Quantification of Environmental Effects on PV Module Degradation: A Physics-Based Data-Driven Modeling Method

Arun Bala Subramaniyan, *Member*, *IEEE*, Rong Pan, *Member*, *IEEE*, Joseph Kuitche and GovindaSamy TamizhMani

***Abstract*— This paper explains the fusion of the physics-based material degradation mechanism with the statistics-based data modeling approach for predicting the degradation rate of photovoltaic (PV) modules. The degradation of PV module is mainly associated with the module construction type and climatic condition at its use location. The aim of this paper is to quantify the effect of dynamic environmental stresses (dynamic covariates) on the power degradation of the module over its lifetime. There are various physics-based models, such as Arrhenius model, for understanding the physical or chemical reaction-related root causes of PV degradation. But, to estimate the underlying material properties, such as activation energy (Ea), statistical modeling plays a key role. In addition, instead of being continuously monitored, the performance characteristics of PV modules are often measured only at quarterly or annually intervals, which makes it difficult to model the complete degradation path of the module. On the other hand, the information on dynamic covariates are recorded more frequently with the development of sophisticated sensors and data acquisition systems. This information can be integrated through physics-based models to study the effects of environmental variables in degradation processes. Hence, in this paper, a cumulative exposure model is used to link the module degradation path and the environmental variables, including module temperature (both static and cyclic), ultra-violet radiation, and relative humidity, which are recorded as multivariate time series.**

***Index Terms*— Cumulative effects model, Environmental effects on PV degradation, PV module reliability quantification.**

I. Introduction

THE power degradation in PV modules can be due to numerous degradation modes that are built up within the module over time. Due to the rapid evolution of PV technology and relatively young age of most of the PV systems around the world, there is not sufficient data available to study various failure modes occurring in the field. A brief review of PV degradation modes, field degradation rates, and their respective accelerated test procedures, standards, and results are summarized in [1], [2], and [3]. Around 86 different types of defects have been reported in the field due to poor manufacturing, dynamic environmental stresses in the field etc. A survey on the risk levels related to different defects that occurred in field aged PV module has been reported in [4], [5]. Dynamic stresses due to varying climatic conditions in the field, especially due to static and cyclic temperature, ultra-violet (UV) radiation, and relative humidity (RH) are the primary causes of various degradation modes. Also, extensive field analysis from [5], [6] indicates that these degradation modes affect individual performance parameters such as short circuit current (Isc), open circuit voltage (Voc) and fill factor (FF), which directly contribute to output power (Pmax) as Pmax is the product of Isc, Voc and FF. Overall, the power degradation in PV modules can be traced to the cumulative effect of all these numerous degradation modes due to the module exposure to harsh environmental conditions.

Though the use environment has significant impacts on the long-term performance of PV systems, there has been not much work done on quantifying the effect of environmental factors on PV module degradation because of the sparsity of PV field performance data in the past. However, the long-term market penetration of crystalline silicon modules and technological advancements in data collection systems have recently made more and more performance data of PV modules available to the research community. Koyler, Mann and Farrar [7] incorporated the environmental effects on the degradation modeling of PV modules (especially for solar cell encapsulants) using accelerated test data. In their study, a discrete environmental cell approach was used with each cell covering the distinct interval of various environmental factors. The number of hours over which the combination of weather conditions occurred was calculated. For instance, if an environmental cell range is arbitrarily defined to be 60°C to 70°C module temperature and 60% to 70% relative humidity, the number of hours over which these conditions occur in field can be obtained from weather data. The degradation rate for the corresponding cell can be determined by using the physics-based rate equations. Then, the degradation rate is multiplied by

This work was primarily funded by the SunShot Program of Department of Energy under PREDICTS 2 and SERIIUS.

The authors are with Arizona State University, AZ 85281 (e-mail: abalas18@asu.edu, rong.pan@asu.edu, kuitche@asu.edu, manit@asu.edu).

number of hours to determine the amount of degradation in that specific environmental cell. This process is repeated for each cell and the summation of all cells determines the cumulative degradation.

Combining the degradation information collected over a period of time, along with the environmental cell approach and rate equations, the effects of various environmental conditions like temperature (activation energy), UV, etc., can be estimated. The importance of this estimation is that, once the model estimates are known, then the degradation of corresponding module technology can be predicted for any location using the available climatic information. References [3] and [8] provide a detailed investigation of degradation of various module technologies installed across diverse climatic conditions. But this study has not reported the quantitative effects of environmental variables on degradation rates, which requires the development of appropriate accelerated life tests to study individual degradation modes. A pitfall in using the discrete cell or environmental binning approach, as in [7], is that this method provides only an approximation for overall degradation. Improved results can be obtained using continuous degradation rate model since the dynamic environmental factors (or dynamic covariates) are measured continuously over time. However, in most cases, the module performance measurements are taken only in discrete intervals like quarterly or annually, making it difficult to build a continuous time degradation rate model.

The incorporation of time-varying covariates into degradation models have been previously discussed via Bayesian linear degradation path models [9] and stochastic degradation models [10]. Nelson [11] introduced the idea of incorporating the time-varying covariates into the induced cumulative distribution function (cdf) of the degradation process through a cumulative damage model. The general statistical models and data analysis methods of using accelerated test data to predict population reliability in service are explained in [12], [13]. References [13] and [14] extended this approach by using B-spline models to incorporate the unknown effect of dynamic covariates on the degradation behavior. But this approach cannot be directly adopted for PV module degradation modeling. The spline model cannot reveal the underlying factors like activation energy and other parameters. As these factors are material properties, any estimates should be reasonable to the specific material under study. In most cases, the unavailability of these parameters pose a serious hindrance to the determination of acceleration factors for various acceleration tests involving temperature, UV, relative humidity, etc. In addition, a good review of physics-based models such as Arrhenius model, Peck Model, Coffin-Manson model, etc. for studying the effects of various environmental factors is discussed by Escobar and Meeker [15]. Since, the lifetime of PV modules depends on dynamic environmental factors, these physics-based models could aid in determining the functional form for covariate effects in building the cumulative damage model for PV modules.

In literature, Pan, Kuitche and TamizhMani [16] showed that information about dynamic covariates can be incorporated to model the lifetime of PV modules with known degradation path. But, the study considered the average effect of dynamic covariates with the only factor being ambient temperature. Most of the failure modes occur due to multiple varying environmental stresses such as temperature, UV irradiance, and humidity. In real world, all these environmental stresses apply on PV modules simultaneously and have an effect on module performance. In addition, considering module temperature and plane of array (POA) irradiance will provide a better information than considering ambient temperature or horizontal irradiance. Hence, this paper presents an approach to PV degradation modeling that includes the effects of module-specific stresses (module static temperature, module cyclic temperature) and other environmental stresses (ambient relative humidity and plane of array UV irradiance). At first, various models to determine these module-specific stressors are discussed, and then, reliability estimation using cumulative damage model is presented. Ordinary Least Squares (OLS) method is used for parameter estimation. The M55 mono-crystalline silicon modules are considered for this study. Note that the c-Si technology accounts for more than 92% of PV-market [17]. But, our methodology can be applied for any type of module, provided that the degradation data and covariate information are available. We demonstrate that this hybrid approach of combining both physics-based models and data-driven statistical inferences is effective for incorporating the joint effect of dynamic covariates and quantifying the field degradation rate of PV module.

In brief, the motivation of this paper is to build a good model for lifetime prediction of outdoor PV modules. The challenge is to estimate the effects of dynamic environmental variables on module degradation with inconsistent degradation data. Our method has two major advantages. First, for the field deployed modules, integration of environmental factors into the degradation model helps with a thorough understanding of the degradation process behavior. Secondly, the estimates of activation energy and other factors will be a useful metric for designing accelerated life testing profiles for PV modules.

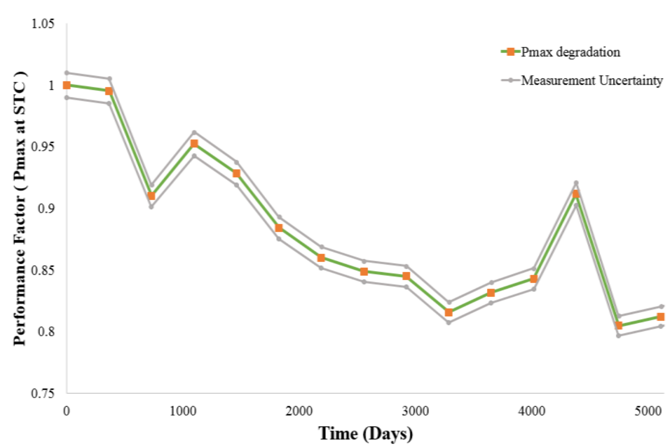

Fig. 1. Power Degradation from 1998 to 2012 (5100 days)

II. Data Description

The performance data of PV systems with mono-crystalline silicon modules installed at Solar Energy Research Facility (SERF) at National Renewable Energy Laboratory (NREL) is used for this research. The site under study consists of two identical, 6 kW (AC) grid-connected photovoltaic systems located on the roof of SERF building in Golden, Colorado. The details of the system are given in Table I and the system performance parameters are available at the PVDAQ website [18].

The data from the period of 1998 to 2012 is used for our study. The annual dc power degradation is calculated for STC conditions, using temperature coefficient of -0.4334 %/°C for power-temperature correction [19]. To calculate the power degradation, the average STC dc power at first month's data output was considered as the initial value, which was only 6600 W, which confirms the fact that the system might be overrated by around 10% [19]. Overall, there are 280 modules, but the performance measures of individual modules are not available, only the overall system performance is available. It is assumed that performances of modules are independent and identically distributed (iid) and the percentage degradation of the entire system is proportional to the percentage degradation of individual module. But in reality, the module degradation percentage will be lower than the overall system degradation percentage due to subcomponents, connections, and other external reasons such as invertor failure, module mismatch, and soiling. They all contribute to additional degradation of the system. The average degradation of two systems with measurement system uncertainty of 1% (280 modules) is shown in Fig. 1. Note that the average degradation is around 1.3% per year with y-axis representing the percentage of initial power available (converted to a scale of 0 to 1) with respect to time along x-axis.

The 30-minute-interval weather data for the site, where the modules are installed, are retrieved from National Solar Radiation Database (NSRDB) [20]. These data are split to two parts – one for building the weather model and the other for testing the weather model. The training dataset consists of data between 1998 and 2010 and the test dataset for validation is from 2011 and 2012.

The Sandia module temperature model given in (1) is used for calculating the module temperature [21]. This model accounts for ambient temperature, POA irradiance and wind speed which plays a major role in influencing module temperature. Reference [22] confirms the accuracy of Sandia module temperature model.

$$T_{mod}=T_{amb}+E_{POA}(e^{a+b(WS)}) \tag{1}$$

where T_{mod} is the module temperature in Celsius, T_{amb} is the ambient temperature in Celsius, and E_{POA} is the plane of array (POA) irradiance, WS is the wind speed (in m/s), a and b are constants equal to -3.75 and -0.075 respectively. The UV radiation is taken to be 5% of the POA irradiance (E_{POA}), which is calculated using (2)-(6) given by Sandia PV performance model [23].

TABLE I
System Details obtained from NREL database [18]

Description	NREL SERF-EAST (Site #50)	NREL SERF-WEST (Site #51)
Latitude	39.742 N	39.7416 N
Longitude	105.1727 W	105.1734 W
Elevation	1994.7 meters	1994.7 meters
Azimuth	158 °	158 °
Tilt	45 °	45 °
Rating (AC/DC)	6 kW/ 7.43 kW	6 kW/ 7.43 kW
Module Manufacturer	Siemens	Siemens
Module Type	M55 mono-crystalline silicon	M55 mono-crystalline silicon
Inverter Manufacturer	Fronius	Fronius
Inverter Type	IG 4500 - LV	IG 4500 - LV
Number of Modules	140	140
Latitude	39.742 N	39.7416 N
Longitude	105.1727 W	105.1734 W
Elevation	1994.7 meters	1994.7 meters
Azimuth	158 °	158 °

$$E_{POA}=E_b+E_g+E_d \tag{2}$$

where E_b is the plane of array beam component, E_g is the POA ground reflected component, and E_b is the POA sky-diffuse component. The POA beam component is calculated utilizing direct normal irradiance (DNI) and the angle of incidence (AOI) given in (3), where the AOI between PV array and sun's rays is determined using (4).

$$E_b=DNI\times \cos(AOI) \tag{3}$$

$$AOI=\cos^{-1}[\cos(\theta_Z)\cos(\theta_T)+\sin(\theta_Z)\sin(\theta_T)\cos(\theta_A-\theta_{A,array})] \tag{4}$$

where θ_A, θ_Z, θ_T and $\theta_{A,array}$ denotes solar azimuth angle, solar zenith angle, array tilt angle and array azimuth angle, respectively. The ground reflected component E_g, which is the part of radiation hitting a tilted surface and being reflected off the ground, is calculated using (5). The calculation involves three components: Global Horizontal Irradiance (GHI), albedo (the reflectivity of ground surface) and surface tilt angle ($\theta_{T,surf}$).

$$E_g=GHI\times albedo\times\frac{1-\cos(\theta_{T,surf})}{2} \tag{5}$$

There are several models available for calculating the sky diffusivity component of POA, such as Isotropic sky diffuse model, Hay and Davies sky diffuse model, Reindl sky diffuse model, Perez sky diffuse model, Sandia sky diffuse model, etc., [23], [24]. In this paper, the Sandia sky diffuse empirical model is used since it has been found to work well at Sandia facility compared to other models. The diffuse irradiance (E_d) using diffuse horizontal irradiance (DHI), global horizontal irradiance (GHI), surface tilt angle (θ_T), and sun zenith angle (θ_Z) is given by (6). Note that all angles are in degrees.

$$E_d = DHI \times \frac{(1+\cos(\theta_T))}{2} + GHI \times \frac{(0.012\theta_Z - 0.04) \times (1-\cos(\theta_T))}{2} \quad (6)$$

In our model, both static and cyclic effects of module temperature are included. The daily maximum module temperature is taken for static effect and the cyclic effect is calculated from difference between daily maximum and minimum module temperature.

There are various degradation rate equations available for studying the effects of environmental factors [15] and various studies report the use of the aforementioned physics-based equations in modeling individual failure modes due to environmental stressors [25], [26], [27]. For instance, the log-normal model (Arrhenius) is most commonly used to provide the functional form for temperature effect on PV modules. But most of these studies are based on only individual failure modes and they have been done mostly for indoor accelerated testing. This paper links the proven physics-based rate equations through cumulative damage model to predict the field reliability of PV modules. Next section will explain the cumulative damage model and the functional forms for dynamic covariates.

A. Data Notation

Before proceeding to the methodology section, the notations used in our degradation data model are explained herein. Let $\boldsymbol{X}(t) = [X_1(t), \dots, X_r(t)]'$ be the vector of covariates, i.e., the environmental stress variables in our study, and r is the number of covariates. Each covariate is actually a stochastic process; i.e., $\boldsymbol{X}(t) = \{\boldsymbol{X}(h): 0 \leq h < t\}$, which denotes the dynamic information recorded from time 0 to t. So, the value of covariate 'v' for module 'i' at time 'u' is denoted by $x_{iv}(u)$. Suppose there are 'z' modules, then for module 'i', the degradation measurements at time t_{ij} is denoted by $y_i(t_{ij})$, $i = 1, \dots, z$ and $j = 1, \dots, z_i$, where z_i is the number of time points when the degradation measurements were taken on module 'i'. The covariate process history for module 'i' is represented by $x_i(t_{i,z_i}) = \{x_i(h): 0 \leq h < t_{i,z_i}\}$. As there are r covariates, we have $x_i(h) = [x_{i1}(h), \dots, x_{ir}(h)]'$. In our case, the degradation is measured for the entire SERF system with a combined 280 modules and only overall performance of the systems is available. The average P_{max} degradation is taken as performance statistic and hence the value of z is 1. If individual module measurements were available, the value of 'z' could be larger than 1. References [13] and [14] discussed some random effect models that can be utilized then for modeling a group of modules in the same system.

III. Cumulative Exposure Model

This section explains the cumulative exposure model for quantifying the degradation path of PV modules. Once the modules are installed in field, they start producing power and, at the same time, degrade due to the exposure to sunlight and other environmental stresses, because continuous exposure to these environmental stress factors initiate defects such as encapsulant browning, solder bond degradation, etc., and over time the module power production decreases. We can use the information about environmental factors to calculate the instantaneous degradation using physics-based stress-effect function and then accumulate these stress effects over time to study the cumulative damage of environmental stress on module power output. The functional form of stress-effect function will be described in next subsection. So, if the power degradation (current-voltage curve) of field-installed modules is measured quarterly or annually and the local weather information is recorded continuously, the effects of weather stress variables on the module degradation need to be accumulated using the cumulative exposure model. Ideally, the predicted cumulative degradation should be equal to the actual degradation that is measured on quarter or annual basis. Using this cumulative damage model, we will be able to estimate the unknown model parameters, such as activation energy. In general, the actual degradation of the photovoltaic modules installed in the field is given by (7).

$$y(t) = G(t) + \varepsilon(t) \quad (7)$$

where $y(t)$ is module power degradation measurement at time t, $G(t)$ is the actual power degradation path, and $\varepsilon(t)$ is the model error term.

A PV module is considered failed once the cumulative degradation reaches a certain threshold specified by the customer (usually, a 20% drop from initial power output is considered as failure). The model given in (7) does not provide any information about the effects of environmental factors that are actually responsible for degradation. In particular, the actual degradation path of PV module is affected by the exposure to several environmental factors over time. Therefore, the cumulative exposure G for a module is given by (8) [28].

$$G = \int_0^t f[\boldsymbol{X}(s), \beta]\, ds \quad (8)$$

where the function $f[\boldsymbol{X}(s), \beta]$ represents the instantaneous effect of stress on performance of the module from both usage and environmental conditions at time 's'.

We can use the cumulative exposure model described in [11]–[14] to model the module degradation behavior with the incorporation of dynamic covariate information, recorded as multivariate time series. Therefore, a module degradation path is modeled and given in (9) and (10).

$$y(t_{ij}) = G[t_{ij}, x_i(t_{ij})] + \varepsilon_i(t_{ij}) \quad (9)$$

$$G[t_{ij}, x_i(t_{ij})] = \beta_{ini} + \int_0^t \prod_{l=1}^{p} f_l[x_{il}(s), \beta_l]\, ds \quad (10)$$

where β_{ini} is the initial level of power degradation, β_l the parameters in the covariate effect function $f(.)$, and $\varepsilon_i(t_{ij})$ the error term. Note that the function $f_l[x_{il}(s), \beta_l]$ denotes the instantaneous effect of the stress factor x_{il} on the module degradation at time s.

This cumulative damage modeling approach was studied by Nelson [11], [12] for the accelerated failure time model and adopted by Hong and Meeker [13] by modeling the effects through a general additive model. Pan, Kuitche and TamizhMani [16] simplified the cumulative damage model by using the average statistic of stress effect over the period of degradation measurement. The error term $\varepsilon_i(t_{ij})$ has been modeled as independent and identically distributed with normality assumption in [13], [14], but in this paper we do not assume any distribution and use Ordinary Least Squares (OLS) for parameter estimation. Now, the next challenge is to determine functional form for covariate effect transformation function $f(.)$ on module degradation, which is explained in the following subsection.

A. Functional Forms for Covariate Effects

There are two approaches to define the functional forms for covariate effects. The first approach is motivated by the models based on physics or chemistry based degradation rate equations. For instance, the Coffin-Manson model is most commonly used to provide the functional form for the effect of cyclic temperature on PV modules, provided that the information on temperature is available. The second approach is to use non-parametric methods, especially when the physics-based models do not work well for the given dataset or data are inadequate for validating the functional form. For example, Hong and Meeker [13] and Meyer [29] defined shape restricted spline basis functions to model the effects of dynamic covariates. But, the underlying parameters for failure mechanisms, such as activation energy, are difficult to describe by using spline basis functions. Since, the functional forms for all the environmental factors under study are available and have been used by various previous studies [16], [25] - [27], the physics-based functional forms are adopted for our research.

In terms of static and cyclic module temperature effect functions, the Arrhenius model given in (11) and Coffin Manson model given in (12) are used, respectively. The effect of relative humidity is modeled through Peck model given in (13) [15], [30]. The effect of UV can be modeled using a log-linear model as a function of wavelength, especially integrating over UV irradiance band. Since the information about wavelength and UV is not available directly, the UV irradiance is taken to be 5% of POA irradiance [31] and the functional form of its effect is modeled similar to Peck's model with Arrhenius term. This method has been used in accelerated testing to study the effect of encapsulant browning in PV modules due to UV exposure [31].

$$Rate(T)=\gamma_0 \times exp\left(\frac{-E_a}{k\times T}\right) \tag{11}$$

$$N=\frac{1}{(\Delta T)^{\alpha}} \tag{12}$$

$$Rate(T,RH)=\xi_0 \times exp\left(\frac{-E_a}{k\times T}\right)\times (RH)^{\beta} \tag{13}$$

Where "$Rate(.)$" denotes the reaction rate due to environmental factors such as temperature, humidity, etc. "N" is the number of cycles to failure, "E_a" is the Activation Energy (in eV) based on product or material characteristics. "k" is the Boltzmann Constant (8.62×10^{-5} eV/K), "T" is the Temperature (in Kelvin), "ΔT" is the cyclic temperature (in Kelvin), "RH" is the Relative Humidity (in %) and "$\gamma_0, \xi_0, \alpha, \beta$" are the constants based on product or material characteristics.

Motivated by the Norris-Landzberg approach [32], a combined functional effect model can be built. Reference [26] developed a similar model to study the effect of cyclic temperature on solder bond degradation. Hence, the functional form for the environmental stress-effect transformation function $f(.)$ is given by the physics-based model given in (14).

$$Rate(T,\Delta T, UV, RH)=\beta_0 \times exp\left(\frac{-\beta_1}{k\times T_{max}}\right)\times\left(\Delta T_{daily}\right)^{\beta_2}\times \left(UV_{daily}\right)^{\beta_3}\times\left(RH_{daily}\right)^{\beta_4} \tag{14}$$

Where "$Rate(T, \Delta T, UV, RH)$" is the reaction rate due to environmental factors such as static module temperature, cyclic module temperature, UV radiation and relative humidity. The parameter "k" is the Boltzmann constant (8.62×10^{-5} eV/K), "T_{max}" is the daily maximum temperature of module (in Kelvin), "ΔT_{daily}" is the daily cyclic temperature of module (in Kelvin), "UV_{daily}" is the daily daytime average UV irradiance (in W/m^2), and "RH_{daily}" is the daily average relative humidity (in %). In addition, the parameters to be estimated are "β_0" which is the frequency factor (in sec^{-1}), "β_1" the activation energy (in eV), "β_2" the effect of cyclic temperature, usually in the range of 2 to 5 [15]. "β_3" is the effect of ultra violet radiation, usually in the range of 0.6 to 1 [5], [31]. "β_4" is the effect of relative humidity, usually in the range of 0 to 2 [15], [26].

IV. Parameter Estimation

The method of Ordinary Least Squares (OLS), which aims at minimizing the squared loss, is used to obtain the estimates for unknown model parameters. However, the evaluation of integral in (10) is not trivial and involves special numerical quadrature methods to solve this problem. Since the degradation data and dynamic covariate information are measured as discrete data points, the discretized data model is given in (15).

$$y_i(t_{ij})=\beta_{ini}+\sum\beta_0\times\prod_{l=1}^{p} f_l[x_{il}(t_{ij});\beta_l]+\varepsilon_{ijl} \tag{15}$$

In our case, the value of 'p' is 4 since we are considering four environmental variables. The constraints on the parameters are given in (17)-(20). Note that all these parameters are based on material properties of the PV module. In our case, we assume that the module is new at installation, hence the value of 'β_{ini}' is 0. In some cases, the measured power is greater than the specified nameplate ratings, especially during the infant life period [3]. Reference [16] used a 10% threshold to account for this inherent uncertainty. In our model, 'β_{ini}' is set to -0.1 to account for various hidden uncertainties and get a reasonable prediction. The optimization problem is formulated as follows.

Objective:

$$Minimize \ \ (Y\text{-}\boldsymbol{X}\beta)^2 \tag{16}$$

Subject to:

$$0 \leq \beta_1 \leq 2 \tag{17}$$

$$2 \leq \beta_2 \leq 5 \tag{18}$$

$$0.6 \leq \beta_3 \leq 1 \tag{19}$$

$$0 \leq \beta_4 \leq 2 \tag{20}$$

where Y is the actual response and $\boldsymbol{X}\beta$ is the predicted value, with $\boldsymbol{X}$ being the covariate matrix and β the corresponding effects.

It is to be noted that the regression model cannot be linearized using natural log function since it involves cumulative effects; i.e., the integral (summation) hinders the possibility to linearize the model. There are several methods available to solve the constrained non-linear regression problem [33]. The algorithms based on the Nelder-Mead method, variable metric method, directed search method, and gradient method are mostly appropriate for specific set of problems, but they may require restrictive assumptions relating to existence of derivatives, unimodality, etc. [34], [35]. In addition, using these methods for non-linear regression may not converge to a global solution and it is complicated to deal with the gradients when the number of parameters becomes high. In addition, since the functional forms for covariate effects include constraints on parameters, it will be difficult to include all these constraints and simultaneously come up with reasonable initial starting point. A reasonable approach is to follow an evolutionary strategy for optimization. The genetic algorithm (GA) fits this purpose and a wide range of nonlinear regression models in finance, engineering, chemical sciences have been solved using GA [34]-[37]. Hence, a GA-based nonlinear parameter estimation is used in our study. The algorithm works by generating a possible set of parameter values termed as population, based on the given upper and lower bound. There is no requirement to calculate the gradient of performance functions due to which, the probability for a GA to become trapped in a local optimum is less compared to conventional gradient based methods.

In our study, the population size of 100 is used. Once the population set is determined, the individuals of the generation are evaluated for the given fitness function, in our case, fitness corresponds to sum of squared errors. The evolution process of the generation is driven by genetic operators, which follow the Darwinian principle, i.e., survival of the fittest. The common genetic operators are reproduction, crossover and mutation. The evolution operation is carried out once the fitness proportion (10% for our case) of the population is determined. Some members of the population is transformed using genetic operators namely reproduction, crossover and mutation, with assigned probabilities of 0.1, 0.2 and 0.1 respectively. A set of new population is formed and the process is repeated until convergence. Note that when the number of variables in genetic algorithm increases, the computational cost grows rapidly since more function evaluations will be needed and the rate of convergence will be slow.

TABLE II
PARAMETER ESTIMATES OF CUMULATIVE EFFECTS MODEL ALONG WITH CONFIDENCE INTERVAL

Parameter	Corresponding Description	Estimate	Lower 95% CI	Upper 95% CI
β0	Frequency Factor	0.35	0.21	0.50
β1	Activation Energy	0.70	0.63	0.76
β2	Cyclic Temperature effect	2.41	2.34	2.46
β3	UV effect	0.75	0.68	0.89
β4	RH effect	1.52	1.27	1.80

V. RESULTS & VALIDATION

The results of parameter estimation, uncertainty interval, predicted degradation plots are given in this section. The estimate of activation energy (β_1) is 0.70 ± 0.1 eV, which is reasonable for crystalline silicon module family. It is noticed that a very small change in the order of 0.01eV in the activation energy can produce a significant difference in degradation rate prediction, because β_1 is used in an exponential function. The values of model parameters are tabulated in Table II. The predicted degradation path and the actual degradation path of the training data set are shown in Fig. 2. Note that the parameter estimates are based on the training data set from 1998 to 2010. In order to validate the model and make it useful for prediction, the test data set containing the environmental stress information from 2011 to 2012 is used. Note that, there is an abnormal measured data shown in Fig. 2 but the exact reason is not known to eliminate it. Even though we could not explore more about the abnormal data point, it did not have significant influence in our regression model estimates as seen from the predicted path. The predicted degradation path, as well as its prediction interval, and the actual degradation path of the test dataset are also shown in Fig. 2.

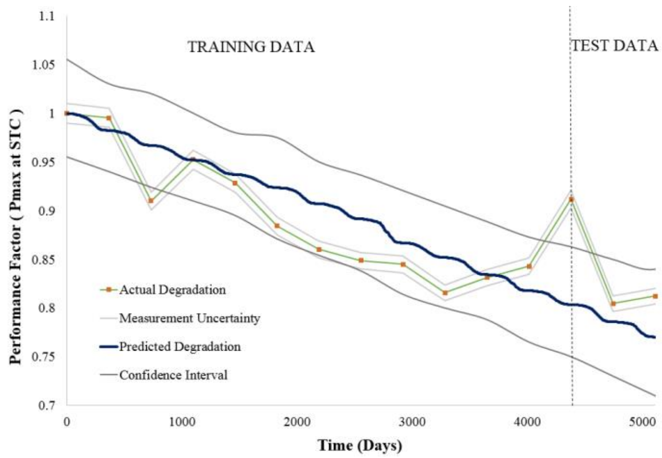

Fig. 2. Actual vs Predicted degradation of dataset from year 1998 to 2012

The predicted values are satisfactory given that only limited information about the degradation path is available. If the information of all the 280 modules are individually available, then a more accurate estimate can be obtained. In addition, there is not much information on defects, such as browning, soiling, etc., from the available field measured data. These hidden effects could also be affecting the module performance and might be causing variation in output power. Now, using our cumulative effect model it is possible to predict the behavior of monocrystalline silicon modules for any location in the world, as long as the local weather data are available. Fig. 3 shows the degradation paths for various locations across U.S., categorized to different climate zones. Table III provides additional details of these predictions.

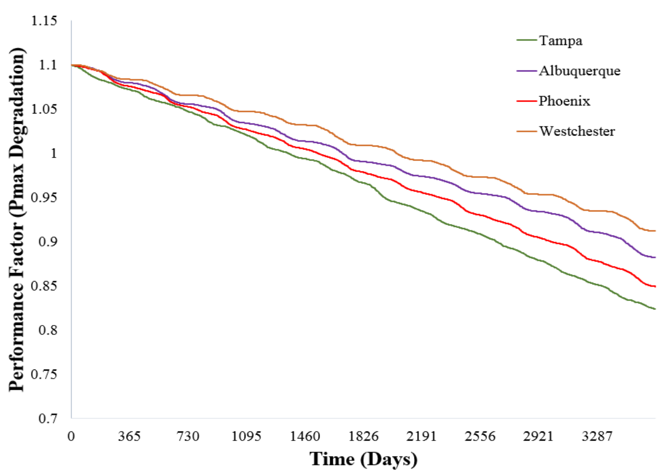

Fig. 3. Degradation plot of M55 x-Si modules at selected locations for 10 years (2006 - 2015)

In order to additionally validate the model results with real data, the M55 modules installed at two sites in Arizona (Mesa and Tempe) is considered, with data available at the PRL, ASU. The first site has 9 modules aged 18 years with a mean Pmax degradation of 1.45 ± 0.05% per year and median Pmax degradation of 0.96 ± 0.03% per year. The second site has 200 modules aged 18 years with a mean Pmax degradation of 2.35 ± 0.1% per year and median Pmax degradation of 2.40 ± 0.1% per year. Though both sites are close to each other and have similar weather characteristics, there is considerable difference in degradation rates. This can be due to several reasons like low sample size for one site, other losses due to soiling, browning, solder bond failures etc. As it is shown in Table III, the predicted degradation for Phoenix, AZ is about 1.50% per year with a wide interval and appears to be a reasonable estimate. The data for other regions are not available to validate, so it is not possible to arrive at a solid conclusion for other sites. But the model results show that the degradation is higher for hot-humid regions than for hot-dry regions, denoting that humidity plays a role in determining the degradation path.

For comparison, Tampa has similar weather conditions when compared to Phoenix but has a much higher humidity level, which resulted in a higher degradation rate in Tampa. But, an important point to remember is that the model is based on ambient relative humidity rather than the internal humidity inside the module. In addition, the data used for modeling and estimation is based on the degradation data of M55 modules with low sample size. A good robust estimate of factors like activation energy and other factors is still needed to generalize the results for the entire c-Si module family, due to difference in module construction (like wafer thickness, cell area and module area, etc). Also, the proposed methodology is tested only on four factors, but in reality there are other factors that could play a significant role in the performance data (e.g. microcracks). We assumed here the major environmental factors were temperature (Static and Cyclic), UV, and humidity which cause failure modes such as encapsulant browning, solder bond fatigue etc. However, power degradation can also occur due to failure modes like microcracks, especially due to significant reduction in wafer thickness of modern c-Si glass/polymer modules. The regression model established in the paper can be extended to incorporate such effects and this will be researched in our future study. Nevertheless, the objective of this paper was to establish a methodological approach for estimating the combined effects of multiple environmental factors and the prediction result presented in the paper validates our approach.

TABLE III
PERCENTAGE POWER DEGRADATION PER YEAR FOR M55 X-SI MODULES AT VARIOUS LOCATIONS

Region	Classification	Estimated degradation per year (%)	Lower 95% CI	Upper 95% CI
Phoenix	Hot and Dry	1.50 %	0.90 %	1.92 %
Westchester	Cold	0.88 %	0.63 %	0.95 %
Tampa	Hot and Humid	1.76 %	1.48 %	2.04 %
Albuquerque	Semi-Arid	1.17 %	0.63 %	1.47 %

VI. CONCLUSION

In this paper, a practical approach for integrating stochastic outdoor weather information to study PV module degradation using physics-based and data-driven modeling method is presented. The proposed approach is modeled and validated using actual monocrystalline silicon modules data collected over a long term. The advantage of cumulative effects modeling method is that it can be used to predict the degradation behavior of any module construction. Using this approach, it is easy to identify the influence of environmental factors and levels that can be used in designing or improving accelerated life testing conditions necessary for module lifetime predictions. The estimates of cumulative damage model provided in this paper holds for general class of monocrystalline silicon M55 module family and might vary for different manufacturers. In order to accommodate for different manufacturers, a random effects model can be incorporated. Also, the proposed methodology was demonstrated based on the performance data obtained with one-year time resolution. The accuracy of the prediction based on the proposed methodology may be improved by using the performance data obtained on a quarterly basis. But the quarterly performance measurements over multiple decades would become cost prohibitive along with frequent interruption of plant production. Since the quarterly degradation rate of c-Si modules is negligibly small and falls within the experimental measurement error, the one-year time resolution data is used. If higher-resolution, high-volume, multi-decade dataset is to be used, the user can run the proposed algorithm with a number of parallel clusters to get faster convergence. In addition, there is not much information about the individual performance parameters like short circuit current, open circuit voltage to study the specific defects like encapsulant discoloration, solder bond degradation etc., thus making it hard to understand the type of defects that were reasonable for the power degradation of the modules in our study.

To predict for future, the multivariate time series forecasting methods should be used for obtaining covariate information and then predict the module degradation. There are several univariate and multivariate time series modeling approaches available to forecast weather conditions. But it is not possible to conclude a single model will work for all locations. Also, it is highly complex but not impossible to forecast the weather conditions for 25 years and beyond. As an alternative approach, the user can take advantage of the Typical Meteorological Year (TMY) datasets for approximate and long-term predictions. These TMY datasets are constantly updated and are available at NSRDB database [20]. In addition, although the covariates such as temperature, UV radiation and relative humidity are commonly used as the major factors in module degradation, there are several other factors such as wind, snowfall, etc., that should not be ignored for certain climatic conditions. Also, in some cases, if the functional form of the covariates is not known explicitly, a complete data driven approach for quantifying the degradation path could be implemented.

References

[1] G. TamizhMani and J. Kuitche, "Accelerated lifetime testing of photovoltaic modules," *Solar America Board for Codes and Standards*, pp. 1-33, July 2013.

[2] J. Kuitche, R. Pan and G. TamizhMani, "Investigation of dominant failure modes for field-aged crystalline silicon PV modules under desert climatic conditions," *IEEE Journal of Photovoltaics*, vol. 4, no. 3, pp. 814-826, May 2014.

[3] D. C. Jordan and S. R. Kurtz, "Photovoltaic degradation rates—an analytical review," *Progress in Photovoltaics: Research and Applications*, vol. 21, pp. 12-29, Jan. 2013.

[4] M. K. Moorthy, "Automation of risk priority number calculation of photovoltaic modules and evaluation of module level power electronics," MS Thesis, Arizona State University, USA, 2015.

[5] S. Pore, "Reliability of PVmodules: Dependence on manufacturing quality and field climatic conditions," MS Thesis, Arizona State University, USA, 2017.

[6] J. H. Wohlgemuth, D. W. Cunningham, A. M. Nguyen and J. Miller, "Long term reliability of PV modules," *In Proc. EUPVSEC*, Paris, 2005.

[7] J. M. Kolyer, N. R. Mann and J. Farrar, "Final report on accelerated/abbreviated test methods for predicting life of solar cell encapsulants" DOE/JPL/954458-10, 1978.

[8] C. Raupp, C. Libby, S. Tatapudi, D. Srinivasan, J. Kuitche, B. Bicer and G. TamizhMani, "Degradation rate evaluation of multiple PV technologies from 59,000 modules representing 252,000 modules in four climatic regions of the United States," *In Proc IEEE PVSC*, Portland, USA, 2016

[9] N. Gebraeel and J. Pan, "Prognostic degradation models for computing and updating residual life distributions in a time-varing environment," *IEEE Transactions on Reliability*, Vol. 57, pp. 539–550, Dec. 2008.

[10] V. Bagdonavicius, I. Masiulaityte and M. S. Nikulin, "Reliability estimation from failure-degradation data with covariates," *Advances in Degradation Modeling*, pp. 275-291, 2010.

[11] W. Nelson, "Prediction of field reliability of units, each under differing dynamic stresses from accelerated test data," *Handbook of Statistics*, vol. 20, pp. 611-621, 2001.

[12] W. Nelson, "Accelerated Lifetime Testing: Statistical models, Test Plans and Data Analyses," John Wiley & Sons, USA, 1990.

[13] Y. Hong, Y. Duan, W. Q. Meeker, D. L. Stanley and X. Gu, "Statistical Methods for Degradation Data with Dynamic Covariates Information and an Application to Outdoor Weathering Data," *Technometrics*, vol. 57, no. 2, pp. 180-193, Apr. 2014.

[14] Z. Xu, Y. Hong and R. Jin, "Nonlinear general path models for degradation data with dynamic covariates," *Applied Stochastic Models in Business and Industry*, vol. 32, no. 2, pp. 153-167, Mar. 2016.

[15] L. A. Escobar and W. Q. Meeker, "A review of accelerated test models," *Statistical science*, pp. 552-577, Nov. 2006.

[16] R. Pan, J. Kuitche and G. TamizhMani, "Degradation analysis of solar photovoltaic modules: Influence of environmental factor," *Reliability and Maintainability Symposium*, pp. 1-5, Jan. 2014.

[17] Photovoltaics Report, "Fraunhofer Institute for Solar Energy Systems, ISE with support of PSE AG," Freiburg, 2017.

[18] PVDAQ: Photovoltaic data acquisition system website: https://maps.nrel.gov/pvdaq/.

[19] E. E. Van Dyk, T. R. Strand and R. Hansen, "SERF Photovoltaic Systems: Technical Report on System Performance for the period 1 August 1994 - 31 July 1995," National Renewable Energy Laboratory, USA, 1995.

[20] National Renewable Energy Laboratory's NSRDB Website: https://nsrdb.nrel.gov/.

[21] D. L. King, W. E. Boyson and J. A. Kratochvil, "Photovoltaic array Performance model," Sandia National Laboratories report, USA, 2004.

[22] S. Kurtz, K. Whitfield, D. Miller, J. Joyce, J. H. Wohlgemuth, M. Kempe, N. Dhere, N. Bosco, T. Zgonena, "Evaluation of High Temperature Exposure of Rack-Mounted Photovoltaic Modules," *In Proc. IEEE PVSC*, pp. 2399–2404, Philadelphia, 2009.

[23] PV Performance Modeling Collaborative website: https://pvpmc.sandia.gov/.

[24] M. Lave, W. Hayes, A. Pohl and C. W. Hansen, "Evaluation of global horizontal irradiance to plane-of-array irradiance models at locations across the United States," *IEEE Journal of Photovoltaics*, vol. 5, no. 2, pp. 597-606, Mar. 2015.

[25] A. Bala Subramaniyan, R. Pan, J. Kuitche and G. TamizhMani, "Prediction of acceleration factor of photovoltaic modules installed around the world," *Reliability and Maintainability Symposium*, pp. 1-6, Jan. 2018.

[26] N. Bosco, T. J. Silverman and S. Kurtz, "Climate specific thermomechanical fatigue of flat plate photovoltaic module solder joints," *Microelectronics Reliability*, vol. 62, pp. 124-129, Jul. 2016.

[27] G. M. Kimball, S. Yang and A. Saproo, "Global acceleration factors for damp heat tests of PV modules," *In Proc. IEEE PVSC*, Portland, June 2016.

[28] Y. Hong and W. Q. Meeker, "Field-failure predictions based on failure-time data with dynamic covariate information," *Technometrics,* vol. 55, no. 2, pp. 135-149, May 2013.

[29] M. C. Meyer, "Inference using shape-restricted regression splines," The Annals of Applied Statistics, vol. 2, pp. 1013–1033, Mar. 2008.

[30] D. S. Peck, "Comprehensive model for humidity testing correlation," *In Proc. International Reliability Physics Symposium*, pp. 44–50, Apr. 1986.

[31] ASU-PRL Report, "Prediction of climate-specific degradation rate for photovoltaic encapsulant discoloration," Photovoltaic Reliability Laboratory, Arizona State University, USA, 2017.

[32] K. C. Norris and A. H. Landzberg, "Reliability of controlled collapse interconnections," *IBM journal of research and development*, vol. 13, no. 3, pp. 266-271.

[33] J. C. Nash and M. Walker-Smith, "Nonlinear parameter estimation," Marcel Decker, New York, 1987.

[34] Z. Pan, Y. LishanKang, and G. Z. Liu, "Parameter estimation by genetic algorithms for nonlinear regression," *High Technology*, pp. 946 – 953, 1995.

[35] T. J. Vandernoot and I. Abrahams, "The use of genetic algorithms in the non-linear regression of immittance data," *Journal of Electroanalytical Chemistry*, vol. 448, no. 1, pp. 17–23, May 1998.

[36] S. Mohan, "Parameter estimation of nonlinear muskingum models using genetic algorithm," *Journal of Hydraulic Engineering*, vol. 123, no. 2, pp. 137-142, Feb. 1997.

[37] D. B. Hibbert, "Genetic algorithms in chemistry," *Chemometrics and Intelligent Laboratory Systems*, vol. 19, no. 3, pp. 277-293, Jan. 1993.