Chapter

STATISTICAL MODELING, PARAMETER ESTIMATION AND MEASUREMENT PLANNING FOR PV DEGRADATION

***Dazhi Yang*[1,*], *Licheng Liu*[2], *Carlos David Rodríguez-Gallegos*[3], *Zhen Ye*[4], *Li Hong Idris Lim*[5] *and Omid Geramifard*[1]**
[1]Singapore Institute of Manufacturing Technology (SIMTech),
Agency for Science, Technology and Research (A*STAR),
Singapore, Singapore
[2]Saferay Pte. Ltd., Singapore, Singapore
[3]Solar Energy Research Institute of Singapore,
National University of Singapore, Singapore, Singapore
[4]Modules Division, REC Solar Pte Ltd., Singapore, Singapore
[5]Department of Electronic Systems,
University of Glasgow (Singapore),
Singapore, Singapore

ABSTRACT

Photovoltaics (PV) degradation is a key consideration during PV performance evaluation. Accurately predicting power delivery over the course of lifetime of PV is vital to manufacturers and system owners. With many systems exceeding 20 years of operation worldwide, degradation rates have been reported abundantly in the recent years. PV degradation is a complex function of a variety of factors, including but not limited to climate, manufacturer, technology and installation skill. As a result, it is difficult to determine degradation rate by analytical modeling; it has to be measured.

As one set of degradation measurements based on a single sample cannot represent the population nor be used to estimate the true degradation of a particular PV technology, repeated measures through multiple samples are essential. In this chapter, linear mixed effects model (LMM) is introduced to analyze longitudinal degradation data. The framework herein introduced aims to address three issues: 1) how to model the difference in degradation observed in PV modules/systems of a same technology that are installed at a shared location; 2) how to estimate the degradation rate and

*E-mail address: yangdazhi.nus@gmail.com

quantiles based on the data; and 3) how to effectively and efficiently plan degradation measurements.

1. Introduction

The installed photovoltaic (PV) capacity has exceeded 227 GW around the globe in April 2016 [1], with the majority of installations realized only in the past few years. Aside from various environmental benefits of adopting PV power, the increasing PV installation is also motivated by financial benefits that have become apparent in the recent years, when grid parity has occurred in many places in the world. PV degradation is a key consideration during PV performance evaluation, not only concerning the manufacturers, but also system developers and owners. This chapter studies the PV degradation through a statistical model. More importantly, the degradation measurement planning is discussed.

1.1. Motivations

Typically, the installation of a PV power plant consists of three phases, namely, the development, construction, as well as operations and maintenance (O&M) phases. Having an accurate estimation of PV degradation is highly critical for the owners and investors of PV power plants, especially during the development and the O&M phases. Despite the substantial drop in the cost of PV power world-wide, the initial capital cost required to set up a sizable, utility-scale PV power plant can still be very taxing on the liquidity of PV power plant owners. Consequently, the owners typically turn to development banks or other financial institutions with deep pockets for financing. The estimation of PV degradation rate thus plays an important role in securing significant equity and/or debt financing, which is a key step in the development phase of a project. The development banks or financial institutions would generally involve third-party technical advisors to estimate the levelized cost of electricity (LCOE), the ratio of the total cost of the PV power plant to the total energy yield produced by the PV power plant over its entire lifecycle, and thus calculate the return on investment for a PV power plant. As the estimation of the degradation rate directly affects the LCOE of a PV power plant, it relates to the amount of financing obtainable.

In the event of an overestimation of the degradation rate (predicted degradation is greater than the actual degradation), resulting in a lower predicted total energy yield produced over the lifetime of a PV power plant, the financing can be undesirably decreased, or in the worst case scenario, unapproved. In the case of an underestimation of the degradation rate (predicted degradation is smaller than the actual degradation), the predicted total energy yield may become higher than the actual yield. This seemingly favors the securing of financing in the development phase, but ultimately it will be detrimental for the O&M phase, which typically stretches over a long period of 20 to 25 years. Although it is possible to justify a couple of years of underperformance of the PV power plant with the unforeseen "bad" weather, an underestimation of the degradation rate would potentially imply consecutive years of underperformance of the PV power plant. In such situation, depending on the financing terms and conditions from the development banks or financial institutions, a full or partial lock-down of the revenues from the sale of electricity of the PV power plant through power purchased agreements, or feed-in tariffs, will be triggered until certain

release conditions are fulfilled. Thus, in order to secure a significant amount of financing, while protecting the owners of the PV power plants from the possibility of a revenue lockdown, an accurate estimation of degradation rate is required.

Aside from being the interest of PV plant owners and investors, PV degradation is also important for the manufacturers to set their warranty. In the increasingly competitive market, it has become popular for the manufacturers to provide peak power warranty for their PV modules. The typical module manufacturer power output warranty increased from 5 years to 25 five years since 1985 [2]. Such warranty usually comes in one of two forms: 1) warranting the maximum annual power decline of PV modules to be less than a certain percentage (such as 0.7%), and 2) warranting the peak power to be above a certain percentage of the nameplate power (such as 80%) at the end of a period (such as 25 years). Both warranty schemes require accurate information on degradation, or more specifically, the information on PV power degradation quantile. In statistics and probability theory, quantiles are points that divide a probability distribution into contiguous intervals with equal probabilities. The definition itself may be less known, but the examples of quantiles, namely, quartiles and percentiles, are well-known. As degradation rate is only a point estimate, i.e., a number; it does not provide the information on PV degradation distribution. Manufacturers would thus need to consider the degradation quantile in their warranty setting. For example, if the degradation rate is estimated to be 0.7% per year (this can be thought of as the median, if a normal distribution is assumed), it is impractical for the manufacturers to set their warranty exactly at 0.7%, since 50% of the modules are likely to experience a degradation larger than that. Such improper warranty would lead to mass replacement or repair of modules, and thus harm the manufacturers financially.

1.2. PV Degradation Preliminaries

As degradation rate is receiving more attention, many researchers have reported degradation rates based on available data. Comprehensive reviews of published degradation rates can be found in Refs. [2, 3]. Ref. [4] reviewed some of the mechanisms which cause PV degradation. Degradation in PV can be quantified at the module level [5, 6] and system level [7, 8]. Based on the study by Jordan and Kurtz [2], the mean degradation rates of modules and systems differ by only small margins, despite their distinct degradation mechanisms. For such reasons, this chapter only provides a simulation example on module degradation. Nevertheless, the methods to measure system degradation are provided for referencing.

PV degradation is studied across different technologies [9–12]. Five mainstream technologies are often seen in the literature, namely, amorphous silicon, cadmium telluride, copper indium gallium selenide, mono-crystalline silicon and multi-crystalline silicon. Among these technologies, crystalline silicon received the most attention at the reported time [2]. Therefore, in the later analysis, without loss of generality, crystalline silicon modules are used. For crystalline silicon technologies, the degradation rates observed in the first year of operation are much higher due to early degradation mechanisms, such as the light induced degradation (LID) [13]. Therefore, it is more amenable to remove this "burn-in" year from the simulation.

As mentioned earlier, the nameplate power measured under standard test conditions (STC) is a commonly used parameter to describe the expected module energy output. How-

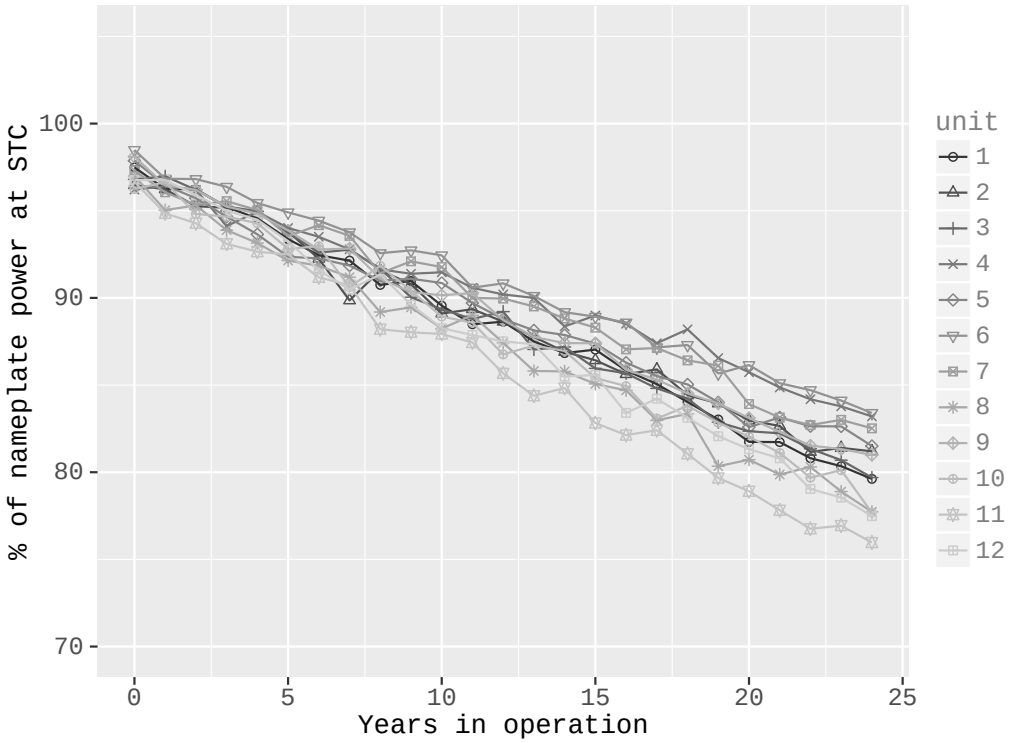

Figure 1. Simulated degradation paths for 12 collocated crystalline silicon modules from the same manufacturing batch. Simulation is performed based on Eq. (6), see below.

ever, even for modules with the same nameplate power, the actual energy production may be significantly different [13]. Some factors that lead to these differences are climate, manufacturing process and installation skill, etc. During a degradation study, these factors can be eliminated by considering PV modules of a same manufacturing batch and installed at a shared location. Nevertheless, other factors that affect the peak power, such as different degrees of LID, as well as the intricate loss propagation in each module, are difficult to control, if not to eliminate at all. The degradation rate estimation thus cannot depend on few measurements of a single sample; repeated measures are essential.

Before we describe the degradation model based on repeated measures, a set of typical degradation curves is shown. Figure 1 shows the degradation curves for 12 crystalline silicon modules; simulation (detailed in Section 3.) is used to generate these curves. It is assumed that these modules have identical nameplate information and are installed at a shared climatic condition. Furthermore, the early degradation effects are removed from the simulation, so the the degradation of the remaining years can be assumed to be linear. Although some publications use an exponential degradation model [14]; it is shown that for a typical starting degradation rate, the models do not differ significantly up to 25 years [15].

In the following sections, we are going to look at: 1) how the degradation curves shown in Figure 1 can be modeled using the linear mixed effect model (LMM); 2) how to estimate the parameter values and quantile information through the LMM; and 3) how to effectively

and efficiently plan the degradation measurement. Degradation measurement planning in this case refers to the design of degradation experiment, which includes design parameters such as the sample size, period of study, measurement frequency and types of measurement.

2. Modeling, Parameter Estimation and Measurement Planning

Repeated measures are defined if an outcome is measured repeatedly within a set of *units* [16]. In our current context, a unit could be a PV module or system, depending on the type of degradation measurements. These collected data are called *longitudinal data* if they are taken over time [17].

2.1. Degradation Model

Suppose we collect m_i degradation measurements of unit i, where $i = 1, \cdots, n$ denoting the samples, and let y_{ij} be the measured degradation of unit i at time t_{ij}, where $j = 1, \cdots, m_i$, the linear degradation model is given by:

$$y_{ij} = b_{0,i} + b_{1,i} t_{ij} + \varepsilon_{ij}, \tag{1}$$

where $b_{0,i}$ and $b_{1,i}$ denote the intercept and the gradient of the linear model for unit i; ε_{ij} denotes a random effect. The term $\mathcal{D}_i = b_{0,i} + b_{1,i} t_{ij}$ is called the *true degradation* path for unit i. Symbol m_i is used to denote the (possibly) different numbers of measurements for each unit, as longitudinal data are often unbalanced [18, 19].

We have discussed earlier that PV degradation is a complex function of a variety of factors. In this chapter, we do not consider contributing factors that may have significant impacts on degradation, such as environmental and climate condition, technology and manufacturer. Under such arrangement, the intercept and gradient in Eq. (1) can be modeled using a bivariate normal distribution, $(b_0, b_1)^\top \sim \mathrm{BVN}(\boldsymbol{\beta}, \boldsymbol{V})$ with mean vector

$$\boldsymbol{\beta} = (\beta_0, \beta_1)^\top \tag{2}$$

and covariance matrix

$$\boldsymbol{V} = \begin{pmatrix} \sigma_{b_0}^2 & \rho\sigma_{b_0}\sigma_{b_1} \\ \rho\sigma_{b_0}\sigma_{b_1} & \sigma_{b_1}^2 \end{pmatrix}. \tag{3}$$

It should be noted that b_0 and b_1 are random variables, and $b_{0,i}$ and $b_{1,i}$ are a particular realization of these random variables, i.e., the intercept and gradient for unit i. The practical relevance and implications of the above BVN assumption are more described in Section 3.. The probability density function of BVN distribution is:

$$f(b_0, b_1; \boldsymbol{\beta}, \boldsymbol{V}) = \frac{1}{2\pi\sigma_{b_0}\sigma_{b_1}\sqrt{1-\rho^2}} \exp\left[-\frac{\kappa}{2(1-\rho^2)}\right], \tag{4}$$

where

$$\kappa = \left(\frac{b_1 - \beta_1}{\sigma_{b_0}}\right)^2 + \left(\frac{b_2 - \beta_2}{\sigma_{b_1}}\right)^2 - 2\rho\left(\frac{b_1 - \beta_1}{\sigma_{b_0}}\right)\left(\frac{b_2 - \beta_2}{\sigma_{b_1}}\right). \tag{5}$$

2.2. Linear Mixed Effects Model

Linear mixed effects model is perhaps the most fundamental model for analyzing longitudinal data. Its general form is given as:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i^* + \boldsymbol{\varepsilon}_i, \tag{6}$$

where $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are matrices of known covariates [18]. In this model, $\boldsymbol{\beta}$ is the *fixed effects* and $\boldsymbol{b}_i^*$ is the *unit-specific effects*. The degradation model, Eq. (1), can be related to the above matrix equation with notations used in Section 2.1., i.e., $\boldsymbol{Y}_i = (y_{i1}, \cdots, y_{im_i})^\top$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \cdots, \varepsilon_{im_i})^\top$ and $\boldsymbol{b}_i^* = (b_{0,i}^*, b_{1,i}^*)^\top$; it can be expressed as:

$$y_{ij} = (\beta_0 + b_{0,i}^*) + (\beta_1 + b_{1,i}^*)t_{ij} + \varepsilon_{ij}, \tag{7}$$

where $(b_0^*, b_1^*)^\top \sim \mathrm{BVN}(\boldsymbol{0}, \boldsymbol{V})$. Furthermore, we have:

$$\boldsymbol{\varepsilon}_i \sim \mathrm{MVN}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_i); \tag{8}$$

$$\mathbb{V}(\boldsymbol{b}_i^*, \boldsymbol{\varepsilon}_i) = \boldsymbol{0}; \tag{9}$$

$$\boldsymbol{X}_i = \boldsymbol{Z}_i = \begin{pmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{im_i} \end{pmatrix} \tag{10}$$

and $\boldsymbol{I}_i$ is an m_i by m_i identity matrix. We do not distinguish between LMM and degradation model hereafter.

Eq. (8) implies that $\boldsymbol{\varepsilon}_i$ is independent and normally distributed. Eq. (9) reveals that $\boldsymbol{b}_i^*$ and $\boldsymbol{\varepsilon}_i$ are independent. Under these settings, $\boldsymbol{Y}_i$ has a multivariate normal distribution with mean vector $\boldsymbol{X}_i\boldsymbol{\beta}$ and covariance

$$\boldsymbol{\Sigma}_i = \mathbb{V}(\boldsymbol{Y}_i) = \boldsymbol{Z}_i\boldsymbol{V}\boldsymbol{Z}_i^\top + \sigma^2\boldsymbol{I}_i, \tag{11}$$

which is a special case of the repeated-measured models in Ref. [19], i.e, $\boldsymbol{Y}_i \sim \mathrm{MVN}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$. The multivariate normal random vector $\boldsymbol{Y}_i$ has pdf:

$$f(\boldsymbol{y}_i; \boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i) = \frac{1}{(\sqrt{2\pi})^{m_i}|\boldsymbol{\Sigma}_i|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right], \tag{12}$$

where $|\boldsymbol{\Sigma}_i|$ is the determinant of $\boldsymbol{\Sigma}_i$.

2.3. Parameter Estimation

We have presented the degradation model in the previous sections. In order to use the model to describe a particular set of data, model parameters need to be estimated. Standard ways for parameter estimation include method of moments, maximum likelihood (ML) method and Bayesian method; we use ML method in this chapter. *Maximum likelihood estimator* (MLE) of a parameter θ is the value of θ that maximizes $\mathcal{L}(\theta)$, the *likelihood function* of θ. As maximizing a *log-likelihood function*, $\ell(\theta)$, leads to the same answer as maximizing $\mathcal{L}(\theta)$, we consider $\ell(\theta)$ here. In this case, it is easier to work with $\ell(\theta)$.

By observing Eq. (12), we know that the MVN is controlled by its mean vector $\boldsymbol{X}_i\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{\Sigma}_i$. These two parameters are controlled by:

$$\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma_{b_0}, \sigma_{b_1}, \rho, \sigma)^\top. \tag{13}$$

Suppose $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n$ are independent observations from $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_n$, respectively, follow Ref. [19], the total log-likelihood for n units is:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}) = \text{const.} - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) - \frac{1}{2}\sum_{i=1}^{n}\log|\boldsymbol{\Sigma}_i|, \tag{14}$$

where $\ell_i(\boldsymbol{\theta})$ is the log-likelihood for observational unit i. We use $\widehat{\boldsymbol{\theta}}$ to denote the ML estimate of $\boldsymbol{\theta}$.

2.4. Degradation Quantiles

The parameters of an LMM are estimated through the ML method. The information attained so far is sufficient for many PV degradation applications. For example, $\widehat{\beta}_1$ (the ML estimate of β_1) can be considered as the estimated degradation rate. Other quantities, such as the degradation quantile, can also be derived from $\widehat{\boldsymbol{\theta}}$.

Recall that the true degradation at time t is $\mathcal{D} = b_0 + b_1 t$, $\mathcal{D}$ can be considered as a function of random variables. For the present case, b_0 and b_1 follow a BVN distribution, their sum can be shown to be normally distributed with mean and variance being:

$$\mathbb{E}(\mathcal{D}) = \mathbb{E}(b_0 + b_1 t) = \beta_0 + \beta_1 t \tag{15}$$

and

$$\mathbb{V}(\mathcal{D}) = \mathbb{V}(b_0 + b_1 t) = \sigma_{b_0}^2 + t^2\sigma_{b_1}^2 + 2t\rho\sigma_{b_0}\sigma_{b_1}, \tag{16}$$

respectively. The p quantile of the degradation distribution at time t is:

$$d_p(t) = \mathbb{E}(\mathcal{D}) + \sqrt{\mathbb{V}(\mathcal{D})}\Phi^{-1}(p), \tag{17}$$

where $\Phi^{-1}(p)$ is the *probit* function. Mathematically, it is the inverse of $\Phi(z)$, the standard normal CDF. For example, $\Phi(-1.96) = 0.025$ and $\Phi^{-1}(0.025) = -1.96$. Let $\widehat{d}_p(t)$ be the ML estimate of $d_p(t)$, it can be directly obtained by evaluating Eq. (17) at $\widehat{\boldsymbol{\theta}}$:

$$\widehat{d}_p(t) = \widehat{\beta}_0 + \widehat{\beta}_1 t + \Phi^{-1}(p)\sqrt{\widehat{\sigma}_{b_0}^2 + t^2\widehat{\sigma}_{b_1}^2 + 2t\widehat{\rho}\widehat{\sigma}_{b_0}\widehat{\sigma}_{b_1}}. \tag{18}$$

2.5. Preliminaries for Degradation Measurement Planning

In statistics, estimators (such as MLE) are random variables; their distributions are called *sampling distribution*. On the other hand, estimates are numbers computed from data. For simplicity, we do not distinguish them in terms of notations in this chapter. However, the concept should not be mixed up.

The estimates $\widehat{\boldsymbol{\theta}}$ and $\widehat{d}_p(t)$ are known as *point estimates*; point estimation provides a single "best guess" of some quantity of interest [20]. However, point estimates are often

insufficient; the knowledge of the sampling distributions is also important, especially during measurement planning where confidence intervals are of interest. The standard deviation of an estimator is called the *standard error*, denoted by $\mathbf{se}$. In particular, we are interested in the standard error of the MLE of the p quantile, namely, $\mathbf{se}(\widehat{d}_p)$. To derive $\mathbf{se}(\widehat{d}_p)$, Fisher information is useful.

2.5.1. Fisher Information

Let random variable X have a distribution $f(x;\theta)$, the *score function* (a function of θ) is defined to be:

$$s(X;\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = \frac{\partial \log f(x;\theta)}{\partial \theta}, \tag{19}$$

i.e., the derivative of the log-likelihood. The *Fisher information* $\mathcal{I}(\theta)$ is defined as:

$$\mathcal{I}(\theta) = \mathbb{E}_\theta \left[s^2(X;\theta) \right] = \int \left(\frac{\partial \ell(\theta)}{\partial \theta} \right)^2 f(x;\theta) dx. \tag{20}$$

As Eq. (20) can be difficult to evaluate sometimes, an alternative definition can be used:

$$\mathcal{I}(\theta) = -\mathbb{E}_\theta \left(\frac{\partial^2 \ell(\theta)}{\partial \theta^2} \right). \tag{21}$$

A proof of Eq. (21) can be found in Ref. [21], page 242.

In our present case, the parameter $\boldsymbol{\theta}$ can be written into $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma_{b_0}, \sigma_{b_1}, \rho, \sigma)^\top = (\boldsymbol{\beta}^\top, \boldsymbol{\vartheta}^\top)^\top$, the Fisher information of unit i can be evaluated by considering the Hessian matrix [19]:

$$\boldsymbol{H}_i = \begin{pmatrix} \boldsymbol{H}_{\boldsymbol{\beta}\boldsymbol{\beta},i} & \boldsymbol{H}_{\boldsymbol{\beta}\boldsymbol{\vartheta},i} \\ \boldsymbol{H}_{\boldsymbol{\vartheta}\boldsymbol{\beta},i} & \boldsymbol{H}_{\boldsymbol{\vartheta}\boldsymbol{\vartheta},i} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}} & \dfrac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\vartheta}} \\ \dfrac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\beta}} & \dfrac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}} \end{pmatrix}. \tag{22}$$

We thus have

$$\mathcal{I}_i(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{X}_i^\top \boldsymbol{\Sigma}_i^{-1} \boldsymbol{X}_i & \mathbf{0} \\ \mathbf{0} & \boldsymbol{M}_i \end{pmatrix}, \tag{23}$$

where the element on row r and column s of the symmetrical 4 by 4 matrix $\boldsymbol{M}_i$ is:

$$M_{i,rs} = \frac{1}{2} tr(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{ir} \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{is}), \tag{24}$$

$r, s = 1, \cdots, 4$ and the explicit representations of $\dot{\boldsymbol{\Sigma}}_{ir}$ or $\dot{\boldsymbol{\Sigma}}_{is}$ are obtained by differentiating

Eq. (11) with respect to each parameter in $\boldsymbol{\vartheta}$:

$$\dot{\boldsymbol{\Sigma}}_{i1} = \frac{\partial \boldsymbol{\Sigma}_i}{\partial \vartheta_1} = \frac{\partial \boldsymbol{\Sigma}_i}{\partial \sigma_{b_0}} = \boldsymbol{Z}_i \begin{pmatrix} 2\sigma_{b_0} & \rho\sigma_{b_1} \\ \rho\sigma_{b_1} & 0 \end{pmatrix} \boldsymbol{Z}_i^\top; \tag{25}$$

$$\dot{\boldsymbol{\Sigma}}_{i2} = \frac{\partial \boldsymbol{\Sigma}_i}{\partial \vartheta_2} = \frac{\partial \boldsymbol{\Sigma}_i}{\partial \sigma_{b_1}} = \boldsymbol{Z}_i \begin{pmatrix} 0 & \rho\sigma_{b_0} \\ \rho\sigma_{b_0} & 2\sigma_{b_1} \end{pmatrix} \boldsymbol{Z}_i^\top; \tag{26}$$

$$\dot{\boldsymbol{\Sigma}}_{i3} = \frac{\partial \boldsymbol{\Sigma}_i}{\partial \vartheta_3} = \frac{\partial \boldsymbol{\Sigma}_i}{\partial \rho} = \boldsymbol{Z}_i \begin{pmatrix} 0 & \sigma_{b_0}\sigma_{b_1} \\ \sigma_{b_0}\sigma_{b_1} & 0 \end{pmatrix} \boldsymbol{Z}_i^\top; \tag{27}$$

$$\dot{\boldsymbol{\Sigma}}_{i4} = \frac{\partial \boldsymbol{\Sigma}_i}{\partial \vartheta_4} = \frac{\partial \boldsymbol{\Sigma}_i}{\partial \sigma} = 2\sigma \boldsymbol{I}_i. \tag{28}$$

The Fisher information for all n units is the sum of the Fisher information for each unit:

$$\mathcal{I}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathcal{I}_i(\boldsymbol{\theta}). \tag{29}$$

2.5.2. Standard Error of MLE

The Fisher information matrix is used to calculate the standard error and covariance matrices associated with ML estimates. Wasserman [20] states the following theorem:

Theorem 1. *(Asymptotic Normality of the MLE) Let* $\mathbf{se} = \sqrt{\mathbb{V}(\widehat{\theta})}$. *Under appropriate regularity conditions, the following hold:*
1. $\mathbf{se} \approx \sqrt{1/\mathcal{I}(\theta)}$ *and*

$$\frac{(\widehat{\theta} - \theta)}{\mathbf{se}} \rightsquigarrow \mathrm{N}(0, 1). \tag{30}$$

2. Let $\widehat{\mathbf{se}} = \sqrt{1/\mathcal{I}(\widehat{\theta})}$. *Then,*

$$\frac{(\widehat{\theta} - \theta)}{\widehat{\mathbf{se}}} \rightsquigarrow \mathrm{N}(0, 1). \tag{31}$$

Symbol $\rightsquigarrow$ *denotes convergence in distribution. (end of theorem)*

The first statement of the theorem says that $\widehat{\theta} \approx \mathrm{N}(\theta, \mathbf{se})$. The second statements says that even when the standard error is replaced by the estimated standard error, the asymptotic normality is still true. This theorem can thus be used to construct the asymptotic confidence interval for $\widehat{\theta}$. If we extend the theorem to multi-parameter cases [20], we have:

Theorem 2. *Under appropriate regularity conditions,*

$$(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \approx \mathrm{MVN}(\mathbf{0}, \boldsymbol{J}(\boldsymbol{\theta})), \tag{32}$$

where $\boldsymbol{J}(\boldsymbol{\theta}) = [\mathcal{I}(\boldsymbol{\theta})]^{-1}$ *is the inverse of the Fisher information matrix. (end of theorem)*

The details of those regularity conditions mentioned in the above theorems can be found in Chapter 12 of Ref. [21]. Following Theorem 2, we have:

$$\mathbb{V}(\widehat{\boldsymbol{\theta}}) = [\mathcal{I}(\boldsymbol{\theta})]^{-1}, \tag{33}$$

where $\mathbb{V}(\cdot)$ denotes the *approximated variance-covariance matrix* of the MLE. In other words, the $\widehat{\mathrm{se}}^2$ of each parameter is given by the corresponding diagonal term of $\boldsymbol{J}(\boldsymbol{\theta})$; the covariance between the parameters are given the off-diagonal terms of $\boldsymbol{J}(\boldsymbol{\theta})$. An estimate of $\mathbb{V}(\cdot)$ at the ML estimates is denoted by $\widehat{\mathbb{V}}(\widehat{\boldsymbol{\theta}})$.

2.5.3. Standard Error and Confidence Interval of the Degradation Quantiles

The estimate of the approximated variance-covariance matrix of the MLE can be obtained through the inverse of the Fisher information matrix. With this information, together with the degradation quantile d_p evaluated at the ML estimates (denoted by $\widehat{d}_p(\widehat{\boldsymbol{\theta}})$ or simply $\widehat{d}_p$), the standard error of the quantile can be estimated through the delta method. In statistics, the *delta method* is used to estimate the approximate probability distribution for a smooth function, g, of an asymptotically normal statistical estimator (such as MLE). In the present case, $g(\widehat{\boldsymbol{\theta}})$ is Eq. (17), namely, the quantile function of the degradation.

Wasserman [20] states the following theorem:

Theorem 3. *(Multiparameter delta method) Suppose that ∇g evaluated at $\widehat{\boldsymbol{\theta}}$ is not 0. Let $\widehat{\tau} = g(\widehat{\boldsymbol{\theta}})$. Then*

$$\frac{(\widehat{\tau} - \tau)}{\widehat{\mathbf{se}}(\widehat{\tau})} \rightsquigarrow \mathrm{N}(0, 1), \tag{34}$$

where

$$\widehat{\mathbf{se}}(\widehat{\tau}) = \sqrt{(\widehat{\nabla} g)^{\top} \widehat{\mathbb{V}}(\widehat{\boldsymbol{\theta}})(\widehat{\nabla} g)}, \tag{35}$$

$\widehat{\nabla} g$ is ∇g evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. (end of theorem)

In our case, ∇g is the (column) vector of partial derivatives of d_p with respect to the parameters. The elements of this vector are:

$$\partial d_p / \partial \beta_0 = 1; \tag{36}$$
$$\partial d_p / \partial \beta_1 = t; \tag{37}$$
$$\partial d_p / \partial \sigma_{b_0} = \zeta(2\sigma_{b_0} + 2t\rho\sigma_{b_1}); \tag{38}$$
$$\partial d_p / \partial \sigma_{b_1} = \zeta(2t^2\sigma_{b_1} + 2t\rho\sigma_{b_0}); \tag{39}$$
$$\partial d_p / \partial \rho = \zeta(2t\sigma_{b_0}\sigma_{b_1}); \tag{40}$$
$$\partial d_p / \partial \sigma = 0, \tag{41}$$

where

$$\zeta = \frac{\Phi^{-1}(p)}{2\sqrt{\sigma_{b_0}^2 + t^2\sigma_{b_1} + 2t\rho\sigma_{b_0}\sigma_{b_1}}}. \tag{42}$$

The estimated standard error of the quantile of degradation distribution at the ML estimates is thus given by:

$$\widehat{\mathbf{se}}(\widehat{d}_p) = \sqrt{\widehat{\boldsymbol{c}}^\top \widehat{\mathbb{V}}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{c}}}, \tag{43}$$

where $\boldsymbol{c}$ is the (column) vector of partial derivatives of d_p. The $1-\alpha\%$ confidence interval of the estimated quantile is thus given by:

$$\widehat{d}_p \pm z_{\alpha/2}\widehat{\mathbf{se}}(\widehat{d}_p), \tag{44}$$

under the normal-based interval.

2.6. Section Summary

This section discusses the modeling and parameter estimation for PV degradation. A linear degradation model is first presented. Its equivalent LMM is then outlined. A total of six parameters, namely, $\beta_0, \beta_1, \sigma_{b_0}, \sigma_{b_1}, \rho$ and σ, are estimated using the ML method. With the estimated parameters, the degradation quantile $d_p(t)$ at time t could be evaluated through Eq. (18). However, as a single value (a point estimate) carries limited information about the distribution of $\widehat{d}_p$, the standard error and confidence interval of $\widehat{d}_p$ are given in Eqs. (43) and (44), respectively. To derive the representations of standard error and confidence interval, the Fisher information and several theorems are used.

3. Simulation Study on PV Degradation Planning

In this section, we use simulation to describe the planning method for PV degradation. In particular, we are going to look at how degradation measurement settings would affect the uncertainties in degradation estimates. In our context, *degradation measurement settings* refer to factors such as type of degradation measurements, number of test units and number of measurements for each unit. We note that this simulation study does not advise on the "optimal" settings for degradation measurements, because the appropriate settings depend on a variety of factors. Furthermore, the tolerance for standard error and the acceptable range for confidence intervals may also vary based on different expert views. We thus present a visual representation of various concepts outlined in the previous section.

In the simulation study below, we consider PV modules as the degradation test units. More specifically, a total of 12 crystalline modules are assumed. By assuming the units are collocated modules from a same manufacturing batch, some factors (e.g., climate, technology and manufacturer) that can affect the degradation are eliminated.

3.1. Low- and High- Accuracy Degradation Experiments

Three methods of measurement are commonly used for PV degradation studies [22]: 1) the regression-based low-accuracy experiment (LE) through outdoor monitoring data, 2) the LE through outdoor I–V measurements, and 3) the high-accuracy experiment (HE) though the indoor I–V measurements. Similar to many other engineering problems (see Refs. [23–25] for other examples on low- and high-accuracy experiments), the LE data in

PV degradation are more accessible as compared to the HE data. To fully utilize the results from LE, the outcomes from the HEs are often used to benchmark various LEs to determine their accuracies [22]. Although LEs in general have a lower cost as compared to HE, the limitation of the LEs is obvious during the decision making process of the manufacturers, for example, setting the degradation warranty based on inaccurate degradation rates leads to financial risks [26].

Among the three degradation experiments listed above, the first experiment is commonly used to determine the degradation of PV systems or sub-systems, while the latter two experiments are used when the test units are PV modules. Although the units used in this simulation study are set to be modules, we describe all three experiments below to give the readers a general understanding on each experiment.

3.1.1. Low-Accuracy Degradation Methods

Low-accuracy degradation experiments use regressions and outdoor monitoring data to determine degradation rates. The regressor (explanatory variable) is usually time, while the regressands (explained variable) are different metrics. Consider a linear regression problem $y = b_0 + b_1 t + \varepsilon$, the term *metric* is used by Jordan and Kurtz [22] to denote y. After the regression, the degradation rate is reflected by the fitted value $\widehat{b}_1$, i.e., the gradient of the fitted line.

When the test units are PV systems, examples of metrics include performance ratio (PR), PR with temperature correction [27], $\mathrm{DC}/G_{\mathrm{POA}}$ [28] and PVUSA [29]. These four metrics were summarized by Jordan and Kurtz [22]. While these metrics depend on irradiance data (see below) which is not always available, other irradiance-independent metrics have been proposed [30]. On the other hand, when the units are modules, examples of metrics include maximum power ($P_{\max}$), open-circuit voltage (V_{oc}), short-circuit current (I_{sc}) and fill factor (FF). These metrics were described and used by Smith *et al.* [31].

The core idea of LEs is to use the drops in certain metrics to represent degradation in PV modules/systems over time. It is therefore important to consider various types of corrections and data filtering. PR is perhaps the most commonly used metric to measure system performance. In mid-latitude regions, PR varies within a year with winter showing a relatively higher PR than summer. This observation is mainly due to the lower module temperature in winter. Module temperature is thus commonly used to adjust the seasonal variation in PR. For example, Dierauf *et al.* [32] proposed a PR correlation method where PR is normalized by removing the weather dependency. Conventional PR is given by:

$$\mathrm{PR} = \frac{\sum_t EN_{\mathrm{AC},t}}{\sum_t \left[P_{\mathrm{STC}} \left(\frac{G_{\mathrm{POA},t}}{G_{\mathrm{STC}}} \right) \right]}, \tag{45}$$

while the weather-corrected performance ratio, PR_{corr}, is:

$$\mathrm{PR}_{\mathrm{corr}} = \frac{\sum_t EN_{\mathrm{AC},t}}{\sum_t \left\{ P_{\mathrm{STC}} \left(\frac{G_{\mathrm{POA},t}}{G_{\mathrm{STC}}} \right) \left[1 - \gamma (T_{\mathrm{mod\_typ\_avg}} - T_{\mathrm{mod},t}) \right] \right\}}, \tag{46}$$

with γ being the temperature coefficient for power, with a typically value of $-0.4\%/^{\circ}\mathrm{C}$; EN_{AC} being the measured AC power generation (in kW); P_{STC} being the nameplate power

(in kW); G_{POA} being the in-plane irradiance (in kW/m^2); G_{STC} being the STC irradiance (1 kW/m^2); T_{mod} being the module temperature (in $^\circ$C) and $T_{\mathrm{mod\_typ\_avg}}$ being the average cell temperature computed from a typical meteorological year.

The summation in the above equations can be calculated over any defined period of time, may it be days, weeks, months or years. It is shown that the seasonal cycles in the PR can be effectively removed using this weather correction regardless if monthly or daily PR is used [32]. Besides the corrections in PR, data filtering is also commonly used to remove certain data points. For example, an irradiance filter can be applied to remove data points far from STC; a module temperature filter can be used to remove data points which deviate largely from the $T_{\mathrm{mod\_typ\_avg}}$. In addition, outlier filters and stability filters are also frequently involved in the data quality control process [22].

As previously mentioned, the metrics used to determine module degradation are usually different from those for systems. In order to compute metrics such as P_{max}, V_{oc} and I_{sc}, I–V measurements (I–V curves) are required. The commercial I–V curve tracers are commonly equipped with the capability of generating the values of these metrics. More advanced mathematical models [33] that are only suitable for off-chip computation are also available. On this point, we refer the readers to another chapter of this book, namely, *PV Panel Modeling and Identification*, for a detailed tutorial. Comparing to the indoor measurement (see below), a drawback of the outdoor I–V measurement is that the ambient condition is uncontrolled in this experiment. This leads to the need of minimizing the effect of weather on degradation rate determination; data filtering is essential in this LE. Data filtering here aims at identifying the data points that are close to the STC, for example, an irradiance filter that includes only data points with deviations of $\pm$20 W/m^2 from the STC irradiance could be considered. We refer the readers to [31] for a detailed example on data filtering, in which the authors studied the degradation of 12 crystalline silicon modules over a period of 17 years.

The advantage of conducting LEs is that the experiments can be performed remotely once the measurement devices are in place. As compared to the HE, the cost of LE can be lower as well, especially when the monitoring equipment becomes more affordable and reliable. On the other hand, although appropriate measurement equipment setup and data filtering could minimize the uncertainty in the LEs to a certain extent, the outdoor experiments still suffer from soiling and other degradation mechanisms that are hard to trace through monitoring data. In the later sections, we further discuss the trade-off between cost and accuracy for different degradation experiments.

3.1.2. High-Accuracy Degradation Methods

High-accuracy degradation experiments rely on measuring the I–V curves of a PV panel at fixed time intervals under indoor conditions [2]. Due to the dependency on indoor test facilities, HE is usually performed by shipping the modules back to the manufacturers. This embeds high costs into the degradation experiment. For such reasons, it is common (about 74% of all degradation rates reported in the literature) to conduct indoor I–V measurements only once; the measurements are then compared to the nameplate rating of a module [3]. The inherent assumption is the accuracy of the nameplate rating, which may possess significant variance, especially when early degradation mechanisms are not excluded from the

experiments.

3.2. Examples on Parameter Estimation Using MLE

In Section 3.1., low- and high-accuracy degradation experiments are discussed. As the degradation study can be based on different metrics, without loss of generality, we consider percentage of nameplate power at STC as the regressand. We begin our simulation study by setting the degradation parameters. A total of six parameters, namely, β_0, β_1, σ_{b_0}, σ_{b_1}, ρ and σ, are used to parameterize the degradation model. The numeric values of these parameters are justified and set as follows.

PV modules experience early degradation such as LID during the first year of operation. To simulate the approximate 3% drop in the first year, the intercept of true degradation curve β_0 is set at 97. It was reported that some crystalline modules may have more than 4% power loss after the first weeks of operation [13], $\sigma_{b_0} = 0.5$ is used to represent the variations of early degradation among the sample modules. This means that the PV modules under simulation preserve 97% of the nameplate power at STC at time $t = 0$ with a standard deviation of 0.5. It is noted that $t = 0$ denotes the beginning of the simulation, one can consider this to be the beginning of the second actual operating year. Only the simulation time reference t will be used hereafter.

In Ref. [2], a rich literature review is presented on the degradation rate of crystalline silicon technology. It was found that the average degradation rate of crystalline silicon technology is 0.7%/year, i.e., $\beta_1 = -0.7$. Further to that, $\sigma_{b_1} = 0.1$ is interpreted from Ref. [2] to denote the variation in the degradation rate distribution, see Fig. 5 from Ref. [2] for this interpretation.

We assume that the intercept and gradient of the degradation path can be modeled using a BVN distribution with correlation ρ. In reality, this parameter does not carry significant physical implication. However, it is reasonable to assume modules with higher β_0 values degrade slower. Therefore, a small positive correlation between b_0 and b_1, $\rho = 0.3$, is set.

The final model parameter σ represents the error term. It should be set differently for HE and LE. In HE, it can be assumed that the error is small; $\sigma = 0.5$ is set to represent the year-to-year experimental variations in practicing the HE. In LE, experimental errors are higher; $\sigma = 2$ is set to represent the uncertainties due to imperfect data filtering, soiling and other mechanisms that may affect the variance in module power output.

Considering an expected PV lifetime of 25 years, the simulated output power (in % of the nameplate power) for the 12 modules over a period of 24 years (the burn-in year is not simulated) is plotted in Fig. 1 using the HE assumptions and Eq. (1). To obtain these curves, 12 sets of (b_0, b_1) values are first drawn from the BVN distribution parameterized by $\boldsymbol{\beta}$ and $\boldsymbol{V}$. The true degradation paths ($\mathcal{D}_i$) are then produced using $\mathcal{D}_i = b_{0,i} + b_{1,i}t_{ij}$, where $i = 1, \cdots, 12$, $j = 1, \cdots, m_i$ and $m_i = 24, \forall i$. Noise terms are then added to $\mathcal{D}_i$ using random samples drawn from $\mathrm{N}(0, \sigma^2)$, where $\sigma = 0.5$. Using MLE, the estimated HE parameters are: $\widehat{\beta}_0 = 96.982$, $\widehat{\beta}_1 = -0.706$, $\widehat{\sigma}_{b_0} = 0.481$, $\widehat{\sigma}_{b_1} = 0.087$, $\widehat{\rho} = 0.443$ and $\widehat{\sigma} = 0.516$. These estimated parameters agree with our earlier simulation settings, showing the precise estimations from MLE. Similarly, using the LE simulated data, the ML estimates are $\widehat{\beta}_0 = 96.858$, $\widehat{\beta}_1 = -0.709$, $\widehat{\sigma}_{b_0} = 0.405$, $\widehat{\sigma}_{b_1} = 0.086$, $\widehat{\rho} = 0.631$ and $\widehat{\sigma} = 2.062$.

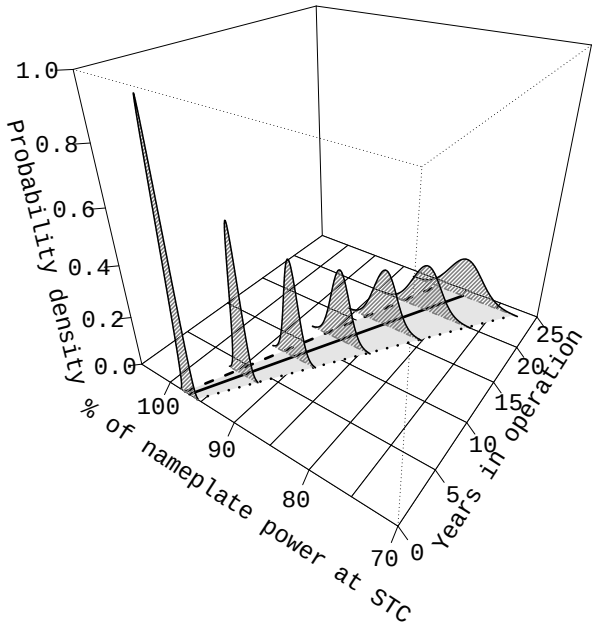

Figure 2. Evaluations of the 0.001 (dotted line on x–y plane), 0.50 (solid line), 0.95 (dashed line) degradation quantiles at ML estimates of the LE.

3.3. Degradation Quantiles Evaluated at ML Estimates

True degradation $\mathcal{D}$ is normally distributed. This is visualized in Figure 2; the data used for plotting are computed with the LE settings. At any time t, the expected value and variance of the normal distribution can be evaluated at the ML estimates through Eqs. (15) and (16). Figure 2 shows the pdfs of $\mathcal{D}$ at $t = 0, 4, \cdots, 24$, the evolution of the PDFs of $\mathcal{D}$ is apparent. In Section 2.4., degradation quantiles are formulated based on the distribution of $\mathcal{D}$. Using the ML estimates we obtained earlier, degradation quantiles can be computed via Eq. (18). Three example quantiles, namely, $\widehat{d}_{0.001}$ (dotted line on x–y plane), $\widehat{d}_{0.5}$ (solid line) and $\widehat{d}_{0.95}$ (dashed line), are shown. This quantile information is critical for manufacturers or system developers to set their warranty.

As mentioned earlier, point estimates have their limitations. Considering standard errors and confidence intervals of the degradation quantiles may help manufacturers or system developers to make their decisions. Standard errors and confidence intervals can be computed via Eqs. (43) and (44), respectively. Figure 3 shows the 95% confidence intervals for $\widehat{d}_{0.5}$ based on 5 and 15 years of LE data.

The computation for the confidence intervals depends on the amount of information available. By examining Eq. (44), it is noted that the confidence interval depends on $\widehat{d}_p$ and $\widehat{\mathbf{se}}(\widehat{d}_p)$. After the model parameter $\boldsymbol{\theta}$ is estimated based on whatever available data, may it be 5, 10 or 15 years of data, $\widehat{d}_p$ can be evaluated for any time t. The solid black lines in

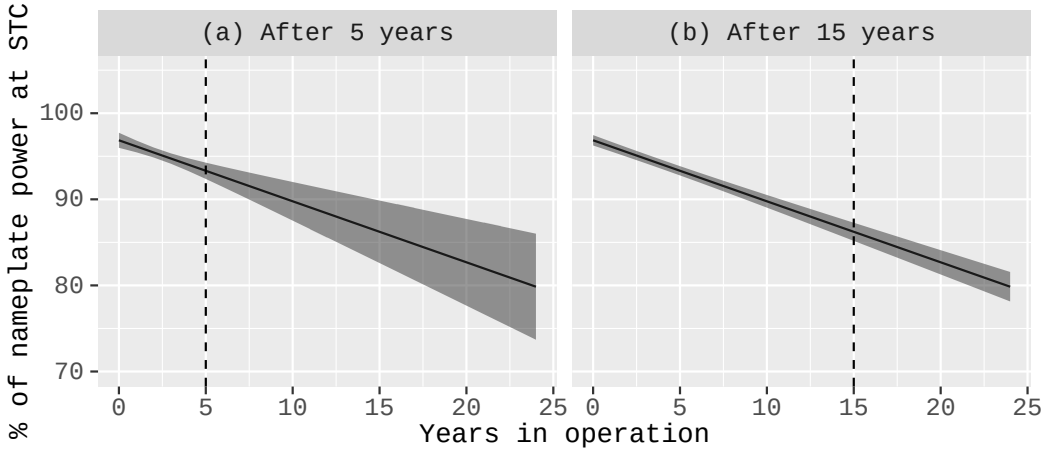

Figure 3. The 95% confidence intervals of 0.50 degradation quantile based on 5 and 15 years of data. The estimated 0.5 quantiles are shown as the solid black line. The shaded regions denote the confidence intervals.

Figure 3(a) and 3(b) are $\widehat{d}_{0.5}$ for $t = 1, \cdots, 24$; and the two lines are identical. In other words, quantiles do not depend on the amount of information but depend only on $\widehat{\boldsymbol{\theta}}$, see Eq. (18). On the other hand, $\widehat{\text{se}}(\widehat{d}_p) = \sqrt{\widehat{\boldsymbol{c}}^\top \widehat{\mathbb{V}}(\widehat{\boldsymbol{\theta}}) \widehat{\boldsymbol{c}}}$ depends on the amount of information, since $\widehat{\mathbb{V}}(\widehat{\boldsymbol{\theta}})$ is the inverse of the estimated Fisher information matrix. From Eq. (23), we know that $\mathcal{I}(\boldsymbol{\theta})$ depends on the amount of data, i.e., $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ represent the available data. Thus, the confidence intervals estimated based on 5 and 15 years of data are different.

Going back to Figure 3, as expected, the width of the confidence interval is narrower when more information is available. It should be noted that the confidence intervals plotted in Figure 3 are for the 0.5 quantile; they should not be mixed up with the gray band plotted on the x–y plane of Figure 2.

3.4. Degradation Measurement Planning Using a Simple Test Plan

The discussion so far has assumed that the degradation measurements are available at the time of analyses. However, from a planner's perspective, it is necessary to decide the type of experiments, number of measurements and number of test units, as degradation monitoring is often associated with financial constraints. The cost of LE includes equipment cost and overheads. While it is common to employ monitoring systems at PV farms, when the farm scale becomes large, setting up a single weather station is most likely insufficient. On the other hand, gauging every sub-system may be an overkill. A trade-off needs to be made between degradation estimation accuracy and cost. The composition of the HE cost is different from that of the LE. HE in PV degradation is indoor I–V measurements. Once the modules/systems are deployed, it becomes difficult to conduct I–V measurements especially when the installation is remote. If the manufacturers were to conduct the degradation studies, shipping cost is predominant. In either of the LE or HE case, the number of

measurements and number of units are of interest.

We use the standard error of degradation quantile to reflect the degradation estimation accuracy. As shown in Figure 2, the confidence interval of degradation quantile is a function of time t. From Eq. (44), it is seen that this time dependency is originated from the standard error, i.e., $\widehat{\mathbf{se}}(\widehat{d}_p)$ is also a function of time. However, it should be noted that although the $\widehat{\mathbf{se}}$ at different time $t = 1, \cdots, 24$ are different, the discussion below is applicable for all t. We use $t = 15$ as an example, i.e., the degradation quantile at the end of 15 years is of interest.

Figure 4 (a) shows the contour plot of the estimated standard error of the estimated 0.50 quantile for an HE, $\widehat{\mathbf{se}}(\widehat{d}_{0.50})$, at the end of the evaluation period. The contour plot can be interpreted as follows. The case of $n = 3$ and $m = 3$ corresponds to the situation whereby three units are each measured three times during the course of 15 years at $t = 0$, $t = 7.5$ and $t = 15$, respectively. Under this setting, the estimated standard error is $\widehat{\mathbf{se}}(\widehat{d}_{0.50}) = 0.95$, reflected by the contour line at the bottom left corner of Fig. 4 (a). Similarly, $\widehat{\mathbf{se}}(\widehat{d}_{0.50}) = 0.5$, a smaller standard error, is found for the setup with $n = 11$ units and $m = 3$ measurements. It can be concluded from the "vertical" contour lines that in the HE simulation, fewer number of measurements (per module) can be used without losing much accuracy. Instead, the degradation standard error relies more on the number of units used in HE.

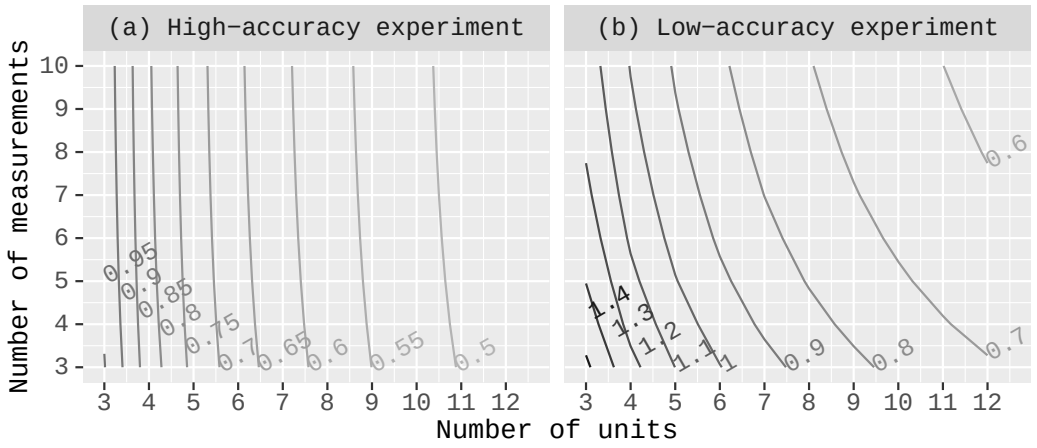

Figure 4. Contour plot of $\widehat{\mathbf{se}}(\widehat{d}_{0.50})$ at the end of evaluation period using different number of units and different number of measurements.

For a comparison of HE and LE errors, Figure 4 (b) shows the contour plot of the estimated standard error for an LE setting. The time of evaluation is also set as $t = 15$; and the contour lines correspond to $\widehat{\mathbf{se}}(\widehat{d}_{0.50})$. In contrast to the HE, standard error for the LE is more sensitive to the number of measurements. The improvement in accuracy can be achieved in the LE by increasing either the number of measurements or the number of units. However, it is observed that for same m and n values, LE yields a higher standard error than HE. This is largely owing to the higher measurement uncertainties embedded in the LE.

Figure 4 shown above demonstrates how the number of units (n) and measurements (m) can affect the standard error estimation, and thus affect the planning method for degradation studies. However, 15 years (as used in Figure 4) may be considered long for a PV degradation study; a shorter period of study such as 5 or 10 years may be more appropriate. On this point, we analyze the extrapolation of degradation quantile standard error below. In other words, we are going to look at how $\widehat{\mathbf{se}}(\widehat{d}_p)$ will be projected (up to 25 years for example) when different runtimes of the experiment are considered.

Fig. 5 (a) shows the standard error of the 0.5 quantile, $\widehat{d}_{0.5}$, as a function of time, for different HE studies. The HE degradation studies are assumed to be conducted over periods of 5, 10 and 15 years. In all three simulations, the number of units is set as 12, and the degradation for each unit is measured once every year throughout the period of study. For example, for the 5 year study, a total of 60 measurements are made to compute the standard error curve shown in Fig. 5 (a). Based on the available data of each case, the standard errors are extrapolated using the degradation model for the remaining years of a typical PV lifetime. It can be seen that the standard error projection at 25 years decreases as the runtime of the experiment increases.

The simulation result of the LE with the same degradation measurement settings is shown in Fig. 5 (b). With no surprise, the standard errors of the LE are higher than those of HE. Nevertheless, it is found that the standard error from the LE is comparable to that of the HE when the study period is long enough, such as a period of 15 years.

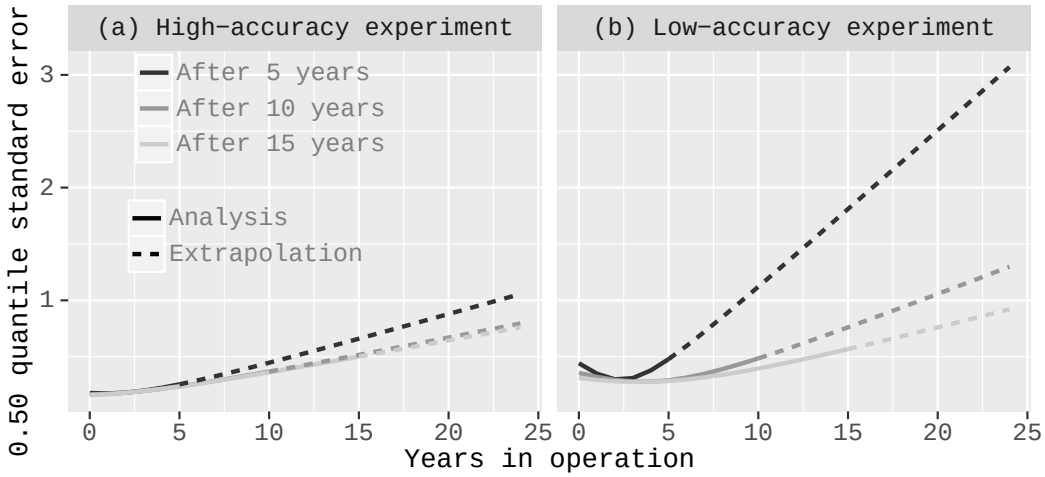

Figure 5. $\widehat{\mathbf{se}}(\widehat{d}_{0.50})$ as functions of time for the HE and LE.

The above simple test plan enables PV module manufacturers to plan the degradation studies effectively. Although the examples were given for 0.5 quantile, the approach can be readily applied to any quantile. For example, a similar analysis on the 0.001 quantile can be useful during warranty policy making. The particular choice of experiment and setup can be decided by experts based on some specific tolerable upper bound of the standard error. Together with the above mentioned cost constraints for HE and LE, the problem can be considered as a multi-objective optimization task. However, the solution to this task is

not within the scope of this chapter.

4. Conclusion

A practical PV degradation model is introduced in this chapter. The degradation model is based on the linear mixed effect model and contains six parameters; it enables flexible simulation and design exercises for photovoltaic degradation. Instead of using the conventional regression based methods for gradient estimation, maximum likelihood estimation is used to identify the degradation rate together with other parameters, simultaneously. The degradation model used here not only provides accurate estimates of the most used parameter, the degradation rate, it also gives quantile estimates of PV degradation. The quantile information gives additional insights for making warranty policy.

Three types of PV degradation experiments are discussed in detail. The low-accuracy experiments are performed in the outdoor conditions; their accuracy is limited by various uncertainties primarily due to ambient weather condition and soiling, which can otherwise be controlled in an indoor measurement environment. It is found that the accuracy of the high-accuracy experiment is rather independent of the number of measurements made over the period of a degradation study. To improve the accuracy of an HE, more degradation test units should be considered. On the other hand, the accuracies of the LEs depend on both the number of measurements and number of units.

All the degradation experiments described in this chapter depend on the length of the evaluation period, number of measurements and cost. Therefore, due the design of an degradation experiment, a multi-objective optimization could be formulated.

References

[1] E. Meza, "IEA PVPS: Installed PV capacity at 227 GW worldwide." http://www.pv-magazine.com/news, Apr 2016.

[2] D. C. Jordan and S. R. Kurtz, "Photovoltaic degradation rates – an analytical review," *Progress in Photovoltaics: Research and Applications*, vol. 21, no. 1, pp. 12–29, 2013.

[3] D. C. Jordan, S. R. Kurtz, K. VanSant, and J. Newmiller, "Compendium of photovoltaic degradation rates," *Progress in Photovoltaics: Research and Applications*, vol. 24, no. 7, pp. 978–989, 2016.

[4] A. Ndiaye, A. Charki, A. Kobi, C. M. Kèbé, P. A. Ndiaye, and V. Sambou, "Degradations of silicon photovoltaic modules: A literature review," *Solar Energy*, vol. 96, no. 0, pp. 140–151, 2013.

[5] A. Skoczek, T. Sample, and E. D. Dunlop, "The results of performance measurements of field-aged crystalline silicon photovoltaic modules," *Progress in Photovoltaics: Research and Applications*, vol. 17, no. 4, pp. 227–240, 2009.

[6] E. D. Dunlop and D. Halton, "The performance of crystalline silicon photovoltaic solar modules after 22 years of continuous outdoor exposure," *Progress in Photovoltaics: Research and Applications*, vol. 14, no. 1, pp. 53–64, 2006.

[7] J. H. So, Y. S. Jung, G. J. Yu, J. Y. Choi, and J. H. Choi, "Performance results and analysis of 3-kW grid-connected PV systems," *Renewable Energy*, vol. 32, no. 11, pp. 1858–1872, 2007.

[8] S. H. Alawaji, "Evaluation of solar energy research and its applications in saudi arabia-20 years of experience," *Renewable and Sustainable Energy Reviews*, vol. 5, no. 1, pp. 59–77, 2001.

[9] T. Ishii, K. Otani, and T. Takashima, "Effects of solar spectrum and module temperature on outdoor performance of photovoltaic modules in round-robin measurements in Japan," *Progress in Photovoltaics: Research and Applications*, vol. 19, no. 2, pp. 141–148, 2011.

[10] G. Makrides, B. Zinsser, G. Georghiou, M. Schubert, and J. Werner, "Degradation of different photovoltaic technologies under field conditions," in *Photovoltaic Specialists Conference (PVSC), 2010 35th IEEE*, pp. 002332–002337, June 2010.

[11] G. Makrides, B. Zinsser, G. Georghiou, M. Schubert, and J. Werner, "Evaluation of grid-connected photovoltaic system performance losses in cyprus," in *Power Generation, Transmission, Distribution and Energy Conversion (MedPower 2010), 7th Mediterranean Conference and Exhibition on*, pp. 1–7, Nov 2010.

[12] B. Marion, J. Adelstein, K. Boyle, H. Hayden, B. Hammond, T. Fletcher, B. Canada, D. Narang, A. Kimber, L. Mitchell, G. Rich, and T. Townsend, "Performance parameters for grid-connected PV systems," in *Proceedings of the 31st PV Specialists Conference*, (Lake Buena, FL, USA), pp. 1601–1606, 2005.

[13] N. Cereghetti, E. Bura, D. Chianese, G. Friesen, A. Realini, and A. Rezzonico, "Power and energy production of PV modules statistical considerations of 10 years activity," in *Photovoltaic Energy Conversion, 2003. Proceedings of 3rd World Conference on*, vol. 2, pp. 1919–1922 Vol.2, May 2003.

[14] S.-L. Chuang, A. Ishibashi, S. Kijima, N. Nakayama, M. Ukita, and S. Taniguchi, "Kinetic model for degradation of light-emitting diodes," *Quantum Electronics, IEEE Journal of*, vol. 33, pp. 970–979, Jun 1997.

[15] M. Vázquez and I. Rey-Stolle, "Photovoltaic module reliability model based on field degradation studies," *Progress in Photovoltaics: Research and Applications*, vol. 16, no. 5, pp. 419–433, 2008.

[16] G. Verbeke, G. Molenberghs, and D. Rizopoulos, "Random effects models for longitudinal data," in *Longitudinal research with latent variables* (K. van Montfort, J. H. L. Oud, and A. Satorra, eds.), pp. 37–96, Berlin: Springer, 2010.

[17] J. J. Faraway, M. A. Tanner, J. Zidek, and C. Chatfield, *Extending the Linear Model with R*. Routledge, 2004.

[18] G. Verbeke and G. Molenberghs, *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag, 2000.

[19] R. I. Jennrich and M. D. Schluchter, "Unbalanced repeated-measures models with structured covariance matrices," *Biometrics*, vol. 42, no. 4, pp. 805–820, 1986.

[20] L. Wasserman, *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2003.

[21] R. B. Millar, *Maximum Likelihood Estimation and Inference: With Examples in R, SAS and ADMB*. John Wiley & Sons, 2011.

[22] D. Jordan and S. Kurtz, "The dark horse of evaluating long-term field performance–data filtering," *Photovoltaics, IEEE Journal of*, vol. 4, no. 1, pp. 317–323, 2014.

[23] P. Z. G. Qian and C. F. J. Wu, "Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments," *Technometrics*, vol. 50, no. 2, pp. 192–204, 2008.

[24] Q. Zhang, X. Deng, P. Z. G. Qian, and X. Wang, "Spatial modeling for refining and predicting surface potential mapping with enhanced resolution," *Nanoscale*, vol. 5, pp. 921–926, 2013.

[25] R. A. Escobar, C. Cortés, A. Pino, E. B. Pereira, F. R. Martins, and J. M. Cardemil, "Solar energy resource assessment in Chile: Satellite estimation and ground station measurements," *Renewable Energy*, vol. 71, pp. 324–332, 2014.

[26] D. C. Jordan, "NREL PV module reliability workshop." http://www.nrel.gov/pv/pvmrw.html, Feb 2011.

[27] H. Haeberlin and C. Beutler, "Normalized representation of energy and power for analysis of performance and on-line error detection in PV systems," in *the 13th Eur. Photovoltaic Sol. Energy Conf.*, (Nice, France), 1995.

[28] D. C. Jordan and S. R. Kurtz, "PV degradation risk," in *World Renewable Energy Forum*, (Denver, Colorado), May 2012.

[29] C. Jennings, "PV module performance at PG&E," in *Photovoltaic Specialists Conference, 1988., Conference Record of the Twentieth IEEE*, pp. 1225–1229 vol.2, 1988.

[30] S. Pulver, D. Cormode, A. Cronin, D. Jordan, S. Kurtz, and R. Smith, "Measuring Degradation Rates Without Irradiance Data," Tech. Rep. NREL/CP-5200-47597, National Renewable Energy Laboratory, Golden, CO, USA, 2010.

[31] R. M. Smith, D. C. Jordan, and S. R. Kurtz, "Outdoor PV module degradation of current-voltage parameters," Tech. Rep. NREL/CP-5200-53713, National Renewable Energy Laboratory, Golden, CO, USA, 2012.

[32] T. Dierauf, A. Growitz, S. Kurtz, J. L. B. Cruz, E. Riley, and C. Hansen, "Weather-corrected performance ratio," tech. rep., NREL, April 2013.

[33] L. H. I. Lim, Z. Ye, J. Ye, D. Yang, and H. Du, "A linear identification of diode models from single I–V characteristics of PV panels," *IEEE Transactions on Industrial Electronics*, vol. 62, pp. 4181 – 4193, July 2015.