and Hemispherical on* 37° *Tilted Surface*. ASTM International, West Conshohocken, PA, USA, 2012.

[4] Clayton, Paul R. *Fundamentals of Electric Circuit Analysis*. John Wiley & Sons, 2001.

[5] Floyd, Thomas L. *Electronics Fundamentals: Circuits, Devices & Applications*. Prentice Hall, 2009.

[6] C.W. Hansen, A. Luketa-Hanlin, and J.S. Stein. Sensitivity of single diode models for photovoltaic modules to method used for parameter estimation. *28th European Photovoltaic Solar Energy Conference and Exhibition*, pages 3258 – 3264, 2013.

[7] Adelmo Ortiz-Conde, Francisco J Garca-Snchez, Juan Muci, and Andrea Sucre-Gonzlez. A review of diode and solar cell equivalent circuit model lumped parameter extraction procedures. *Facta Universitatis Series: Electronics and Energetics*, 27 (1):57–102, 2014.

[8] G. Farivar and B. Asaei. Photovoltaic module single diode model parameters extraction based on manufacturer data sheet parameters. *Power and Energy (PECon)*, 2010 IEEE International Conference on, 929–934. http://dx.doi.org/10.1155/2013/230771.

[9] Aaron P. Dobos. An improved coefficient calculator for the california energy commission 6 parameter photovoltaic module model. *Journal of Solar Energy Engineering*, 134, 2012.

[10] S.K. Datta, K. Mukhopadhyay, S. Bandopadhyay, and H. Saha. An improved technique for the determination of solar cell parameters. *Solid State Electron*, 35:1667–1673, 1992.

[11] C.W. Hansen, A. Luketa-Hanlin, and J.S. Stein. Estimation of parameters for single diode models using measured iv curves. *39th IEEE Photovoltaics Specialists Conference*, 2013.

[12] M.R. AlRashidi, K.M. El-Naggar, and M.F. AlHajri. Parameters estimation of double diode solar cell model. *World Academy of Science, Engineering and Technology*, 74:100–103, 2013.

[13] M. Chegaar, Z. Ouennoughi, and A. Hoffmann. A new method for evaluating illuminated solar cell parameters. *Solid State Electronics*, 45:293–296, 2001.

[14] J.A. Jervase, H. Bourdoucen, and A. Al-Lawati. Solar cell parameter extraction using genetic algorithms. *Measurement Science and Technology*, 12:1922–1925, 2001.

[15] M.R. AlRashidi, M.F. AlHajri, K.M. El-Naggar, and A.K. Al-Othman. A new estimation approach for determining the i–v characteristics of solar cells. *Solar Energy*, 85:1543–1550, 2011.

[16] S. R. Lele, B. Dennis, and F. Lutscher. Data cloning: easy maximum likelihood estimation for complex ecological models using bayesian markov chain monte carlo methods. *Ecology Letters*, 10:551–563, 2007.

[17] California Energy Commission. CEC PV Module Database, 2014. [Online; accessed 22-April-2014].

[18] PVsyst. `http://www.pvsyst.com`.

[19] Paul G Constantine, Eric Dow, and Qiqi Wang. Active subspace methods in theory and practice: applications to kriging surfaces. *SIAM Journal of Scientific Computing*, 2014. In press.

[20] Constantine, Paul G. *Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies*. SIAM, In Press.

[21] Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global Sensitivity Analysis: The Primer*. John Wiley & Sons, 2008.

[22] Max D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991.

[23] F. Campolongo, J. Cariboni, and A. Saltelli. An effective screening design for sensitivity analysis of large models. *Environmental Modeling and Software*, 22:1509–1518, 2007.

[24] Paul G Constantine and David Gleich. Computing active subspaces. *arXiv preprint arXiv:1408.0545*, 2014.

[25] Trent M. Russi. *Uncertainty quantification with experimental data and complex system models*. PhD thesis, UC Berkeley, 2010.

[26] Hongmei Tian, Fernando Mancilla-David, Kevin Ellis, Eduard Muljadi, and Peter Jenkins. A cell-to-module-to-array detailed model for photovoltaic panels. *Solar Energy*, 86:26952706, 2012.

[27] Jieming Ma, Ka Lok Man, T.O. Ting, Nan Zhang, Sheng-Uei Guan, and Prudence W.H. Wong. Approximate single-diode photovoltaic model for efficient i-v characteristics estimation. *The Scientific World Journal*, 2013, 2013. http://dx.doi.org/10.1155/2013/230471.

[28] W. De Soto, S. A. Klein, and W. A. Beckman. Improvement and validation of a model for photovoltaic array performance. *Solar Energy*, 80:78–88, 2006.

[29] B.R. Zaharatos, M. Campanelli, C. Hansen, K. Emery, and L. Tenorio. Likelihood methods for single-diode model parameter estimation from noisy I-V curve data. *40th IEEE Photovoltaics Specialists Conference*, 2014.

[30] Carl R. Osterwald. Translation of device performance measurements to reference conditions. *Solar Cells*, 18:269–279, 1986.

[31] Amit Jain and Avinashi Kapoor. Exact analytical solutions of the parameters of real solar cells using Lambert W-function. *Solar Energy Materials & Solar Cells*, 81:269–277, 2004.

[32] S. R. Lele, K. Nadeem, and B. Schmuland. Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association*, 105(492):1617–1625, 2010.

[33] Oana Teodora Chis, Julio R. Banga, and Eva Balsa-Canto. Structural identifiability of systems biology models: a critical comparison of methods. *PloS one*, 6, 2011.

[34] Thomas J. Rothenberg. Identification of model parameters. *Econometrica*, 39 (3):577–591, 1971.

[35] D. Campbell and S. Lele. An anova test for parameter estimability using data cloning with application to statistical inference for dynamic systems. *Computational Statistics & Data Analysis*, 70:257–267, 2014.

[36] H. Haario, M. Laine, A. Mira, and E. Saksman. Dram: Efficient adaptive mcmc. *Statistics and Computing*, 16:339–354, 2006.

[37] Joshua S. Stein, Daniel Riley, and Clifford W. Hansen. *PV_LIB Toolbox (Version 1.1)*. Sandia National Laboratories, Albuquerque, NM, USA, 2014. `pvpmc.org/pv-lib/`.

[38] MATLAB Statistics Toolbox. *Version 8.3 in MATLAB R2013b (8.2.0.701)*. The MathWorks, Inc., Natick, MA, USA, 2013.

[39] J. W. Bishop. Computer simulation of the effects of electrical mismatches in photovoltaic cell interconnection circuits. *Solar Cells*, 25:73–89, 1988.

[40] G. Casella and R. L. Berger. *Statistical Inference*. Duxbury, New York, NY USA, 2002.

[41] A. Agresti and B. Coull. Approximate is better than 'exact' for interval estimation of binomial proportions. *The American Statistician*, 52:119–126, 1998.

[42] R Dennis Cook. *Regression Graphics: Ideas for Studying Regressions Through Graphics*, volume 482. John Wiley & Sons, 2009.

[43] Paul G. Constantine, Qiqi Wang, Alireza Doostan, and Gianluca Iaccarino. *A surrogate accelerated Bayesian inverse analysis of the HyShot II flight data*. American Institute of Aeronautics and Astronautics, 2014/06/29 2011.

[44] Qiqi Wang, Han Chen, Rui Hu, and Paul G. Constantine. *Conditional sampling and experiment design for quantifying manufacturing error of transonic airfoil*. American Institute of Aeronautics and Astronautics, 2014/06/29 2011.

[45] Trent W. Lukaczyk, Paul G. Constantine, Francisco Palacios, and Juan J. Alonso. *Active subspaces for shape optimization*. American Institute of Aeronautics and Astronautics, 2014/06/29 2014.

[46] Hany S. Abdel-Khalik, Youngsuk Bang, and Congjian Wang. Overview of hybrid subspace methods for uncertainty quantification, sensitivity analysis. *Annals of Nuclear Energy*, 52(0):28 – 46, 2013. Nuclear Reactor Safety Simulation and Uncertainty Analysis.

[47] Alexander M Samarov. Exploring regression structure using nonparametric functional estimation. *Journal of the American Statistical Association*, 88(423):836–847, 1993.

[48] Bradley Efron and Robert J Tibshirani. *An Introduction to the Bootstrap*. CRC press, 1994.

[49] Ian Jolliffe. *Principal Component Analysis*. Springer, 2002.

[50] Gene H Golub and Charles F Van Loan. *Matrix Computations*. Johns Hopkins University Press, 3 edition, 1996.

[51] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, 2nd edition, 2006.

[52] Bruno Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964 – 979, 2008. Bayesian Networks in Dependability.

[53] R. D. Cook and Chih-Ling Tsai. Residuals in nonlinear regression. *Biometrika*, 72:23–29, 1985.