A Bayesian Inference Approach to Extract Circuit Model Parameters and Analyze Photovoltaic Degradation from Power Production Data

Joseph Chakar,[1,2,*] Jean-Paul Calin,[1] Marko Pavlov,[3] Jordi Badosa,[4] Jean-Baptiste Puel,[2,5] and Yvan Bonnassieux[1,2]

[1] Laboratoire de Physique des Interfaces et des Couches Minces (LPICM), Centre National de la Recherche Scientifique (CNRS), Ecole Polytechnique, Institut Polytechnique de Paris (IP Paris), Palaiseau, 91120, France

[2] Institut Photovoltaïque d'Île-de-France (IPVF), 18 Boulevard Thomas Gobert, 91120, Palaiseau, France

[3] Feedgy, Paris, 75009, France

[4] Ecole Polytechnique, Institut Pierre-Simon Laplace (IPSL), IP Paris, Ecole Normale Supérieure (ENS), Sorbonne Université, CNRS, Laboratoire de Météorologie Dynamique (LMD), Palaiseau 91120, France

[5] EDF, R&D, 18 Boulevard Thomas Gobert, 91120, Palaiseau, France

* Correspondence: joseph.chakar@polytechnique.edu

Abstract

Power production data can be a valuable resource to analyze photovoltaic (PV) performance without the need for field surveys. Recent work has demonstrated the exciting possibility of leveraging this data to extract circuit model parameters and current-voltage properties of a PV system, but further development is needed to promulgate these findings. Here, instead of using the classical frequentist approach, we switch to the Bayesian framework to solve this complex problem. This allows us to construct probability distributions over the model parameters, get a comprehensive picture of the solution space, and quantify prediction uncertainty. As a result, we can define confidence intervals for the system's electrical properties and consistently track their daily evolution. Our results are validated with laboratory measurements for five silicon and thin-film modules, and our scalable approach works with onsite as well as online weather data, which opens new prospects for remote PV monitoring, modeling, and degradation analysis for real-life applications.

Keywords

Solar energy, Photovoltaics, Bayesian inference, Machine Learning, Single-diode model, Performance analysis, Degradation, Silicon modules, Thin-film modules

Introduction

As the world photovoltaic (PV) fleet continues to grow and age, greater attention is being accorded to operations and maintenance (O&M) practices to ensure the technical and economic viability of solar energy systems over their lifetime. In order to meet production thresholds over more than 20 years, it is becoming increasingly critical to effectively monitor these systems to identify and address performance issues quickly. Today, thanks to new data analytics techniques, PV stakeholders can perform insightful remote analyses using limited but easily-accessible data, instead of relying on information-rich but expensive, expert-dependent, and interruptive field surveys.

One valuable online strategy is the modeling of deployed PV systems using operational data, as it allows us to understand deviations from the expected behavior and assess revamping opportunities. In this context, a simple yet effective tool that is often used to represent solar cells as well as PV modules is the single-diode model (SDM).[1] Even though the values of its parameters are normally not provided by panel manufacturers, several methods exist to estimate them from easily-accessible current-voltage (IV) properties.

On the one hand, analytical techniques[2–8] can fit IV data using mathematical techniques and physical models that describe the behavior of a solar cell. They are generally simple and easy to implement, but rely on specific assumptions that may not always hold. On the other hand, numerical techniques[9–11] use iterative methods like the Newton-Raphson and Lambert W-function methods to extract model parameters from more complex IV information instead of handpicked data points. They can handle more realistic scenarios that cannot be solved analytically, but they require more computational power and suffer from convergence issues due to their sensitivity to initial conditions.

Alternatively, (meta)heuristic algorithms[12–22] – such as the particle swarm, flower pollination, and teaching-learning-based optimization algorithms – offer several advantages over analytical and iterative techniques. They are significantly less sensitive to the initialization process, more robust to noise, and able to handle larger quantities of multi-dimensional IV data (e.g. with temperature dependence). Their speed, flexibility, and scalability have made them a popular choice among researchers for extracting circuit model parameters from IV curves. However, despite being a valuable tool for evaluating PV performance, IV curves are rarely measured in actual PV installations, which makes it difficult to implement these studies in real life.

Instead, we recently demonstrated how such techniques can be used to extract the SDM parameters and infer IV curves from typical power production data[23] (i.e. time-series that list the operating current and voltage of a PV system at varying irradiance and temperature) measured by commercial solar inverters and data loggers. Specifically, we used the teaching-learning-based optimization (TLBO) algorithm to find an optimal combination of SDM parameters that can model a monocrystalline silicon (mono c-Si) module using a representative sample of its production data. Similarly, other researchers have proposed different optimization algorithms[24,25] to achieve this using data from the National Renewable Energy Laboratory (NREL) and National Institute of Standards and Technology (NIST).

Albeit promising, these initial findings have several common shortcomings. First, they all rely on classical optimization techniques, which are limited to finding only one "best" fit for the problem at hand. However, based on the desired level of accuracy, complex models like the SDM can admit multiple solutions, especially when fitted to maximum power point (MPP) data only. As a result, for the same domain space, these algorithms can yield different and often stochastic results depending on their starting point, convergence properties, and tolerance to noise. Moreover, these studies have so far only covered crystalline silicon, which is the predominant but not exclusive PV module technology. Last but not least, reliable indoor measurements are still needed to verify these results, which have been assessed using less dependable outdoor measurements.

Here, we reframe the task of extracting circuit model parameters from production data as a Bayesian optimization problem. Using this probabilistic approach, we can explore the parameter search space more intelligently by considering a range of possible outcomes. We can then update these beliefs as more data become available, which enables us to gradually improve the accuracy of our predictions and reduce their uncertainty over time. We also extend our study on monocrystalline silicon to four other silicon and thin-film module technologies, test our new method using both onsite and satellite-based weather data, and further validate our results using laboratory measurements.

Methodology

Data

We rely on the data collected at the SIRTA (Site Instrumental de Recherche par Télédétection Atmosphérique) observatory's[26] PV test bench[27] located at Ecole Polytechnique in Palaiseau, France, and illustrated in Figure 1. The test bench was installed in 2014 and hosts five free-standing

South-oriented commercial solar panels of different technologies. The nameplate IV properties of these panels at Standard Test Conditions (STC) of 1000 W/m^2 and 25°C are reported in Table 1.

Table 1: Current and voltage properties of the installed modules at Standard Testing Conditions (STC). From left to right: output power (P_{MPP}), voltage (V_{MPP}), and current (I_{MPP}) at the maximum power point; open-circuit voltage (V_{oc}); and short-circuit current (I_{sc}). From top to bottom: tandem amorphous/microcrystalline silicon (a-Si/μ-Si), monocrystalline silicon (c-Si), Copper Indium Selenide (CIS), mono c-Si Heterojunction with amorphous silicon Intrinsic Thin layer (HIT), and Cadmium Telluride (CdTe).

	P_{MPP} [W]	V_{MPP} [V]	I_{MPP} [A]	V_{oc} [V]	I_{sc} [A]
a-Si/μ-Si module	128 (+10%/-5%)	45.40	2.82	59.8	3.45
c-Si module	250 (±3%)	30.52	8.21	37.67	8.64
CIS module	150 (+10%/-5%)	81.5	1.85	108.0	2.20
HIT	240 (+10%/-5%)	43.7	5.51	52.4	5.85
CdTe	82.5 (±10%)	48.30	1.71	60.80	1.94

We consider the one-year period (from 01/04/2021 to 31/03/2022) preceding the flash test measurements that were performed for all the modules in early April 2022. As before,[23] we limit ourselves to the MPP data derived from the IV curves measured by the Chroma electronic loads as well as the irradiance and module backsheet temperature data measured by a Class A CMP22 pyranometer (Kipp & Zonen) and four-wired class A platinum sensors (PT100), respectively. A full description of the measured data and installed equipment is available on the open-access GitLab repository of the test bench (see Data & Code Availability). More information on data pre-processing can also be found in the Supplemental Information.

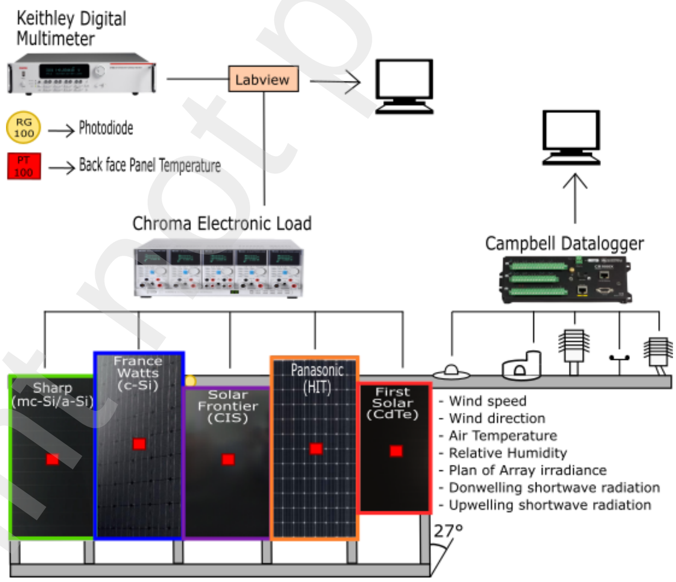

Figure 1: Schematic of the SIRTA PV Test Bench

Single Diode Circuit Model

We also keep the single-diode model (SDM) as our choice of equivalent circuit model. Essentially, the SDM is a simple physics-based model[1] that is commonly used to describe the behavior of a solar cell. Although solar cells are the primary power generation unit, PV modeling is usually done at the module level to reduce computational complexity and because solar panel specifications are more accessible. Assuming all the N_s series-connected cells of a PV module to be identical and under uniform and equal irradiance and temperature (i.e. they generate the same current and voltage), the governing equation[28] of the SDM can be written as:

$$I_M = I_L - I_o\left[exp\left(\frac{V_M + I_M N_s R_s}{nN_s V_T}\right) - 1\right] - \frac{V_M + I_M N_s R_s}{N_s R_{sh}} \tag{1}$$

Given a set of operating conditions (i.e. solar irradiance and cell temperature), the SDM thus depends on five parameters that relate the module's output current I_M to its operating voltage V_M: the light-induced current (I_L), saturation current (I_o), series resistance (R_s), shunt resistance (R_{sh}), and ideality factor (n). $V_T = \frac{k_B T_c}{q}$ is the thermal voltage of the cell, where k_B is Boltzmann's constant, T_c the cell temperature, and q the elementary charge.

The reference values of the SDM parameters are computed at STC and adjusted to the working conditions according to the De Soto model[3] we implemented[23] using the *pvlib* Python package.[29] The module's IV curve can then be computed to find its IV properties, namely the DC output current (I_{MPP}), voltage (V_{MPP}), and power (P_{MPP}) at the MPP, as well as the open-circuit voltage (V_{oc}) and short-circuit current (I_{sc}). We can thus assess the ability of a given parameter combination to model a chosen PV module (or array) by comparing its simulated IV properties with the measured data.

Bayesian Inference Approach

Instead of using a classical optimization algorithm to determine the single "best" parameter values for the SDM, we opt for the Bayesian point of view. This allows us to get probability distributions over every parameter, find a set of parameter combinations that can model the PV system at hand, and analyze how these parameters evolve over time. This approach is based on Bayes' theorem, which lets us update our prior beliefs about the parameters $P(\theta)$ as more data d becomes available.

$$P(\theta \mid d) = \frac{P(\theta)P(d|\theta)}{P(d)} \tag{2}$$

We first define the parameter search space based on typical datasheet specifications and the module's energy production during the analysis period and assume all the possible parameter combinations (hypotheses) to be initially equiprobable. We then calculate the likelihood $P(d|\theta_i)$ of each parameter combination by comparing its corresponding MPP current and voltage with the module's actual output current and voltage at the same measured operating conditions. The normalization constant $P(d)$ being the sum of $P(\theta)P(d|\theta)$ over all the hypotheses, we can compute the posterior $P(\theta \mid d)$ using Bayes' theorem. The full procedure is detailed in the Supplemental Information and the codes are publicly available on GitLab (see Data & Code Availability).

If we repeat this process on a regular basis while transferring the knowledge gained over time (i.e. by setting the posterior distribution obtained on the first day as the prior distribution of the following one and so on), we can progressively narrow down the 90% confidence interval for our predictions and monitor their evolution over time.

Results

Estimation of Circuit Model Parameters

We first consider the mono c-Si panel. Starting with the first day of the analysis period, we apply our Bayesian inference process on a weekly basis until the end of this one-year time frame (i.e. over a total of 52 sunny and cloudy days that are equally spaced).

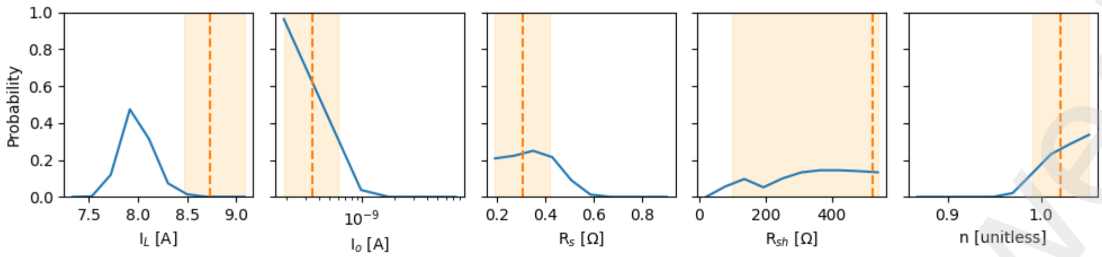

Figure 2: Probability distribution of the model parameters at the end of the analysis period.
The dashed line represents the value estimated from the datasheet properties. The shaded area accounts for the datasheet power tolerance.
From left to right: light-induced current (I_L), diode reverse saturation current (I_o), series resistance (R_s), shunt resistance (R_{sh}), and ideality factor (n).

Figure 2 shows the module SDM parameters' posterior probability distribution on the last day of the analysis period, which in this case also corresponds to the state of the PV module after roughly eight years of operation. The orange dashed line represents the reference STC value estimated from the module datasheet properties using the California Energy Commission (CEC) estimation method,[5] and the shaded area accounts for the manufacturer power tolerance (±3% in this case). In other terms, this area shows how much each parameter would need to change to increase (or decrease) the module's rated power by 3%, assuming the remaining parameters are kept constant. Note that this module's output power becomes insensitive to changes in the shunt resistance once the latter's value exceeds a few hundred ohms.

In the posterior distribution of I_L, the shift of the peak from the initial datasheet estimate indicates that the module's degradation is most likely induced by a drop in its light-induced and thus short-circuit current. This observation is consistent with field experience, which suggests that long-term current loss frequently impacts crystalline silicon modules.[30,31] Moreover, it appears that the dark saturation current I_o and series resistance R_s have not degraded over time, while the ideality factor n has remained more or less constant (~ 1). These observations reaffirm the ones we previously made[23] – even if the analysis periods are roughly one year apart – since this rather stable mono c-Si module has been well-maintained since it was installed. In contrast, even though the expected value of the shunt resistance R_{sh} is below its estimated reference value, there is considerable uncertainty around this parameter. Indeed, the impact of R_{sh} on the module's electrical output is imperceptible once it crosses a certain threshold, as illustrated by the shape of its posterior distribution.

Though the resulting drop in performance is in this instance well within the manufacturer's performance warranty, these plots can provide us valuable insights into the module's degradation modes. For example, a lower-than-expected value of I_{sc} can generally be an indication of uniform soiling or delamination, while shunt resistance degradation may be a symptom of Potential Induced Degradation (PID).[32] One of the main advantages of switching to the Bayesian framework is that it enables us to identify a set of possible solutions by exploring the full parameter search space. In comparison, traditional heuristic algorithms such as the ones used in prior work[12–25] are often designed to find the single best result using a fixed set of rules. These algorithms generally work well for deterministic problems where the outcome of a given input is always the same, which is not the case for extracting the parameters of a complex circuit model using only MPP data.

Estimation of Current-Voltage Properties

Since we know the probability of each parameter combination included in our search space, we can monitor the evolution of the 90% confidence interval for their associated module IV properties throughout the analysis period. This way, we can better understand how changes in the SDM parameters translate to changes in the module behavior (and vice versa) in case of a steady or sudden degradation.

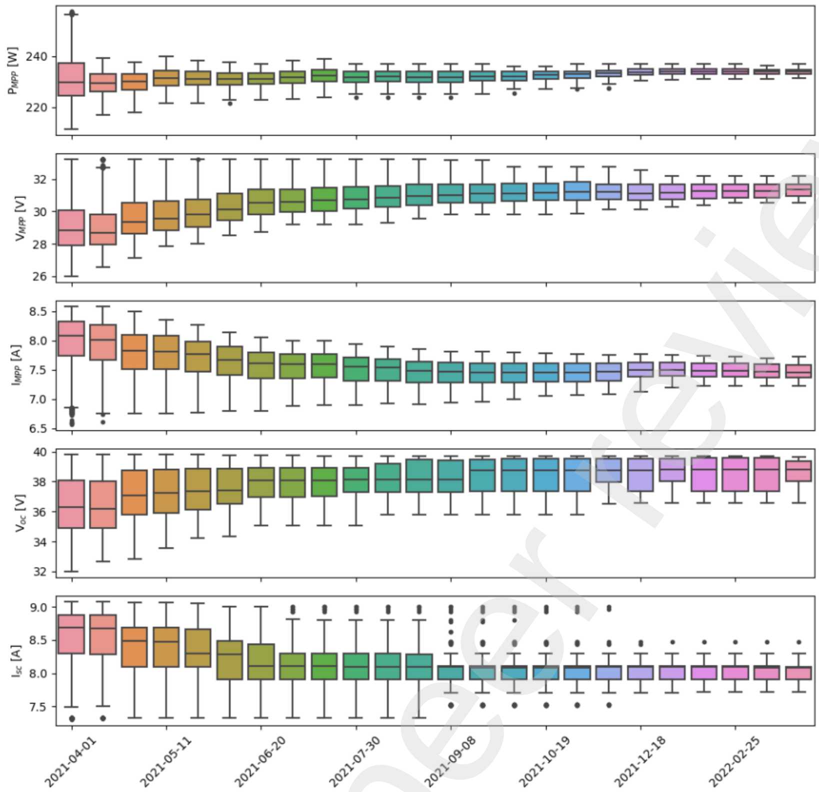

Figure 3: Evolution of the 90% confidence interval for the module's current-voltage properties during the analysis period. Black dots represent outliers. The colors of the boxes and number of days shown are set for visualization purposes only. From top to bottom: output power (P_{MPP}), voltage (V_{MPP}), and current (I_{MPP}) at the maximum power point; open-circuit voltage (V_{oc}); and short-circuit current (I_{sc}).

Naturally, the size of the confidence interval starts large and then progressively shrinks as more production data becomes available. After analyzing several days of production data, the 90% probability mass becomes confined to less than 5% of the search space. Towards the end of the analysis period, it sits in less than 1% of this space. Interestingly, the voltage converges to higher values, and the opposite is true for the current. This reaffirms our previous observation that this module's power degradation is the result of a drop in the current rather than a voltage degradation. Moreover, there is more confidence around the values at the MPP than at the open-circuit voltage V_{oc} and short-circuit current I_{sc}. This comes as no surprise since we can pair the two latter points with different R_s and R_{sh} values to get IV curves that have different shapes but pass through the same MPP. Indeed, it is difficult to determine unique values for the SDM parameters and IV properties from MPP data alone, therefore making it more fitting to have confidence intervals to reflect the estimation uncertainty.

Besides quantifying uncertainty, another major advantage of using Bayesian inference is that it allows us to study short as well as long time frames and progressively update the output in a speedy yet consistent manner – without having to exclude cloudy days. As we previously noted,[23] it would be difficult to obtain similar results using classical optimization algorithms since they are designed to find a single best and independent fit for each batch of data. This becomes even more evident once we look at Li et al.'s results,[24] which exhibit seasonal variability and stochastic patterns that are atypical of rather stable c-Si panels and can thus yield different and perhaps misleading conclusions depending on the length and size of the analysis window. Bayesian inference overcomes this issue because it continuously updates its (prior) beliefs about the system using the best consecutive data fits (i.e. the computed likelihoods).

As for the module SDM parameters, we can inspect the distribution of the predicted module IV properties at the end of the analysis period. This can help us better understand the uncertainty surrounding them as well as their positioning relative to their rated values before deployment.

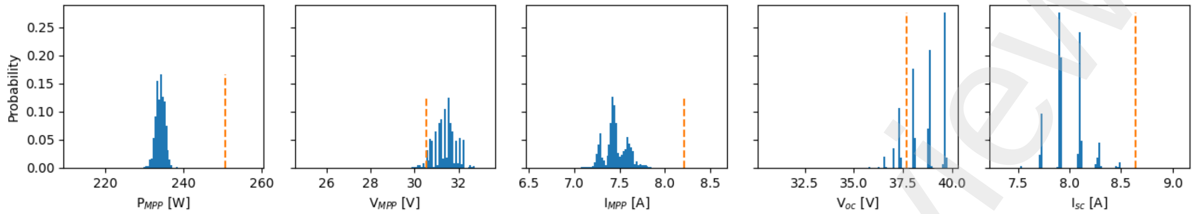

Figure 4: Distribution of the predicted module current-voltage properties.
The dashed line represents the rated value from the panel datasheet.
From left to right: output power (P_{MPP}), voltage (V_{MPP}), and current (I_{MPP}) at the maximum power point; open-circuit voltage (V_{oc}); and short-circuit current (I_{sc}).

At the end of the analysis period, the posterior distributions of the predicted module IV properties are concentrated over a small portion of the initial SDM parameter search space. The 90% confidence intervals for the module's output power P_{MPP}, voltage V_{MPP}, and current I_{MPP} at STC are within ±3W, ±0.83V, and ±0.25A from their respective expected value, which corresponds to impressive precision levels between ±1% and ±3.5%. However, the precision is expectedly lower (relative error ranging between -5.2% and 6.2%) for the open-circuit voltage and short-circuit current. The presence of multiple peaks again shows how different values of V_{oc} and I_{sc} exist to get the same MPP predictions. These findings reaffirm the above observations and show that in this case the module degradation is linked to a drop in current, since the voltage remains stable over time.

Flash Test Validation

Having performed flash tests shortly after the analysis finish date, we can evaluate the accuracy of the predicted module IV properties (at STC) by comparing their expected values (weighted averages) with the results of these indoor measurements.

Table 2: Comparison between the predicted and lab-measured current-voltage properties.
From top to bottom: output power (P_{MPP}), voltage (V_{MPP}), and current (I_{MPP}) at the maximum power point; open-circuit voltage (V_{oc}); and short-circuit current (I_{sc}).

	Datasheet	**Measured**	**Predicted**	**Relative Error (Predicted vs. Measured)**
P_{MPP} [W]	250 (±3%)	243.94	233.95	-4.1%
V_{MPP} [V]	30.52	30.47	31.38	3.0%
I_{MPP} [A]	8.21	8.01	7.46	-6.9%
V_{oc} [V]	37.67	37.56	38.58	2.7%
I_{sc} [A]	8.64	8.49	7.98	-6.0%

The tabulated results show the degradation trends to be predicted correctly. However, it appears that the voltage is once again slightly overestimated while the current is more noticeably underestimated, especially at the MPP. Nevertheless, the error is within the ±7.2% margin of uncertainty for onsite power measurements.[27] The modeling inaccuracies of the SDM (see Supplemental Information), particularly at low irradiance, are another factor to keep in mind when evaluating the quality of these predictions. Last but not least, we suspect that the modules' short-circuit and MPP currents (and thus power outputs) are actually under-measured by the electronic loads because their recorded values do not match the values expected using the flash test results, especially at higher irradiance and temperature. More tests are needed to quantify this error and identify whether the problem comes from the equipment measuring the IV curves and/or the pyranometer overestimating the irradiance, but this issue highlights the importance of having reliable input data, since the prediction quality of any machine learning model strongly depends on the quantity as well as the quality of the data it feeds on.

Results for Other Panel Technologies

Since we have the flash test results for all the modules installed at the PV test bench, we can also check how accurately and precisely this approach can estimate the IV properties of the four remaining panel technologies, namely: tandem amorphous/microcrystalline silicon (a-Si/μ-Si), Copper Indium Selenide (CIS), mono c-Si Heterojunction with amorphous silicon Intrinsic Thin layer (HIT), and Cadmium Telluride (CdTe).

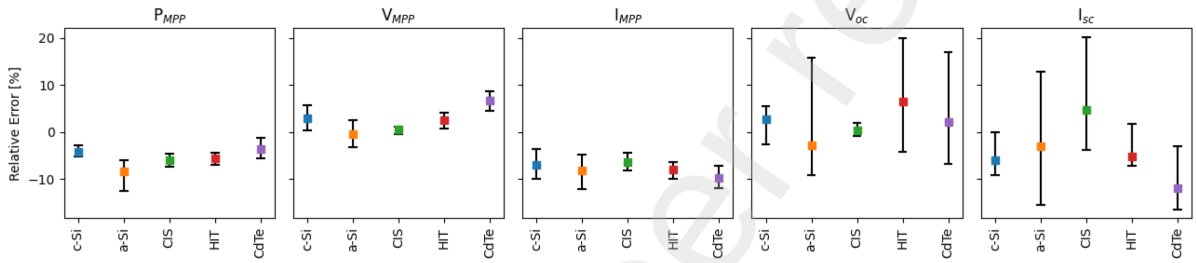

Figure 5: Relative error between the predicted and measured current-voltage properties of all the installed modules. The colored squares represent the error for the expected value and the error bars correspond to the minimum and maximum errors of the 90% confidence interval.
From left to right on the title axis: output power (P_{MPP}), voltage (V_{MPP}), and current (I_{MPP}) at the maximum power point; open-circuit voltage (V_{oc}); and short-circuit current (I_{sc}).
From left to right on the x-axis: monocrystalline silicon (c-Si), tandem amorphous/microcrystalline silicon (a-Si/μ-Si), Copper Indium Selenide (CIS), mono c-Si Heterojunction with amorphous silicon Intrinsic Thin layer (HIT), and Cadmium Telluride (CdTe).

Overall, the errors for the expected values (colored squares) follow similar trend lines across the different PV technologies and are within the ±7.2% measurement uncertainty of the power production data, which would increase if we also consider the meteorological data uncertainty. All the predicted MPP properties have a narrow 90% confidence interval (black error bars), but the output power and current at MPP are consistently underestimated, most likely due to their apparent underestimation in the production data. In contrast, the output voltage at MPP is generally overestimated, which explains why P_{MPP} is slightly less underestimated than I_{MPP} is.

However, it is more difficult to determine a trend line for V_{oc} and I_{sc}, since the error sign and prediction uncertainty vary from one PV technology to the other. The distributions of these predicted IV properties are also considerably but expectedly wider, especially for the thin-film modules. Even for the comparable c-Si and HIT technologies, we notice that the latter has far higher open-circuit voltage uncertainty. This is not only due to the presence of modeling errors (see Supplemental Information) but also a difference in the manufacturer power tolerance (c.f. Table 1) and sensitivity to the different SDM parameters.

While the flash tests were performed twice and have a low measurement uncertainty of ±0.24%, their setup is still mostly adapted for more prevalent silicon modules. The accuracy of the lab results obtained for the thin-film panels may thus be affected by a mismatch with the solar simulator's spectrum and/or mono c-Si reference cell used.

Despite these shortcomings, the average IV-property prediction error is around 5% across all five technologies. Taking the power tolerance and measurement uncertainty into account, we can say that the SDM offers a good compromise between simplicity and accuracy when extracting module IV properties from power production data.

Sensitivity Analysis

Knowing that weather data is seldom measured in the field, we report the results we get when we replace the site measurements for irradiance and temperature with online satellite estimates using the same databases[33–36] and thermal model[37] described in our previous work.[23] This adds another layer of uncertainty to the input data but helps keep this research project grounded in real life.

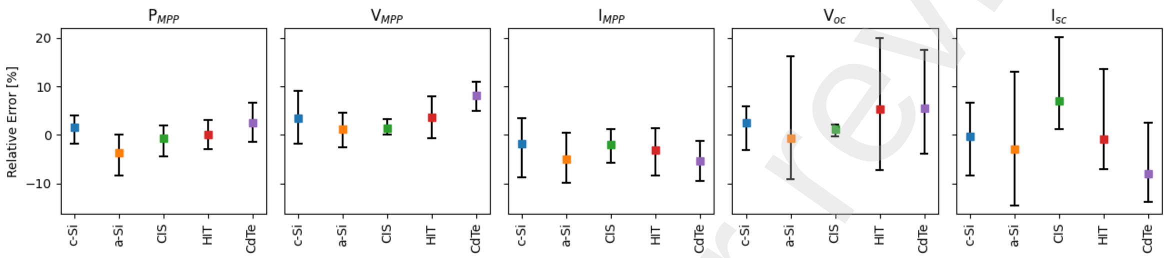

Figure 6: Relative error between the predicted and measured current-voltage properties of all the installed modules (using online weather data).

At first sight, the error profiles closely resemble those obtained when onsite weather data is used. The main difference is that the error is lower here because the Global Plane-of-Array (GPOA) irradiance estimated using satellite data is lower than the one measured by the pyranometer (especially at high irradiance),[23] which results in higher values for the predicted output power and current.

Considering that the choice of databases and physical models is not optimized for this PV test bench, these promising results attest to the robustness of this practical approach, which can be scaled to real PV systems of any size and technology.

Discussion

The ability to effectively extract circuit model parameters and IV properties from readily-available power production and weather data on a daily basis is a powerful remote PV monitoring and modeling tool with a wide range of benefits for PV stakeholders.

First, this type of study – which can take anywhere between a few minutes to a few hours on the average computer (based on the size of the system and length of the analysis period) – allows for continuous monitoring of insightful PV characteristics. This makes it possible to identify and address problems early on, thus minimizing downtime and maximizing energy production. The modules studied here are operating optimally, but for larger and more complex PV systems, one can inspect the evolution of the SDM parameters and/or IV properties over time to trace back the onset of degradation, identify its potential root causes, and assess its severity. As a result, system operators can make more informed decisions remotely, hence reducing the need for site visits and manual inspections.

In addition, the extracted parameters can be coupled with AC modeling to simulate the performance of grid-connected PV systems and forecast energy production. By estimating future energy losses, system owners can see if it makes financial sense to replace their panels and/or inverters for repowering applications. A key feature of using a probabilistic Bayesian approach is the ability to get a range of possible outcomes along with their associated probabilities. This is particularly useful for making business decisions and managing risk because it helps assess the variance around the expected return on investment.

Whether they are related to the modeling error or breadth of the search space, this approach's aforementioned limitations can be addressed to improve the quality of the results. The SDM is widely used to describe the behavior of silicon solar cells, which are the dominant PV technology, but one may consider alternative models to get improved accuracy and physical interpretation, particularly for thin-film technologies. In our proposed Bayesian inference method, we evaluate the same set of parameter combinations for every day of the analysis period, and the initial reference parameter values extracted from the panel datasheet strongly influence the combinations that constitute our search space. One way to enhance the final results is to refine the initial parameter search and rerun this analysis over and over until a pre-defined convergence criterion is met. Or, instead of having a limited number of hypotheses (i.e. a discrete parameter space), we can use an acquisition function to explore a continuous parameter space. Acquisition functions can balance global exploration (e.g. exploring solutions with larger shunt resistance values for the c-Si panel) and local exploitation (e.g. refining the value of the light-induced current) of the search space by relying on a probabilistic model (typically a Gaussian process) that approximates the objective function (the SDM here). While these functions help reduce the number of evaluated hypotheses by exploring the search space more efficiently, they come with a heavy computational cost when compared with *pvlib*'s implementation of the SDM. Furthermore, this adaptive parameter search space strategy typical of Bayesian optimization would take away the ability to carry posterior distributions over time. In any case, there is only little room to improve the accuracy and precision of the V_{oc} and I_{sc} predictions since we are limited to MPP data. Ultimately, the choice of model and search space exploration strategy depends on the desired level of accuracy and computational efficiency for the specific application.

Still, further development is needed to standardize procedures such as data cleaning, translate changes in the SDM parameters to detectable module defects and performance problems, and test this approach on larger and more complex PV systems.

Conclusion

In this work, we present several enhancements to the method we previously developed[23] to extract the single-diode model parameters of a PV system from commoditized production and weather data. We move from the frequentist approach to the Bayesian point of view to find a range of possible solutions along with their associated probabilities. We automate the definition of the search space based on the specifications and performance of the system and incorporate a transfer learning strategy to improve the precision and interpretability of the predictions over short and long time frames. Using flash test measurements, we demonstrate an average prediction error of 5% for the current-voltage properties of two silicon and three thin-film modules. Our approach can be scaled to PV systems of any size and technology and works with field weather data as well as satellite estimates, thus offering exciting opportunities for remote performance and degradation analysis of real solar energy systems.

Experimental Procedures

Data & Code Availability

The data and codes that support the findings of this study are publicly available on GitLab https://gitlab.in2p3.fr/energy4climate/public/sirta-pv1-data.

Flash Tests

The flash tests were performed at TotalEnergies laboratory facilities using a CetisPV-XF2-M Xenon flasher. Flash duration is 58ms, and average results are reported at 1000 W/m² and 25°C for four IV curve measurements (one forward and one reverse sweep per curve). The ±0.24% measurement uncertainty is based on a statistical process control (SPC) obtained by continuously measuring the MPP power output of a certified mono c-Si reference module.

Acknowledgments

The authors acknowledge Ph.D. candidate Moira Torres for helping access the data and flash test results used in this study. This project was first launched at Feedgy, a PV performance analysis and repowering company, and is being further developed at Ecole Polytechnique as part of an ongoing collaboration. This research is also produced within the framework of the Energy4Climate Interdisciplinary Center (E4C) of IP Paris and Ecole des Ponts ParisTech. It is also supported by the 3rd Programme d'Investissements d'Avenir [ANR-18-EUR-0006-02] and by the Foundation of Ecole Polytechnique (Chaire "Défis Technologiques pour une Energie Responsable" financed by Total Energies).

Author Contributions

Conceptualization, J.C. and M.P.; Methodology, J.C.; Software, J.C.; Formal Analysis, J.C.; Data Curation, J.C. and J.B.; Writing – Original Draft, J.C.; Writing – Review & Editing, J-P.C., M.P., J.B., J-B.P., and Y.B.; Visualization, J.C. and J-P.C.; Supervision, M.P., J.B., J-B.P., and Y.B.; Funding Acquisition, M.P., J.B., and Y.B.

Declaration of Interests

The authors declare no competing interests.

References

[1] Gray JL. The Physics of the Solar Cell. In: Luque A, Hegedus S, editors. Handb. Photovolt. Sci. Eng., Chichester, UK: John Wiley & Sons, Ltd; 2005, p. 61–112. https://doi.org/10.1002/0470014008.ch3.

[2] Batzelis EI, Papathanassiou SA. A Method for the Analytical Extraction of the Single-Diode PV Model Parameters. IEEE Trans Sustain Energy 2016;7:504–12. https://doi.org/10.1109/TSTE.2015.2503435.

[3] De Soto W, Klein SA, Beckman WA. Improvement and validation of a model for photovoltaic array performance. Sol Energy 2006;80:78–88. https://doi.org/10.1016/j.solener.2005.06.010.

[4] Piazza MCD, Luna M, Petrone G, Spagnuolo G. Translation of the Single-Diode PV Model Parameters Identified by Using Explicit Formulas. IEEE J Photovolt 2017;7:1009–16. https://doi.org/10.1109/JPHOTOV.2017.2699321.

[5] Dobos AP. An Improved Coefficient Calculator for the California Energy Commission 6 Parameter Photovoltaic Module Model. J Sol Energy Eng 2012;134:021011. https://doi.org/10.1115/1.4005759.

[6] Lo Brano V, Ciulla G. An efficient analytical approach for obtaining a five parameters model of photovoltaic modules using only reference data. Appl Energy 2013;111:894–903. https://doi.org/10.1016/j.apenergy.2013.06.046.

[7] Phang JCH, Chan DSH, Phillips JR. Accurate analytical method for the extraction of solar cell model parameters. Electron Lett 1984;20:406. https://doi.org/10.1049/el:19840281.

[8] de Blas MA, Torres JL, Prieto E, Garćia A. Selecting a suitable model for characterizing photovoltaic devices. Renew Energy 2002;25:371–80. https://doi.org/10.1016/S0960-1481(01)00056-8.

[9] Ishibashi K, Kimura Y, Niwano M. An extensively valid and stable method for derivation of all parameters of a solar cell from a single current-voltage characteristic. J Appl Phys 2008;103:094507. https://doi.org/10.1063/1.2895396.

[10] Ortizconde A, Garciasanchez F, Muci J. New method to extract the model parameters of solar cells from the explicit analytic solutions of their illuminated characteristics. Sol Energy Mater Sol Cells 2006;90:352–61. https://doi.org/10.1016/j.solmat.2005.04.023.

[11] Li Y, Huang W, Huang H, Hewitt C, Chen Y, Fang G, et al. Evaluation of methods to extract parameters from current–voltage characteristics of solar cells. Sol Energy 2013;90:51–7. https://doi.org/10.1016/j.solener.2012.12.005.

[12] Ye M, Wang X, Xu Y. Parameter extraction of solar cells using particle swarm optimization. J Appl Phys 2009;105:094502. https://doi.org/10.1063/1.3122082.

[13] Sheng H, Li C, Wang H, Yan Z, Xiong Y, Cao Z, et al. Parameters Extraction of Photovoltaic Models Using an Improved Moth-Flame Optimization. Energies 2019;12:3527. https://doi.org/10.3390/en12183527.

[14] Xiong G, Zhang J, Shi D, Yuan X. Application of Supply-Demand-Based Optimization for Parameter Extraction of Solar Photovoltaic Models. Complexity 2019;2019:1–22. https://doi.org/10.1155/2019/3923691.

[15] Xiong G, Li L, Mohamed AW, Yuan X, Zhang J. A new method for parameter extraction of solar photovoltaic models using gaining–sharing knowledge based algorithm. Energy Rep 2021;7:3286–301. https://doi.org/10.1016/j.egyr.2021.05.030.

[16] Rajasekar N, Krishna Kumar N, Venugopalan R. Bacterial Foraging Algorithm based solar PV parameter estimation. Sol Energy 2013;97:255–65. https://doi.org/10.1016/j.solener.2013.08.019.

[17] Gude S, Jana KC. Parameter extraction of photovoltaic cell using an improved cuckoo search optimization. Sol Energy 2020;204:280–93. https://doi.org/10.1016/j.solener.2020.04.036.

[18] Alam DF, Yousri DA, Eteiba MB. Flower Pollination Algorithm based solar PV parameter estimation. Energy Convers Manag 2015;101:410–22. https://doi.org/10.1016/j.enconman.2015.05.074.

[19] Patel SJ, Panchal AK, Kheraj V. Extraction of solar cell parameters from a single current–voltage characteristic using teaching learning based optimization algorithm. Appl Energy 2014;119:384–93. https://doi.org/10.1016/j.apenergy.2014.01.027.
[20] Niu Q, Zhang L, Li K. A biogeography-based optimization algorithm with mutation strategies for model parameter estimation of solar and fuel cells. Energy Convers Manag 2014;86:1173–85. https://doi.org/10.1016/j.enconman.2014.06.026.
[21] Xu S, Wang Y. Parameter estimation of photovoltaic modules using a hybrid flower pollination algorithm. Energy Convers Manag 2017;144:53–68. https://doi.org/10.1016/j.enconman.2017.04.042.
[22] Ram JP, Babu TS, Dragicevic T, Rajasekar N. A new hybrid bee pollinator flower pollination algorithm for solar PV parameter estimation. Energy Convers Manag 2017;135:463–76. https://doi.org/10.1016/j.enconman.2016.12.082.
[23] Chakar J, Pavlov M, Bonnassieux Y, Badosa J. Determining solar cell parameters and degradation rates from power production data. Energy Convers Manag X 2022;15:100270. https://doi.org/10.1016/j.ecmx.2022.100270.
[24] Li B, Karin T, Meyers BE, Chen X, Jordan DC, Hansen CW, et al. Determining circuit model parameters from operation data for PV system degradation analysis: PVPRO. Sol Energy 2023;254:168–81. https://doi.org/10.1016/j.solener.2023.03.011.
[25] Sun X, Chavali RVK, Alam MA. Real-time monitoring and diagnosis of photovoltaic system degradation only using maximum power point—the Suns-Vmp method. Prog Photovolt Res Appl 2019;27:55–66. https://doi.org/10.1002/pip.3043.
[26] Haeffelin M, Barthès L, Bock O, Boitel C, Bony S, Bouniol D, et al. SIRTA, a ground-based atmospheric observatory for cloud and aerosol research. Ann Geophys 2005;23:253–75. https://doi.org/10.5194/angeo-23-253-2005.
[27] Dubois AM, Badosa J, Bourdin V, Torres Aguilar MI, Bonnassieux Y. Estimation of the Uncertainty due to Each Step of Simulating the Photovoltaic Conversion under Real Operating Conditions. Int J Photoenergy 2021;2021:1–14. https://doi.org/10.1155/2021/4228658.
[28] Tian H, Mancilla-David F, Ellis K, Muljadi E, Jenkins P. A cell-to-module-to-array detailed model for photovoltaic panels. Sol Energy 2012;86:2695–706. https://doi.org/10.1016/j.solener.2012.06.004.
[29] F. Holmgren W, W. Hansen C, A. Mikofski M. pvlib python: a python package for modeling solar energy systems. J Open Source Softw 2018;3:884. https://doi.org/10.21105/joss.00884.
[30] Jordan DC, Silverman TJ, Wohlgemuth JH, Kurtz SR, VanSant KT. Photovoltaic failure and degradation modes. Prog Photovolt Res Appl 2017;25:318–26. https://doi.org/10.1002/pip.2866.
[31] Pascual J, Martinez-Moreno F, García M, Marcos J, Marroyo L, Lorenzo E. Long-term degradation rate of crystalline silicon PV modules at commercial PV plants: An 82-MWp assessment over 10 years. Prog Photovolt Res Appl 2021;29:1294–302. https://doi.org/10.1002/pip.3456.
[32] Itrac F. Reprinted with permission from SolarPro 2014.
[33] Lefèvre M, Oumbe A, Blanc P, Espinar B, Gschwind B, Qu Z, et al. McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmospheric Meas Tech 2013;6:2403–18. https://doi.org/10.5194/amt-6-2403-2013.
[34] Gschwind B, Wald L, Blanc P, Lefèvre M, Schroedter-Homscheidt M, Arola A. Improving the McClear model estimating the downwelling solar radiation at ground level in cloud-free conditions – McClear-v3. Meteorol Z 2019;28:147–63. https://doi.org/10.1127/metz/2019/0946.
[35] Qu Z, Oumbe A, Blanc P, Espinar B, Gesell G, Gschwind B, et al. Fast radiative transfer parameterisation for assessing the surface solar irradiance: The Heliosat-4 method. Meteorol Z 2017;26:33–57. https://doi.org/10.1127/metz/2016/0781.

[36] Gelaro R, McCarty W, Suárez MJ, Todling R, Molod A, Takacs L, et al. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). J Clim 2017;30:5419–54. https://doi.org/10.1175/JCLI-D-16-0758.1.
[37] Faiman D. Assessing the outdoor operating temperature of photovoltaic modules. Prog Photovolt Res Appl 2008;16:307–15. https://doi.org/10.1002/pip.813.