

Article

Classification of Photovoltaic Failures with Hidden Markov Modeling, an Unsupervised Statistical Approach

Michael W. Hopwood ^{1,2} , Lekha Patel ¹  and Thushara Gunda ^{1,*} 

¹ Sandia National Laboratories, Albuquerque, NM 87123, USA; michael.hopwood@knights.ucf.edu (M.W.H.); lpatel@sandia.gov (L.P.)

² Department of Statistics and Data Science, University of Central Florida, Orlando, FL 32816, USA

* Correspondence: tgunda@sandia.gov

Abstract: Failure detection methods are of significant interest for photovoltaic (PV) site operators to help reduce gaps between expected and observed energy generation. Current approaches for field-based fault detection, however, rely on multiple data inputs and can suffer from interpretability issues. In contrast, this work offers an unsupervised statistical approach that leverages hidden Markov models (HMM) to identify failures occurring at PV sites. Using performance index data from 104 sites across the United States, individual PV-HMM models are trained and evaluated for failure detection and transition probabilities. This analysis indicates that the trained PV-HMM models have the highest probability of remaining in their current state (87.1% to 93.5%), whereas the transition probability from normal to failure (6.5%) is lower than the transition from failure to normal (12.9%) states. A comparison of these patterns using both threshold levels and operations and maintenance (O&M) tickets indicate high precision rates of PV-HMMs (median = 82.4%) across all of the sites. Although additional work is needed to assess sensitivities, the PV-HMM methodology demonstrates significant potential for real-time failure detection as well as extensions into predictive maintenance capabilities for PV.



Citation: Hopwood, M.W.; Patel, L.; Gunda, T. Classification of Photovoltaic Failures with Hidden Markov Modeling, an Unsupervised Statistical Approach. *Energies* **2022**, *15*, 5104. <https://doi.org/10.3390/en15145104>

Academic Editor: Antonio Rosato

Received: 26 June 2022

Accepted: 11 July 2022

Published: 13 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: photovoltaics; failure detection; hidden Markov model; unsupervised statistical learning; classification; operations and maintenance

1. Introduction

Photovoltaic (PV) systems represent one of the fastest growing renewable energy sectors [1]. However, industry analysis shows that there has been a persistent gap between expected and observed energy generation at PV sites, which impacts the overall revenue generated at a PV site [2]. These inconsistencies arise from a combination of deviations, from overestimations in baseline values [3] to failures observed within the field. The latter can arise from diverse causal mechanisms within PV, including delamination within modules [4], communication system problems within inverters [5], and exposure to extreme weather events [6].

Given the notable impact that reduced production can have on overall PV site revenue [2], significant attention has been given to failure detection and diagnosis methods to improve the reliability of these systems and reduce system downtime. Diverse types of data have been used to characterize failures, including current–voltage traces [7], electroluminescence/infrared imagery data [8], and maximum power points [9]. Researchers have also evaluated operations and maintenance (O&M) tickets to understand seasonal patterns in failure frequencies to inform operators of common failure modes and support planning efforts [5].

Machine learning (ML) approaches, such as neural networks (NNs), have emerged as popular techniques for failure characterizations. For instance, artificial NNs have been trained with solar irradiance and PV temperature measurements to classify large deviations

observed in real time between measured and predicted power values [10]. Others have used convolutional NNs to predict the daily electrical power curve of a PV panel using the power curves of neighboring panels to support the detection of panel malfunctions [11]. Alternatively, distance metrics have been used as part of supervised cluster-based analyses to determine failures relative to deviations from normal clusters of events observed from the time series of non-degraded PV sensor data [12–14].

NN methods have demonstrated significant potential for failure analyses. However, the limited nature of labeled datasets used to train and validate these algorithms makes it unclear whether the trained models can generalize well across exogenic variables such as site-specific characteristics, micro-climates, and seasonality trends [15]. Furthermore, interpretability issues surrounding ML architectures render the estimation and uncertainty quantification of explainable parameters that regulate PV systems as difficult [16]. Finally, cluster-based approaches cannot quantify valuable information (e.g., the average time to failure) necessary for both accurate maintenance prediction and risk analyses of a variety of systems.

In contrast, unsupervised statistical methods for failure analyses are able to work with smaller unlabeled data and use a probabilistic approach for identifying patterns within datasets. For example, failure patterns have been characterized using curve fitting methods to generate parameters from the Weibull distribution (parametric approach) and to generate survival functions (non-parametric approach) [17]. Markov processes and stochastic processes with the special property that their future and past realizations (or *states*) are independent given their present state and have also been used to characterize the mean times to failure in PV systems [18]. Although survival rates and times to failure are quantifiable and well-characterized, these approaches focus more on characterization vs. detection.

In this paper, we extend these statistical implementations by introducing the photovoltaic-hidden Markov model (PV-HMM) to support failure detection needs. The popularity of HMMs arises from their ability to relate sequences of observations (or *emissions*) with an underlying Markov process, whose unobserved (i.e., hidden) states form the target of inference. HMMs are able to work with discrete time series (e.g., hourly measurements), which makes this method particularly well suited for characterizing and predicting failures in solar PV. Markov Chains (i.e., Markov processes in discrete time) with probabilistic functions in the form of HMMs have been extensively used in many scientific disciplines, including speech recognition [19], bioinformatics [20], and epidemiology [21]. HMM implementations have also entered energy domains, with researchers using this method to predict energy consumption in buildings [22], evaluate battery usage patterns [23], and even explore fault diagnosis with simulated data in wind energy systems [24]. In this analysis, we extend these implementations to characterize field failures at PV sites.

Specifically, we train site-specific PV-HMMs to distinguish and extrapolate failure states from performance signals in an unsupervised manner with relatively little data. The main benefit of this novel approach are that HMMs do not require labels to train the models and perform robustly even with coarse data [22]. Comparison analyses with threshold levels (TLs) and O&M tickets are used to validate the model outputs. In addition to advancing our understanding of transitions between normal and faulted states, this method can also be used to support predictive maintenance efforts within PV and other energy industries working with similar datasets. The remainder of this paper describes the HMM model and datasets in greater detail (Section 2), presents the results of the trained PV-HMM models (Section 3), and concludes with implications of the analysis and future work opportunities (Section 4).

2. Materials and Methods

This analysis was supported by Sandia National Laboratories' PV reliability, operations, and maintenance (PVRM) database, which contains production, environmental, and O&M ticket data for multiple sites across the United States [6]. More information

about the database contents, as well as the data processing and training of PV-HMMs, is described in the following sub-sections.

2.1. Data and Pre-Processing

The PVRM database contains co-located production and maintenance data for 186 sites across the United States [6]. Production datasets from each site generally capture continuous temporal measurements of observed energy production as well as onsite environmental measurements (e.g., irradiance and temperature) that are measured at regular intervals (e.g., every hour). Site O&M tickets, on the other hand, are discrete data points that are generated whenever an issue is observed that requires tracking or maintenance at a site. O&M tickets generally capture information about the date the issue was first observed, a general description of the issue, and details about how/when the issue was resolved.

Site-level data from PVRM require cleaning and filtering to ensure that the collected data meet quality standards. Following previous analyses [3], various filters were used to remove non-negative energy values and verify that the energy capacity constraints of each site were met. In addition, time periods with irradiance values greater than 400 Watts per square meter irradiance were retained to capture active production hours. Only sites that contained at least 25 data points post data quality checks were considered for further analyses. The final dataset used for PV-HMM analyses consisted of approximately 200,000 hourly data points across 104 sites in nine states across the United States (Figure 1). Generally, the number of data points per site ranged from 170 to 3300.

After data cleaning, data processing steps, involving the normalization of production data and tagging periods with active O&M tickets, were conducted. Specifically, the observed energy production (E) was normalized at each site with the associated expected energy values provided by the industry site operator (\hat{E}). The resulting metric is known as the performance index (PI), which is the primary variable of interest for PV-HMM analyses. Normalizing the energy by a baseline removes the variance caused by normal operation, such as periodic behavior resulting from sunrises and sunsets, and renders a signal which is mainly centered around one. Large perturbations from one, however, may imply a system anomaly and indicate a faulted system. Finally, to support model evaluation activities, PI values were tagged with a flag to indicate whether there was an active O&M ticket for a given hour at a site.

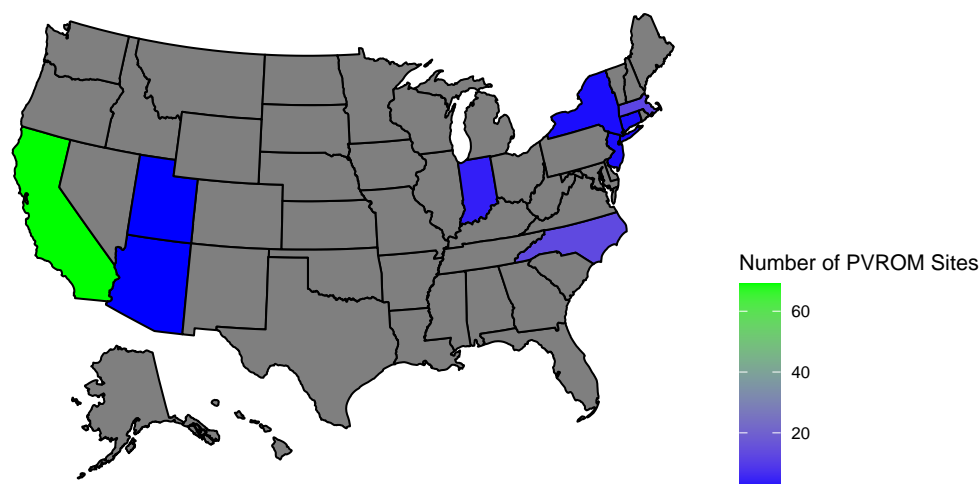


Figure 1. Photovoltaic sites within PVRM that contain both continuous production data and discrete maintenance records spanning nine U.S. states. Grey indicates states with no PVRM sites.

2.2. Photovoltaic Hidden Markov Model (PV-HMM)

Here, we describe the hidden Markov model and its application to PV in greater detail. Specifically, the PV-HMM relates the univariate signal underlying system performance

with what is observed of performance indices. A PV-HMM is trained per site using hourly PI data.

The PV-HMM is intended to learn the states of a 2-state discrete time Markov chain denoted as X_n over time $n = 1, 2, \dots$, where n denotes the hours that PI is analyzed. We assume that its states are binned according to normal (N) and faulted (F) signals, enabling $X_n \in \{N, F\}$. This chain is characterized first by its initial probability of being in state N, $v_N = P(X_1 = N)$ (or equivalently $v_F = 1 - v_N$). Second, due to the *Markov property* or the conditional independence of past and future states on the present state, i.e., $P(X_{n+1}|X_n, X_{n-1}) = P(X_{n+1}|X_n)$, its transition matrix is P_X , meaning that the one-step transition probabilities $P(X_{n+1}|X_n)$ between all states are also required in its characterization. Here,

$$P_X = \begin{pmatrix} P(N|N) & P(F|N) \\ P(N|F) & P(F|F) \end{pmatrix}, \quad (1)$$

whose rows sum to one, providing an alternative representation of the state space diagram for X_n , as shown in Figure 2. The measured performance indices PI_n at each discrete time point $n = 1, 2, \dots$ can be related to the underlying states $X_n \in \{N, F\}$, as depicted in Figure 3, via fitted emission densities of the form:

$$PI_n|X_n = f(PI_n|X_n, \theta). \quad (2)$$

Here, f denotes a general probability density function depending on X_n and other parameters θ and is commonly taken to be the Gaussian distribution for continuous responses or the multinomial distribution for categorical (discrete) responses. However, faulted states could occur at both high and low PI values so a bi-modal distribution is possible that would violate Gaussian assumptions. To deal with this, we utilize

$$f(PI_n|X_n, \theta) = \pi N(PI_n; \mu_{X_n,1}, \sigma_{X_n,1}^2) + (1 - \pi) N(PI_n; \mu_{X_n,2}, \sigma_{X_n,2}^2), \quad (3)$$

a 2-component Gaussian mixture model (GMM) that is able to more effectively capture nuances in the performance indices. Here, $\theta = (\pi, \mu_{X_n,1}, \sigma_{X_n,1}, \mu_{X_n,2}, \sigma_{X_n,2})$, including the mixing proportion of the densities π , and unknown mean (μ) and standard deviation (σ) parameters for each mixture that depend on the state of the underlying Markov chain. Estimating such parameters from the HMM enables predictions to be made between future performance indices and the state of the PV system. This can be achieved by using an iterative quasi-maximization approach called the Baum–Welch algorithm, a type of expectation–maximization (EM) algorithm [25]. At each iteration, this algorithm first performs an expectation (E) step by evaluating the derived expected log-likelihood function at current parameter estimates. Second, a maximization step (M) is pursued to iteratively update parameter estimates until their change becomes negligible and a quasi-maximum is found. Using these parameters, the Viterbi algorithm, a dynamic programming regime, can then be applied to obtain the most likely sequence of labels (N,F) for the hidden states of the PV system. A more thorough introduction to the Viterbi algorithm and its formulation for general inference in HMMs can be found at [26,27].

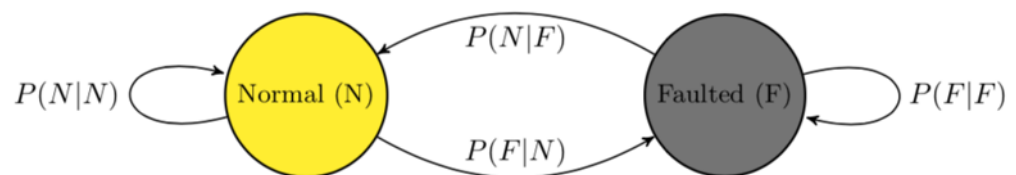


Figure 2. Transition probabilities (P) describe the probability of system state transitions between two states. Possible states are normal (N) and faulted (F).

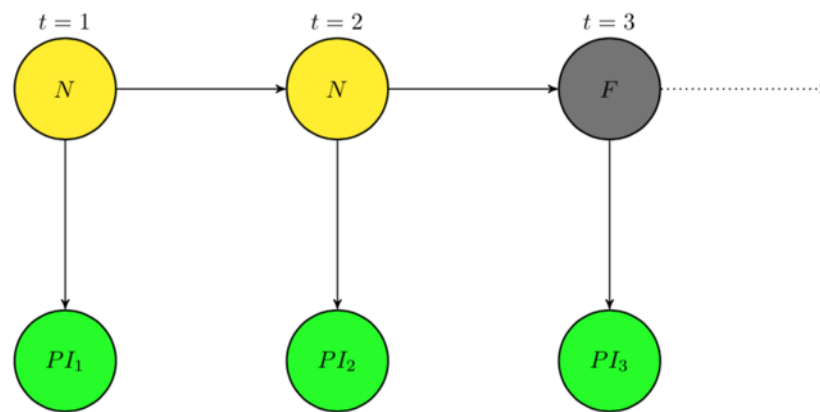


Figure 3. At each time (t), emission probabilities are used to estimate a performance index (PI) given an underlying hidden system state condition (N or F).

2.3. Model Evaluation

The performance of failure detection algorithms is commonly evaluated by way of comparison with other physics-based or NN algorithms [28]. Such approaches, however, only work in simulated settings when the system state is truly known [24,29]. In contrast, one of the biggest challenges for field-based PV failure analyses is the lack of ground truth available for field observations. Since the true system state is not well known for our data, we use two approaches to evaluate PV-HMM generated outputs: TLs and O&M tickets. Thresholds are a common method used in PV failure detection implementations due to their relative simplicity of use [30]. In this analysis, we generate TLs using a percentile approach, wherein data points were compared to different percentiles associated with each site data. At each percentile TL, data that were greater than the associated threshold were flagged as indicating the associated system state. A comparison of these patterns with associated transition probabilities of the PV-HMM enables us to evaluate the general range of TLs, whereby similar probabilities are observed.

As noted above, the TLs provide a basis of comparison but do not necessarily reflect field confirmation. Therefore, the threshold comparisons were augmented by comparing PV-HMMs with O&M tickets, which can capture some observed failures. Previous analyses, however, have indicated that O&M tickets do not capture all possible failures observed at a site [31]. Therefore, evaluation assessments focused on precision analysis to quantify the percentage of O&M tickets that corresponded to a faulted state, as identified by PV-HMMs. A precision metric can be reliable under the assumption that the notated tickets are accurate, without necessarily assuming that the labeling process was comprehensive. The precision metrics are supplemented by site-level visualizations to generate qualitative insights into PV-HMM performance.

3. Results

A total of 104 PV-HMM models were trained as part of this analysis. The transition probabilities evaluated from the Baum–Welch algorithm estimated θ values (not shown) within the PV-HMM models demonstrate a preference towards maintaining a given state (Figure 4). At all sites, when a system is currently in a normal state, it tends to stay normal in its next state. On average, this happens around 93.5% of the time across the sites (blue boxplot). This also indicates that there is a 6.5% probability of a failure in the next state given that the current state is normal (as shown in the red boxplot). On the other hand, we observe a lower probability of 87.1% of the next state being faulted when the current state is faulted (orange boxplot). By extension, this means that a 12.9% probability of returning to a normal state can be achieved given that the current state is faulted (green boxplot). The higher probability percentage from faulted to normal states (vs. normal to faulted) could reflect the site operators' motivation to fix issues quickly to maintain as large an energy output as possible to maximize production.

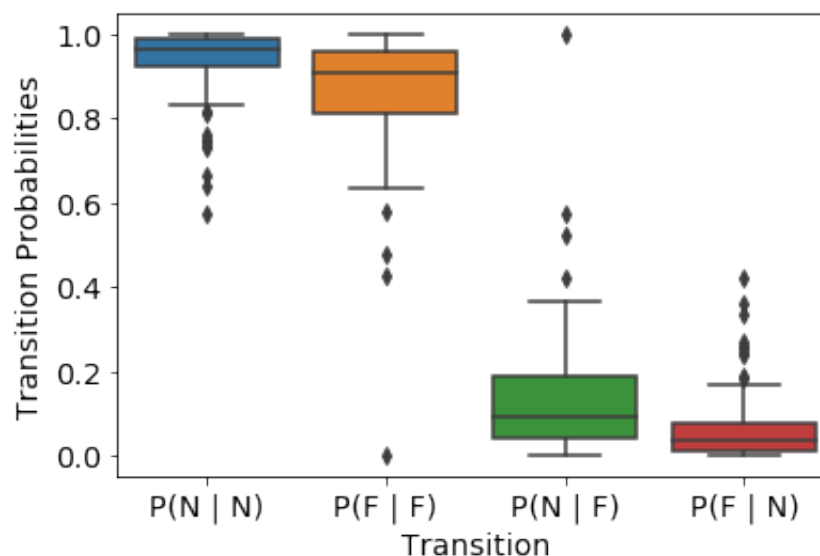


Figure 4. Distribution of PV-HMM transition probabilities across all 104 sites shows the relative preference for maintaining each state within the models (i.e., the state of the next time-step is most likely to be the same as the former).

An evaluation of specific sites indicates that the PV-HMMs capture faulted states to also consider signatures related to heteroskedasticity (i.e., heterogeneous variance) as well as changes in mean values. For example, there is a general decline in PI mean values for one site in October 2018 that is flagged (in red) as faulted by the PV-HMM (Figure 5a). No O&M tickets were observed at this site, so the general decline in PI could reflect environmental stressors that resolved on their own (e.g., snow that melted or soiling removed by rain). Values greater than one PI could reflect under-estimations of expected energy values that arise from irradiance or other sensor errors. Additional faulted states were also identified by the PV-HMM a few months later, capturing high variances in PI. Some sites, however, had more variable PI values in general. Even in such conditions, the PV-HMM was able to recognize faulted conditions (Figure 5b). Finally, in contrast to ML methods which require long periods of data [32], PV-HMMs can be trained on site data with as few as 170 data points (Figure 5c).

TL comparisons generally show that percentile thresholds between 0.80 to 0.98 align with PV-HMM transition probabilities for normal states given that a previous state is also normal (see examples in Figure 6). For faulted to faulted PV-HMM transition probabilities, associated percentile thresholds generally range between 0.18 to 0.42. In addition to the general TL ranges, the site-level plots also highlight the general sensitivity of failure probabilities to specific thresholds, with many sites exhibiting a non-monotonic trend across different percentiles. Furthermore, a lack of ground truth with threshold levels, a common issue with TLs [30], does not provide a complete evaluation for the performance of PV-HMMs.

Therefore, to validate faulted states, precision metrics were calculated for each model (i.e., each site) using associated O&M tickets observed at each site. The calculations demonstrate high precision between the two datasets (median of 82.4% across the sites). This indicates that most of the periods with active O&M tickets were indeed flagged as detected faults by the PV-HMM. Site-level evaluations provide additional insights. For example, we see a deviation in performance identified by the PV-HMM that was missing in the O&M tickets (e.g., April 2018 and October 2018) (Figure 7a). Looking at another system (Figure 7b), the PV-HMM is able to capture more anomalies than identified by O&M tickets, including a notable dip in performance in December 2019. These examples highlight the strengths of the unsupervised learning pursued by PV-HMMs for detecting failures that could be targeted by O&M activities to improve overall site performance.

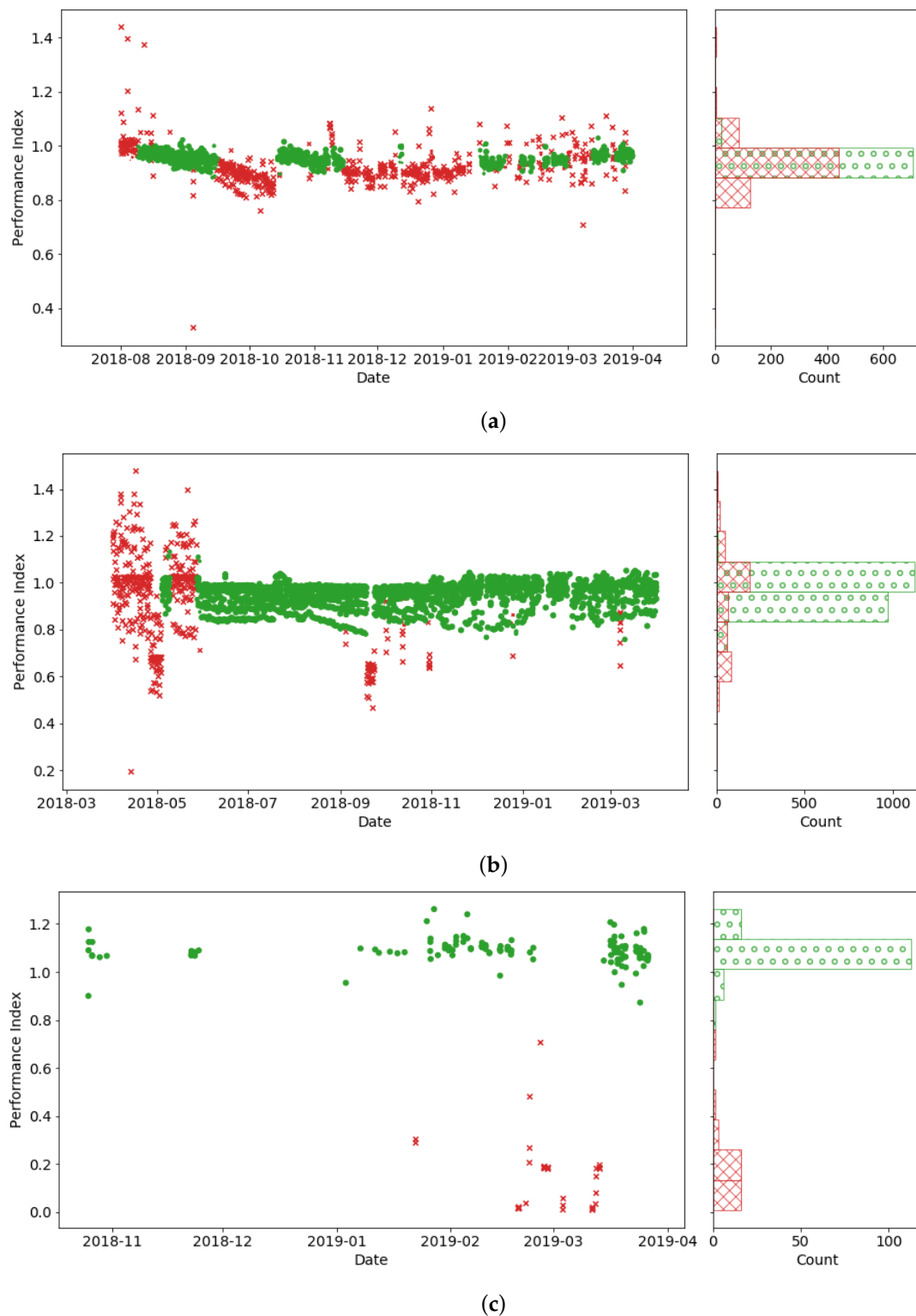
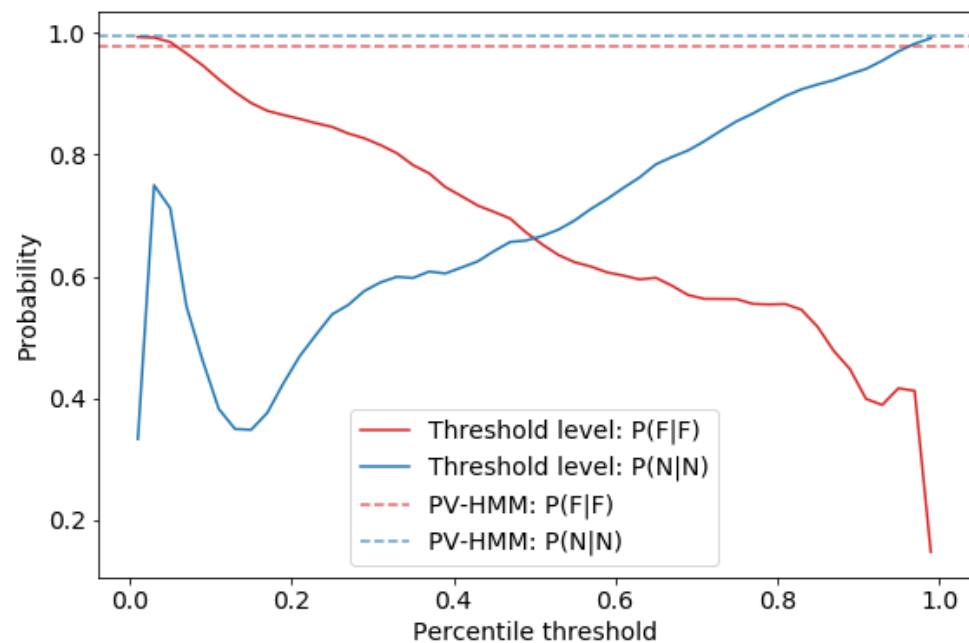
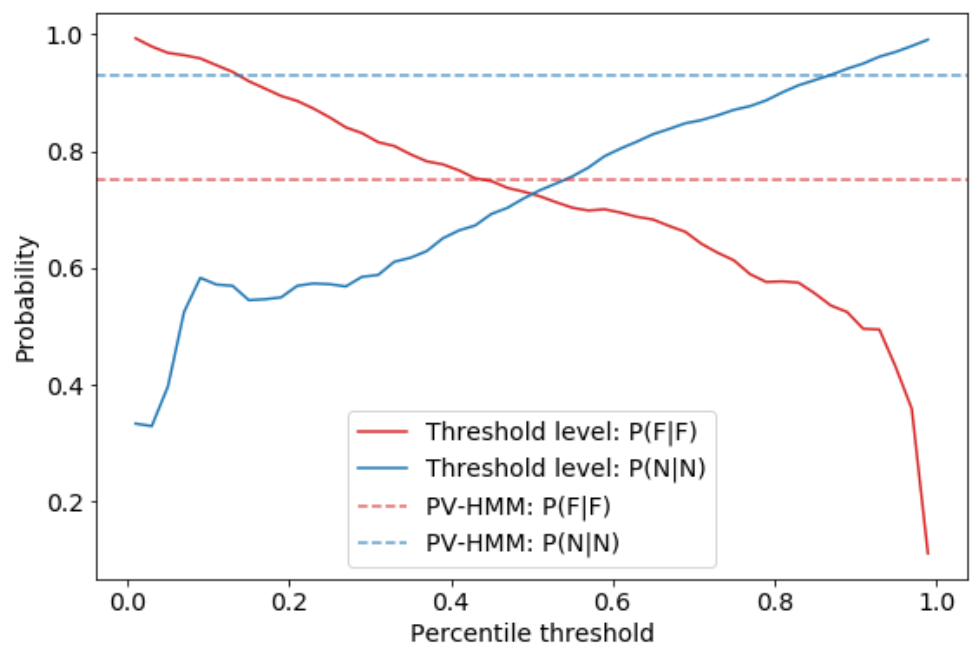


Figure 5. Time series of performance indices and associated emission probabilities for normal states (in green) and faulted states (in red) for three sites. The probability distributions on the right summarize the number of data points at each PI value classified into normal and faulted states. (a) example site with potential soiling issues; (b) example site with high variance in performance index; and (c) example site with bimodal performance index distribution.



(a)



(b)

Figure 6. Visuals of two site-level data (a,b) comparing the probability of failure detection using threshold levels (solid lines) with PV-HMM transition probabilities (dashed lines) plotted for comparison purposes. Lines in red indicate probabilities associated with the faulted (F) state, given that the previous state is also faulted, as while lines in blue indicate probabilities for normal (N) states. Generally, there is a large range of TLs (0.2 to 0.3 percentiles) at which comparable PV-HMM probabilities are observed.



Figure 7. Visuals of two site-level data (a,b) show that more failures (points in yellow) are detected by PV-HMMs (“Detected Outages”) than captured in the maintenance tickets (“O&M overlapping tickets”). The purple points represent samples which have no detected outage.

However, there are limitations in this analysis driven by a lack of ground truth that impacts all field-based PV failure methods. Additional data (across time and geographies) may be useful for comparing predictions to observations in the field. Such insights may also support increased coupling of production and O&M activities to automatically generate O&M tickets when faults are predicted through PV-HMMs. For example, the detected fault states from the PV-HMM also output an associated probability of class, which is used in the Viterbi algorithm to determine the states. Using bootstrapping techniques, these output probabilities can then be associated with confidence bounds that can be used to trigger generation of maintenance review activities in real time. Additionally, having ground truth

data may aid the maximization of PV-HMM accuracy by extending the GMM emission density to incorporate specific exogenic covariates (e.g., weather conditions) whose effects on PV systems' states can be explicitly quantified. The inclusion of these covariates may also provide insight into local conditions that prioritize maintenance activities (e.g., ignoring soiling and snow events). Future analyses may also consider building hierarchical PV-HMMs that take advantage of asset-level data to further support maintenance responses.

As noted above, the PV-HMM leveraged the Viterbi algorithm to calculate the most probable path of the PI into the future as a way to label normal vs. faulted states. These predictions, which were derived from each site's transition and emission profile alone, could be extended into the future to support predictive failure modeling and associated maintenance activities. Although predictive maintenance is common in other industries, general implementation in PV has been limited, in part due to the lack of site-level demonstrations [33]. New methods are needed to effectively quantify cost savings that could arise from predictive maintenance activities. However, solutions will only be successful if they go beyond looking at failure signatures for individual assets (current state of the art; [34]) to consider signatures at the site level (where data are often analyzed). The use of PI as the sole input into the PV-HMM reduces the data requirements for the analysis. Future work should also consider incorporating these methods into open-source package distributions, such as *pvOps* [35], to support ongoing research and application testing of these statistical techniques.

4. Conclusions

This work demonstrates the strengths of an unsupervised statistical approach for labeling faults at PV sites. The PV-HMM methodology was able to identify faults using performance indices across 100+ sites within the United States with high levels of precision relative to O&M ground truth. Unlike ML algorithms, the PV-HMM is able to label historical behaviors, even with limited data points, and does not require a manual setting of threshold levels. This initial implementation of the PV-HMM demonstrates that statistical methods can successfully detect failures using both changes in the mean and variance of datasets. Such approaches have significant potential for supporting O&M activities, such as the detection of failures in real time as well as predictive maintenance activities. Ongoing investigations are needed, however, to evaluate the sensitivity of the model performance to distribution assumptions, explore transition states for specific types of failures, and continue validation with observations in the field. Continued research into these new capabilities for fault diagnosis and prediction within PV systems is warranted to advance the science of PV-HMMs and evaluate the utility of these methods in other (renewable) energy sectors that face similar field challenges with optimizing energy performance.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/en15145104/s1>.

Author Contributions: Conceptualization, M.W.H.; methodology, M.W.H. and L.P.; validation, M.W.H. and T.G.; formal analysis, M.W.H.; writing—original draft preparation, L.P. and T.G.; writing—review and editing, M.W.H.; visualization, M.W.H. and L.P.; supervision, L.P. and T.G.; project administration, T.G.; funding acquisition, T.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by U.S. Department of Energy Solar Energy Technologies Office (Award No. 34172).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data were procured under non-disclosure agreements and thus cannot be shared. However, an anonymized version of the filtered and processed dataset used in this study analysis can be found within the article and Supplementary Information.

Acknowledgments: The authors would like to thank our industry partners for sharing data in support of this analysis, as well as Sam Gilletly for their assistance with reviewing an earlier version of this manuscript. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EM	expectation–maximization
GMM	Gaussian mixture model
ML	machine learning
NN	neural network
O&M	operations and maintenance
PI	performance index
PV	photovoltaics
PV-HMM	PV hidden Markov model
PVROM	PV reliability, operations, and maintenance
TL	threshold level

References

1. International Energy Agency (IEA). *Renewables 2021*; Technical Report; IEA: Paris, France, 2021.
2. Hanawalt, S. *The Challenge of Perfect Operating Data*; PV Magazine: Providence, RI, USA, 2020.
3. Hopwood, M.W.; Gunda, T. Generation of Data-Driven Expected Energy Models for Photovoltaic Systems. *Appl. Sci.* **2022**, *12*, 1872. [[CrossRef](#)]
4. Poulek, V.; Šafránková, J.; Černá, L.; Libra, M.; Beránek, V.; Finsterle, T.; Hrzina, P. PV panel and PV inverter damages caused by combination of edge delamination, water penetration, and high string voltage in moderate climate. *IEEE J. Photovolt.* **2021**, *11*, 561–565. [[CrossRef](#)]
5. Gunda, T.; Hackett, S.; Kraus, L.; Downs, C.; Jones, R.; McNalley, C.; Bolen, M.; Walker, A. A machine learning evaluation of maintenance records for common failure modes in PV inverters. *IEEE Access* **2020**, *8*, 211610–211620. [[CrossRef](#)]
6. Jackson, N.D.; Gunda, T. Evaluation of extreme weather impacts on utility-scale photovoltaic plant performance in the United States. *Appl. Energy* **2021**, *302*, 117508. [[CrossRef](#)]
7. Hopwood, M.W.; Gunda, T.; Seigneur, H.; Walters, J. Neural network-based classification of string-level IV curves from physically-induced failures of photovoltaic modules. *IEEE Access* **2020**, *8*, 161480–161487. [[CrossRef](#)]
8. Akram, M.W.; Li, G.; Jin, Y.; Chen, X.; Zhu, C.; Zhao, X.; Khaliq, A.; Faheem, M.; Ahmad, A. CNN based automatic detection of photovoltaic cell defects in electroluminescence images. *Energy* **2019**, *189*, 116319. [[CrossRef](#)]
9. Harrou, F.; Dairi, A.; Taghezouit, B.; Sun, Y. An unsupervised monitoring procedure for detecting anomalies in photovoltaic systems using a one-class Support Vector Machine. *Sol. Energy* **2019**, *179*, 48–58. [[CrossRef](#)]
10. De Benedetti, M.; Leonardi, F.; Messina, F.; Santoro, C.; Vasilakos, A. Anomaly detection and predictive maintenance for photovoltaic systems. *Neurocomputing* **2018**, *310*, 59–68. [[CrossRef](#)]
11. Huuhtanen, T.; Jung, A. Predictive maintenance of photovoltaic panels via deep learning. In Proceedings of the 2018 IEEE Data Science Workshop (DSW), Lausanne, Switzerland, 4–6 June 2018; pp. 66–70.
12. Lin, P.; Lin, Y.; Chen, Z.; Wu, L.; Chen, L.; Cheng, S. A Density Peak-Based Clustering Approach for Fault Diagnosis of Photovoltaic Arrays. *Int. J. Photoenergy* **2017**, *2017*, 4903613. doi: 10.1155/2017/4903613. [[CrossRef](#)]
13. Liu, S.; Dong, L.; Liao, X.; Cao, X.; Wang, X. Photovoltaic Array Fault Diagnosis Based on Gaussian Kernel Fuzzy C-Means Clustering Algorithm. *Sensors* **2019**, *19*, 1520. doi: 10.3390/s19071520. [[CrossRef](#)]
14. Dimitrievska, V.; Pittino, F.; Muehleisen, W.; Diwald, N.; Hilweg, M.; Montvay, A.; Hirschl, C. Statistical Methods for Degradation Estimation and Anomaly Detection in Photovoltaic Plants. *Sensors* **2021**, *21*, 3733. [[CrossRef](#)] [[PubMed](#)]
15. Kim, J.; Rabelo, M.; Padi, S.P.; Yousuf, H.; Cho, E.C.; Yi, J. A Review of the Degradation of Photovoltaic Modules for Life Expectancy. *Energies* **2021**, *14*, 4278. doi: 10.3390/en14144278. [[CrossRef](#)]
16. Bertsimas, D.; Delarue, A.; Jaillet, P.; Martin, S. The Price of Interpretability. *arXiv* **2019**, arXiv:1907.03419.
17. Gunda, T.; Homan, R. *Evaluation of Component Reliability in Photovoltaic Systems Using Field Failure Statistics*; Sandia National Lab. (SNL-NM): Albuquerque, NM, USA, 2020

18. Cristaldi, L.; Khalil, M.; Faifer, M.; Soulatiantork, P. Markov process reliability model for photovoltaic module encapsulation failures. In Proceedings of the 2015 International Conference on Renewable Energy Research and Applications (ICRERA), Palermo, Italy, 22–25 November 2015; pp. 203–208.
19. Gales, M.; Young, S. The Application of Hidden Markov Models in Speech Recognition. *Found. Trends Signal Process* **2008**, *1*, 195–304. [[CrossRef](#)]
20. Patel, L.; Gustafsson, N.; Lin, Y.; Ober, R.; Henriques, R.; Cohen, E. A hidden Markov model approach to characterizing the photo-switching behavior of fluorophores. *Ann. Appl. Stat.* **2019**, *13*, 1397–1429. [[CrossRef](#)]
21. Watkins, R.; Eagleson, S.; Veenendaal, B.; Wright, G.; Plant, A. Disease surveillance using a hidden Markov model. *BMC Med. Inform. Decis. Mak.* **2009**, *9*, 39. [[CrossRef](#)]
22. Ullah, I.; Ahmad, R.; Kim, D. A prediction mechanism of energy consumption in residential buildings using hidden markov model. *Energies* **2018**, *11*, 358. [[CrossRef](#)]
23. Zhao, D.; Zhou, Z.; Tang, S.; Cao, Y.; Wang, J.; Zhang, P.; Zhang, Y. Online estimation of satellite lithium-ion battery capacity based on approximate belief rule base and hidden Markov model. *Energy* **2022**, *256*, 124632. [[CrossRef](#)]
24. Kouadri, A.; Hajji, M.; Harkat, M.F.; Abodayeh, K.; Mansouri, M.; Nounou, H.; Nounou, M. Hidden Markov model based principal component analysis for intelligent fault diagnosis of wind energy converter systems. *Renew. Energy* **2020**, *150*, 598–606. [[CrossRef](#)]
25. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* **1977**, *39*, 1–38.
26. Bilmes, J.A. A Gentle Tutorial on the EM Algorithm and Its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Technical Report. 1997. Available online: https://f.hubspotusercontent40.net/hubfs/8111846/Unicon_October2020/pdf/bilmes-em-algorithm.pdf (accessed on 10 July 2022).
27. Rabiner, L. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **1989**, *77*, 257–286. doi: 10.1109/5.18626. [[CrossRef](#)]
28. Lazzaretti, A.E.; Costa, C.H.D.; Rodrigues, M.P.; Yamada, G.D.; Lexinoski, G.; Moritz, G.L.; Oroski, E.; Goes, R.E.D.; Linhares, R.R.; Stadzisz, P.C.; et al. A monitoring system for online fault detection and classification in photovoltaic plants. *Sensors* **2020**, *20*, 4688. [[CrossRef](#)]
29. Garoudja, E.; Chouder, A.; Kara, K.; Silvestre, S. An enhanced machine learning based approach for failures detection and diagnosis of PV systems. *Energy Convers. Manag.* **2017**, *151*, 496–513. [[CrossRef](#)]
30. Livera, A.; Theristis, M.; Makrides, G.; Sutterlueti, J.; Georghiou, G.E. Advanced diagnostic approach of failures for grid-connected photovoltaic (PV) systems. In Proceedings of the 35th European Photovoltaic Solar Energy Conference (EU PVSEC), Brussels, Belgium, 24–28 September 2018.
31. Mendoza, H.; Hopwood, M.; Gunda, T. pvOps: Improving operational assessments through data fusion. In Proceedings of the 2021 IEEE 48th Photovoltaic Specialists Conference (PVSC), Fort Lauderdale, FL, USA, 20–25 June 2021; pp. 0112–0119.
32. Müller, B.; Holl, N.; Reise, C.; Kiefer, K.; Kolloch, B.; Branco, P.J. Practical Recommendations for the Design of Automatic Fault Detection Algorithms Based on Experiments with Field Monitoring Data. *arXiv* **2022**, arXiv:2203.01103.
33. Bosman, L.B.; Leon-Salas, W.D.; Hutzler, W.; Soto, E.A. PV system predictive maintenance: Challenges, current approaches, and opportunities. *Energies* **2020**, *13*, 1398. [[CrossRef](#)]
34. Gawre, S.K. Advanced Fault Diagnosis and Condition Monitoring Schemes for Solar PV Systems. In *Planning of Hybrid Renewable Energy Systems, Electric Vehicles and Microgrid*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 27–59.
35. Sandia National Laboratories. pvOps. Available online: <https://github.com/sandialabs/pvOps> (accessed on 25 June 2022).