energies

MDPI

Article

Improved Semi-Supervised Data-Mining-Based Schemes for Fault Detection in a Grid-Connected Photovoltaic System

Benamar Bouyeddou [1,2], Fouzi Harrou [3], Bilal Taghezouit [4,5], Ying Sun [3] and Amar Hadj Arab [4,*]

[1] LESM Lab., Faculty of Technology, University of Saida-Dr Moulay Tahar, Saida 20000, Algeria
[2] STIC Lab., Department of Telecommunications, Abou Bekr Belkaid University, Tlemcen 13000, Algeria
[3] Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
[4] Centre de Développement des Energies Renouvelables (CDER), B.P. 62, Route de l'Observatoire, Algiers 16340, Algeria
[5] Laboratoire de Dispositifs de Communication et de Conversion Photovoltaique, Ecole Nationale Polytechnique Alger, Algiers 16200, Algeria
* Correspondence: a.hadjarab@cder.dz

Abstract: Fault detection is a necessary component to perform ongoing monitoring of photovoltaic plants and helps in their safety, maintainability, and productivity with the desired performance. In this study, an innovative technique is introduced by amalgamating Latent Variable Regression (LVR) methods, namely Principal Component Regression (PCR) and Partial Least Square (PLS), and the Triple Exponentially Weighted Moving Average (TEWMA) statistical monitoring scheme. The TEWMA scheme is known for its sensitivity to uncovering changes of small magnitude. Nevertheless, TEWMA can only be utilized for monitoring single variables and ignoring the correlation among monitored variables. To alleviate this difficulty, the LVR methods (i.e., PCR and PLS) are used as residual generators. Then, the TEWMA is applied to the obtained residuals for fault detection purposes, where the detection threshold is computed via kernel density estimation to improve its performance and widen its applicability in practice. Real data with different fault scenarios from a 9.54 kW photovoltaic plant has been used to verify the efficiency of the proposed schemes. Results revealed the superior performance of the PLS-TEWMA chart compared to the PLS-TEWMA chart, particularly in detecting anomalies with small changes. Moreover, they have almost comparable performance for large anomalies.

Keywords: fault detection; photovoltaic systems; data-driven methods; TEWMA; sensor faults; PLS; PCR; dimensionality reduction

check for updates

Citation: Bouyeddou, B.; Harrou, F.; Taghezouit, B.; Sun, Y.; Hadj Arab, A. Improved Semi-Supervised Data-Mining-Based Schemes for Fault Detection in a Grid-Connected Photovoltaic System. *Energies* **2022**, *15*, 7978. https://doi.org/10.3390/en15217978

Academic Editor: Manolis Souliotis

Received: 5 October 2022
Accepted: 24 October 2022
Published: 27 October 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/).

1. Introduction

The energy supply of the future must be clean, efficient, resource-efficient, and sustainable. Solar Photovoltaic (PV) energy is a strategic option to satisfy environmental concerns and meet the growing electricity demand [1]. It offers a safe, affordable, and renewable solution for generating power on a large scale. The latest statistics show that the installed PV systems reached a total installed capacity of 942 GWdc (Gigawatt direct current) in 2021 [2]. Although there are advanced PV technologies, PV systems are still confronted with different types of anomalies in real working conditions, such as short-circuit, open-circuit, line-line, and partial shading, that decrease their performance and cause power losses [3]. Undoubtedly, such unsatisfactory faulty conditions raise undesirable maintenance costs and inefficiency.

Anomalies and atypical conditions (e.g., dirt, shading) can make the PV system deviate from the desired performance. Broadly speaking, both internal and external anomalies in PV systems can degrade their operating performance, collapsing the system and even resulting in severe security issues [4]. Internal faults include line-to-line faults, short/open circuits,

and defective components (e.g., generators, cables, and inverters). In addition, external anomalies are related to some environmental circumstances, like dusty conditions [5], and subversive acts, such as cyber-attacks (e.g., Denial of services and data integrity), particularly against on-grid plants [6–8]. In the absence of convenient action, the system lifetime can be drastically reduced in the long term [9]. Thus, PV systems need to be continuously monitored for proper functioning, preserving their optimal performances and making them more efficient, reliable, and sustainable [10].

Detecting anomalies in PV systems is essential to avoid energy drops and keep their expected performance. Towards this end, numerous anomaly detection techniques have been developed in the literature last two decades. Such techniques can be grouped into two principal classes, namely, model-based or data-driven methods [11–17]. Model-based approaches involve analytical models to predict normal (i.e., fault-free) PV system outputs. Abnormal measurements are flagged when their deviations from the predictions exceed the set threshold [13,16,17]. Various model-based approaches have been introduced in the literature, such as single-diode-based models [18], double-diode-based models [19], and Kalman filter [20]. For instance, Skomedal et al. [21] compared six monitoring charts, the conventional Shewhart, three modified versions of CUSUM (Cumulative Sum), and a modified modification of EWMA (Exponentially Weighted Moving Average), based on an autoregressive integrated moving average model to detect low-rate losses in sizable grand serial PV farms. They conclude that the best performance is obtained when using CUSUM-based charts in this application. In [18], an approach to detect and identify faults in the DC side of a PV plant is presented. This approach applies the Multivariate EWMA to residuals obtained from the one-diode-based model for fault detection and the univariate EWMA to residuals to discriminate the types of detected anomalies. Of course, the efficiency of the model-based anomaly detection methods depends on the analytical model, which is not often easy to obtain and time-consuming, particularly for large-scale systems.

Data-driven approaches, on the other hand, exploit available measurements when the monitored PV system is functioning properly to learn the underlying reference empirical model. Then, new faulty measurements are uncovered accordingly [22–24]. Data-based anomaly detection techniques enclose statistical and machine learning-based techniques, e.g., Support Vector Machine (SVM) and k-Nearest Neighbors (k-NN) algorithm. Within data-based techniques, in [25], an approach based on Principal Component Analysis (PCA) is proposed to monitor the DC side of a PV system. More specifically, PCA is applied to select the most relevant features and generate residuals. Here, residuals are evaluated via the multivariate charts Hotelling T2 and Squared Predicted Error (SPE) charts for anomaly detection purposes. This approach showed satisfactory detection performance in detecting different real anomalies, including short-circuit, inverter connection, and partial shading. In [26], the PCA approach showed performance in detecting and classifying PV shading under real conditions. In another study [27], a PCA-based image processing method was introduced to detect faults in PV tracking systems. Deceglie et al. [28] proposed the Stochastic Rate and Recovery (SRR) approach to compute losses related to the soiling problem. The SRR defines cleaning and soiling periods using short-circuit current measurements. Then soiling ratios are used to discriminate soiling profiles, and the resulting losses are estimated according to prefixed cleaning thresholds. This study revealed that profile construction is very sensitive to noise, and the estimation threshold should be judiciously set. In [29], Kilic et al. addressed open and short circuit-related failures with the Random Vector Functional Link Networks (RVFLN) and Canonical Correlation Analysis (CCA) as data reduction techniques to reduce the dimension of selected attributes.

In [30], Xiong et al. studied the utility of Discrete Wavelet Transform (DWT) to identify and isolate parallel and series arc faults. Specifically, the DWT-based technique is applied to inspect capacitor current features, and results show significant variations during arc faults, which allow for the revealing of their type and location. Nevertheless, the analysis is based on simplified alleged data and specific arc faults. Indeed, it should be generalized for other fault types onboard real PV systems. The approach proposed in [31] employs

Wavelets Packet Transform (WPT). Here, three PV plant's variables have been chosen to perform de detection procedure: voltage, voltage/energy, and impedance/energy. For each variable, preset thresholds obtained through simulation are used to define the intervals of normal variations. In [32], Edun et al. presented a method to detect disconnections in PV systems using the Spread Spectrum Time Domain Reflectometry (SSTDR). It consists of correlating generated to reflected signals, and the obtained baseline peaks are utilized to reveal and locate the related faults. However, the SSTDR is too sensitive to noise and attenuation, making it relatively adapted to a limited number of cells. In addition, the use of this approach gets more complicated in large-scale PV plants where a large number of reflections can strongly degrade their detection performance.

Recently, several supervised and semi-unsupervised machine leaning-based techniques have been explored for fault detection and classification in PV systems [33–36]. Supervised methods require the availability of labeled data during the training, which is not often easy to obtain, particularly for large-scale systems with various faults. On the other hand, semi-supervised methods need only fault-free data in training and without fault labeling. In [37], individual modules' voltages are estimated by means of multi-layer neural networks to find out short circuit situations. Results showed that detection accuracy increased proportionally with the number of used layers. However, the training of this model requires faulty measurements, which is not always feasible, even if it does not cover all expected faults. Chen et al. [38] presented a multi faults detection approach based on enhanced Semi-Supervised Ladder Networks (SSLN) when using restricted online samples. In [39], an auto-encoder-driven deep neural network approach is designed to deal with operational and cyber-attacks related faults in grid-connected PV systems. The auto-encoder is applied to estimate the normal measurements with the lowest error based on multivariate features. Then faults are reported when estimation error exceeds an empirical detection threshold based on the well-known three-sigma rule. In [40], the SVM algorithm is applied to quantify power quality and isolate eventual disturbances within client-side measurements. First, the DWT selects and extracts the most relevant features. Next, the One-Class SVM (OCSVM) is involved in revealing disturbances. Finally, a Multi-Class SVM (MCSVM) procedure is performed to classify the detected disturbances.

Anomaly detection in PV plants is essential for improving their safety and profitability and optimizing maintenance scheduling. Possible anomalies could be caused by malfunctioning sensors (sensor anomalies) or changes affecting the inspected system (called process anomalies). The need for methods to accurately and quickly uncover abnormal conditions (sensor or process anomalies) is vital to keep the PV system operating with the desired performance. This paper presents a data-driven approach for detecting sensor anomalies and anomalies in the DC side of the PV system. Importantly, the aim is to develop a semi-supervised data-driven detector for PV systems monitoring that does not require labeled data. In this study, we propose a Triple Exponentially Weighted Moving Average (TEWMA)-based strategy to detect anomalies encountered in the DC side of a PV plant. The TEWMA scheme is an effective and sensitive approach, which is found to have better detection of small changes, usually adopted for inspecting only single process variables. However, data from PV systems are multivariate input-output variables and cross-correlated, making the use of the TEWMA chart inappropriate. To bypass this shortcoming, the proposed method integrates Latent Variable Regression (LVR) methods with the TEWMA chart. LVR models, namely, Principal Component Regression (PCR) and Partial Least Square Regression (PLS), demonstrated good capability in modeling multivariate input-output data in different applications by eliminating variables' collinearity and reducing noises as well. In addition, PLS/PCR models have been widely employed for response variables prediction based on predictor variables. They perform well when transforming high dimensional input data to lower dimension outputs that can be effectively used to track abnormal situations. Importantly, the proposed LVR-TEWMA approach employs two complementary steps in detecting anomalies. The first involves the use of the LVR models (i.e., PLS and PCR) to generate the residual sequence and the TEWMA decision threshold based on data from the

PV system under nominal conditions. In the second step, the TEWMA charting statistic is computed using residuals from the reference LVR model based on the newly received data. Comparison of the TEWMA statistic to the determined threshold value yields the anomaly decision (faulty or healthy conditions). Note that the threshold in the conventional TEWMA chart is analytically determined based on the Gaussian assumption of the data distribution. However, this assumption is not valid for the data from PV systems. Here, we extend the TEWMA chart to monitor non-Gaussian data by using the Kernel density estimator to determine the detection threshold.

This work is organized as follows. Section 2 is devoted to the description of the used grid-tied PV system. Then, Section 3 presents the LVR models (i.e., PLS and PCR) for PV system modeling. Section 4 briefly describes the proposed LVR-TEWMA monitoring method and discusses the experimental results. The last section includes this study and provides some future directions.

2. PV Installation Description

This section is devoted to presenting shortly the Grid-Connected Photovoltaic (GCPV) installation used in this study. Indeed, the proposed fault detection algorithm in this work will be verified using the meteorological and electrical data measurement collected from a PV installation at the Centre de Developpement des Energies Renouvelables (CDER) in Algeria, with a total power of 9.54 kWdc in operation since 2004 (Figure 1), where the total solar PV power produced is injected into the low-voltage electrical grid. As shown in Figure 2, the PV system is composed of three identical single-phase PV sub-systems. Each PV sub-system contains a PV sub-array of 30 PV Modules (PVM), a 2.5-kilowatt (kW) grid-tie inverter, and electrical cabinets for protection.

The PV array is divided into three sub-arrays. Each PV sub-array contains two parallel strings of 15 PVM in a series. The main electrical specifications of the PV Modules/sub-array and the grid-tie inverter are shown respectively in Tables 1 and 2.

Where **STC** symbolizes the Standard Test Conditions ($\mathbf{G}$ = 1000 W/m^2, $\mathbf{T_C}$ = 25 °C, and **AM** = 1.5) and **MPP** symbolizes the Maximum Power Point. **G** is the received Irradiance by the PV module during the flash test, $\mathbf{T_C}$ is the temperature of the PV Cell, and **AM** is the Air Mass. $\mathbf{I_{SC}}$ is the short circuit current, $\mathbf{V_{OC}}$ is the open circuit voltage, $\mathbf{I_{MPP}}$ is the current at MPP, $\mathbf{V_{MPP}}$ is the voltage at MPP, and $\mathbf{P_M}$ is the maximum power.

The meteorological and electrical measured data used in this work are recovered by an external monitoring system composed essentially of Sensors, a Data Acquisition unit (Agilent 34970A), and software under PC (Figure 3).



Figure 1. Picture of the PV Array of the GCPV system on the rooftop at CDER.

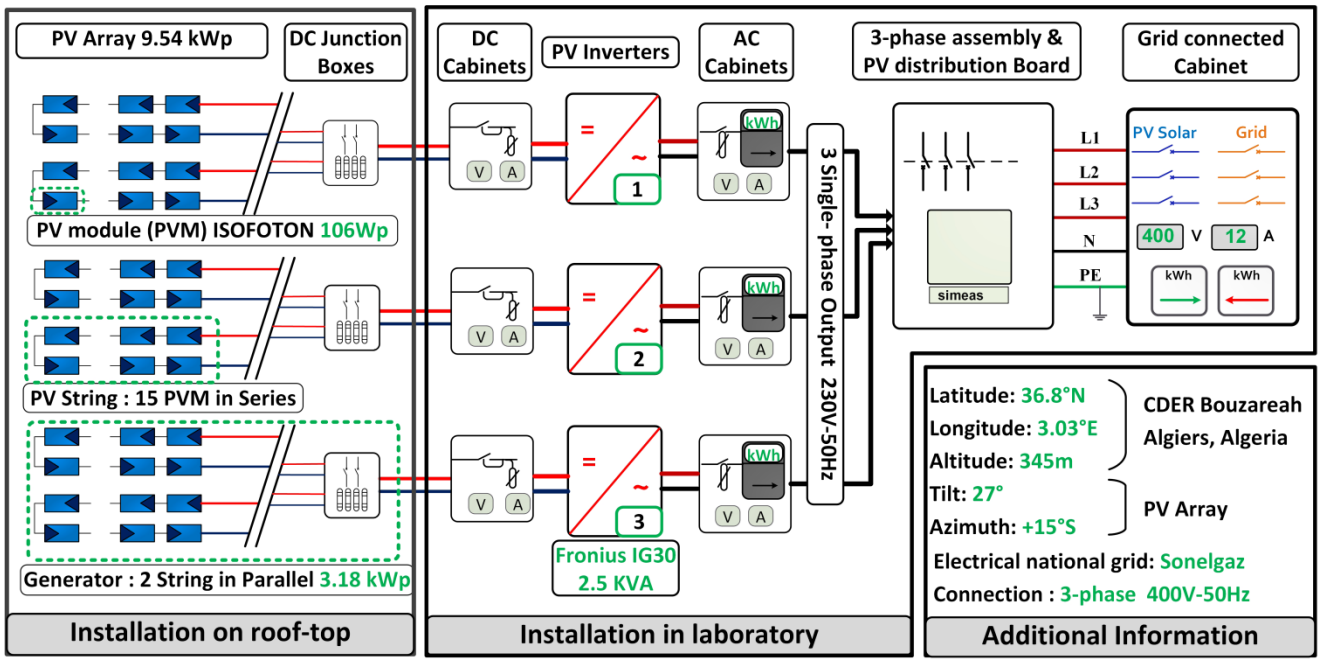


Figure 2. Illustrative diagram of the used grid-connected PV system.

Table 1. Electrical Characteristic of PV Module and PV sub-array Isofoton I106-12.

Parameters	I_{SC} (A)	V_{OC} (V)	I_{MPP} (A)	V_{MPP} (V)	P_M (W)
PV Module	6.54	21.6	6.1	17.4	106
PV Sub-Array	13.08	324	12.2	261	3180

Table 2. Electrical Characteristic of PV Inverter Fronius IG 30 under nominal conditions.

Parameters	Nominal AC Power (W)	DC Voltage Range (V)	Inverter Efficiency (%)	AC Voltage Range (V)	Frequency Range (Hz)
Value	2500	150–400	92.7–94.3	195–253	49.8–50.2

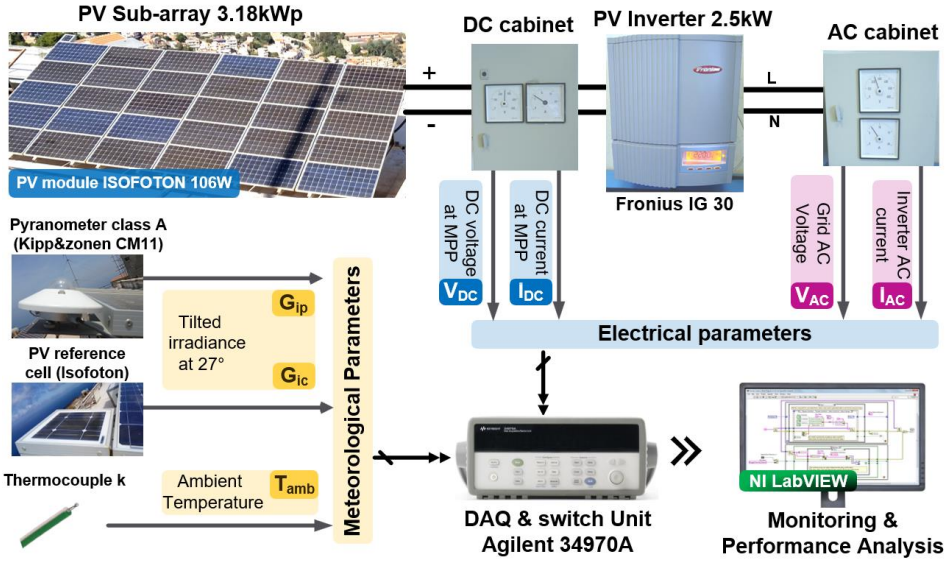


Figure 3. Synoptic diagram of the PV monitoring system.

For the measure of tilted irradiance at 27 °C, a pyranometer (kipp&zonen CM11) and a reference cell are used, and a thermocouple measures the ambient temperature. Two hall-effect sensors (FW BELL CLSM-50S) were used to measure the current on both the Direct current (DC) and Alternating Current (AC) sides of the grid-tied inverter. The DC voltage at the MPP of the PV sub-array is measured by a simple voltage divider circuit; on the other hand, a voltage transformer was used to measure the grid AC voltage.

The data acquisition and switch unit (Agilent 34970A) provides conditioning and measurement of sensor signals. While the monitoring interface is designed under LabVIEW software, this interface can recover, display, record, and analyze the measured data. According to IEC 61724 standard, the sampling time was chosen at 1 min.

3. Materials and Methods

Here, we briefly introduce the KDE-Triple EWMA approach, PCR, and PLS-based anomaly detection schemes.

3.1. PLS (Partial Least Square)

PLS, also called Projection to Latent Structures, is a well-known data-driven dimensionality reduction technique that has been extensively applied to monitor multivariate and highly correlated variables [41–44]. The essence of PLS model consists of transforming input and output data into a lower dimension space so that their corresponding latent variables have the maximum variance [43,44]. Importantly, PLS constructs a model linking the latent variables (LVs) associated with X and Y. The PLS model comprises two complementary models: inner and outer models (Figure 4).

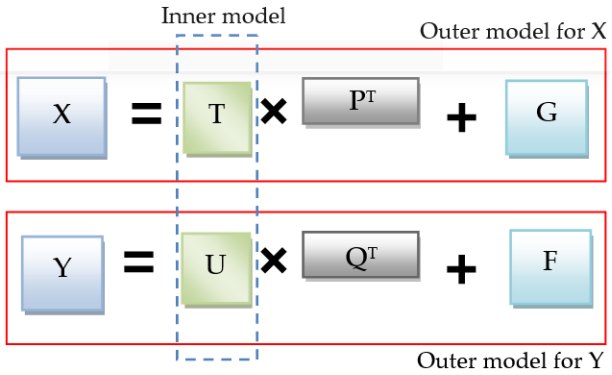


Figure 4. General design of PLS model.

Specifically, for data inputs X with n observations from m variables, $X \in R^{n \times m}$ and data output Y with n observations from p variables, $Y \in R^{n \times p}$, the PLS model (i.e., outer model) can be expressed as follows [42,45]:

$$\begin{cases} X = \sum_{i=1}^{l} tp_i^T = TP^t + G \\ Y = \sum_{i=1}^{l} uq_i^T = UQ^t + F' \end{cases} \quad (1)$$

where $T \in R^{n \times l}$ and $U \in R^{n \times q}$ represent the latent variable matrices associated with input and output matrices, respectively. The loading matrices of input and output space are $P \in R^{m \times l}$ and $Q \in R^{p \times q}$, respectively. G and F represent the PLS model residuals matrices. Here, l is the number of retained principal components (PCs), which can be calculated using the cross-validation method [46]. The retained latent variables u_j and t_j are estimated iteratively [47], and the PLS model establishes the relationship (i.e., inner model) between T and U as:

$$U = T\beta + E, \quad (2)$$

where β denotes the regression matrix and **E** is the residual matrix. Finally, the output variable Y is obtained via the following expression:

$$Y = T\beta Q^T + F^* \tag{3}$$

Various extensions to the linear PLS model have been reported in the literature to widen the applicability of PLS models in practice, including nonlinear PLS [41], wavelet-driven PLS [48], recursive PLS [49], and dynamic PLS [50].

3.2. PCR (Principal Component Regression)

We now describe the PCR model. The PCR method is frequently used in dimensionality reduction and regression to handle the collinearity problem in multivariate data. PCR is performed in two complementary phases. At first, the data reduction procedure is applied to explanatory variables **X** using Principal component analysis (PCA). After that, the resulting retained principal components (CPs) are then linked to the response variables via the Ordinary Least Squares (OLS) regression technique [51,52] (Figure 5).

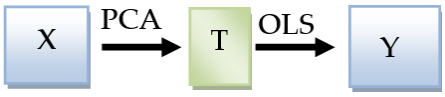


Figure 5. General representation of PCR model.

After applying the PCA algorithm, the input data matrix X is decomposed as follows [52]:

$$X = TW^T = \sum_{i=1}^{k} t_i w_i^T + \sum_{i=K+1}^{m} t_i w_i^T = \hat{X} + E, \tag{4}$$

where $\hat{\mathbf{X}}$, and **E** are the approximated and residual matrices, respectively. Here, $T \in R^{n\times m}$ and $W \in R^{m\times m}$ represent the principal components (PCs) and the loading matrices, respectively. Notice that the cumulative percentage variance (CPV) technique is widely employed to decide the number of PCs to be retained in the model [53]. Therefore, PCR constructs the linear regression between the matrix $\hat{\mathbf{T}}$ of the *k* retained principal components and the response variable **y** as the solution for the following optimization problem:

$$\hat{\beta} = \arg\min_{\beta}(\| \hat{T}\beta - y \|_2^2). \tag{5}$$

The least squares solution is obtained as:

$$\hat{\beta} = (\hat{T}^T\hat{T})^{-1}\hat{T}^T y. \tag{6}$$

More details about PCR and PLS-driven anomaly techniques can be found in [45].

3.3. TEWMA (Triple Exponential Weighted Moving Average)

The TEWMA (i.e., Triple EWMA) is an enhanced variant of the conventional single EWMA monitoring approach, built essentially for conveniently uncovering changes with small and moderate levels. Precisely, the decision statistic is based on applying triple EWMA of the actual and past data points [54,55]. For a given monitored variable X x_i, $i = 1, \ldots n$, samples normally distributed follow normal distribution $N(\mu, \sigma^2)$, the single EWMA statistic E_i is defined as:

$$E_i = \lambda_1 x_i + (1 - \lambda_1)E_{i-1}. \tag{7}$$

Then the double EWMA statistic, DE_i, is expressed as:

$$DE_i = \lambda_2 E_i + (1 - \lambda_2)DE_{i-1}. \tag{8}$$

Finally, the Triple TEWMA statistic, TE_i can be calculated as follow:

$$TE_i = \lambda_3 DE_i + (1 - \lambda_3) TE_{i-1}. \quad (9)$$

The smoothing parameters λ_1, λ_2, and λ_3 can either be equal or different. It has been shown that the Double EWMA (DEWMA) maintains mostly the same performance for the same or different smooth parameters [56,57]. The TEWMA charting statistic with the same smoothing parameter is rewritten as follows [54,58]:

$$\begin{cases} E_i = \lambda x_i + (1 - \lambda) E_{i-1}, \\ DE_i = \lambda E_i + (1 - \lambda) DE_{i-1}, \\ TE_i = \lambda DE_i + (1 - \lambda), \end{cases} \quad (10)$$

where $E_0 = DE_0 = TE_0 = \mu_0$ initialize the three statistics and usually set to be the anomaly-free mean. As a weighted average, statistic E_i can be written as [54,58]:

$$\begin{aligned} E_i &= \lambda x_i + (1 - \lambda) E_{i-1} \\ &= \lambda x_i + (1 - \lambda)[\lambda x_{i-1} + (1 - \lambda) E_{i-2}] \\ &= \lambda [x_i + (1 - \lambda) x_{i-1}] + (1 - \lambda)^2 E_{i-2}. \end{aligned} \quad (11)$$

Hereby, the E_i statistic of the present sample can be computed as pondered sum function of all paste samples *j* [54,56]:

$$E_i = \lambda \sum_{j=1}^{i} (1 - \lambda)^{i-j} x_j + (1 - \lambda)^i E_0, \quad (12)$$

Similarly, DE_i and TE_i can be written as [54,56]:

$$DE_i = \lambda \sum_{j=1}^{i} (1 - \lambda)^{i-j} E_j + (1 - \lambda)^i DE_0, \quad (13)$$

$$TE_i = \lambda \sum_{j=1}^{i} (1 - \lambda)^{i-j} DE_j + (1 - \lambda)^i TE_0. \quad (14)$$

After replacing E_i in Equation (13), DE_i can be expressed in terms of x_i as:

$$\begin{aligned} DE_i &= \lambda \sum_{j=1}^{i} (1 - \lambda)^{i-j} \left[\lambda \sum_{k=1}^{j} (1 - \lambda)^{j-k} x_k + (1 - \lambda)^j E_0 \right] + (1 - \lambda)^i DE_0 \\ &= \lambda^2 \sum_{j=1}^{i} \sum_{k=1}^{j} (1 - \lambda)^{i-k} x_k + \lambda \sum_{j=1}^{i} (1 - \lambda)^i E_0 + (1 - \lambda)^i DE_0 \\ &= \lambda^2 \sum_{k=1}^{i} \sum_{j=k}^{i} (1 - \lambda)^{i-k} x_k + \lambda i (1 - \lambda)^i E_0 + (1 - \lambda)^i DE_0. \end{aligned} \quad (15)$$

We now substitute E_0 into DE_0 to obtain the final pondered form [54]:

$$DE_i = \lambda^2 \sum_{k=1}^{i} (1 - \lambda)^{i-k} (i - k + 1) x_k + (\lambda i + 1)(1 - \lambda)^i DE_0 \quad (16)$$

After that, by substituting (16) in (14), TE_i is expressed as [54]:

$$\begin{aligned} TE_i &= \lambda \sum_{j=1}^{i} (1-\lambda)^{i-j} \left[\lambda^2 \sum_{k=1}^{j} (1-\lambda)^{j-k} (j-k+1) x_k + (\lambda j+1)(1-\lambda)^j DE_0 \right] + (1-\lambda)^i TE_0 \\ &= \lambda^3 \sum_{j=1}^{i} \sum_{k=1}^{j} (1-\lambda)^{i-k} (j-k+1) x_k + \lambda \sum_{j=1}^{i} (1-\lambda)^i (\lambda j+1) DE_0 + (1-\lambda)^i TE_0 \\ &= \lambda^3 \sum_{k=1}^{i} \sum_{j=k}^{i} (1-\lambda)^{i-k} (j-k+1) x_k + \lambda \sum_{j=1}^{i} (1-\lambda)^i (\lambda j+1) DE_0 + (1-\lambda)^i TE_0 \\ &= \frac{\lambda^3}{2} \sum_{k=1}^{i} (1-\lambda)^{i-k} (i-k+1)(i-k+2) x_k + \left(\frac{\lambda}{2}\right) (1-\lambda)^i i(\lambda i + \lambda + 2) DE_0 + (1-\lambda)^i TE_0 \end{aligned} \tag{17}$$

Finally, we substitute DE_0 into TE_0, to obtain the TE_i statistic as:

$$TE_i = \frac{\lambda^3}{2} \sum_{k=1}^{i} (1-\lambda)^{i-k} (i-k+1)(i-k+2) x_k + \left(\frac{(1-\lambda)^i}{2} \right) [\lambda i(\lambda i + \lambda + 2) + 2] TE_0. \tag{18}$$

The monitoring scheme utilizing TEWMA defines the upper and lower control limits as [54]:

$$UCL, LCL = \mu_0 \pm L\sigma \sqrt{\frac{6(1-\lambda)^6 \lambda}{(2-\lambda)^5} + \frac{12(1-\lambda)^4 \lambda^2}{(2-\lambda)^4} + \frac{7(1-\lambda)^2 \lambda^3}{(2-\lambda)^3} + \frac{\lambda^4}{(2-\lambda)^2}}, \tag{19}$$

where TE_i values below the limits indicate that the monitored system operates without anomalies. On the other hand, anomalies can be identified when TE_i values overpass the defined decision limits.

Note that the decision limits in (19) are calculated under the normality distribution assumption of data. Nevertheless, this is not often satisfied, which could result in increased false alarms rate and missed detections. To overcome this problem, kernel density estimation (KDE) is adopted in this study to compute the decision limit.

3.4. KDE-TEWMA (Kernel Density Estimation TEWMA)

In this section, we introduce KDE, which is commonly used to estimate the probability distribution of a given data, and how it can be used to compute the TEWMA threshold. Let us consider $X = x_1; \cdots; x_n$ represents the residual vector using PLS or PCR model. The KDE is applied for estimating the probability density function (PDF) of the residuals by the following formula [59]:

$$p(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h} \right), \tag{20}$$

where x_i is the ith observation. K is the kernel function; the Gaussian kernel is usually used [60]:

$$K(x) = \frac{1}{2\pi} \exp\left(\frac{-\theta^2}{2} \right) \tag{21}$$

In (20), h represents the smoothing bandwidth factor that determines the probability estimation quality. H is the smoothing bandwidth factor that has a central impact on the quality of the probability estimation. Small values lead to under-smoothed PDFs.

On the other hand, large values over-smooth them. As shown in [61], for n observations with the standard deviation σ, the optimal choice can be obtained as:

$$h = 1.06\ \sigma n^{-0.2}. \tag{22}$$

Importantly, in the nonparametric procedure, firstly, the distribution of the TEWMA statistic in (18) is estimated via univariate KDE using anomalies faults-free data. Then, the nonparametric threshold of the mechanism is defined as the $(1 - \alpha)$-th quantile of the estimated distribution. An anomaly is flagged when the TEWMA statistic exceeds the decision threshold.

3.5. Dataset Analysis

The data used in this work has been gathered using the above-described CDER 9.54 kWp PV system (Figure 1). We consider here one month of measurement recorded with a cadence of 10 min during the normal operation of the monitored system. Precisely, the dataset is divided into training data that includes the free first week and used principally during the learning phase; and testing data (week four) which will be applied in the assessment of our proposed models. The provided measurements concern the following variables: solar irradiance, ambient temperature, cell temperature, DC power, DC current, DC voltage, AC power, AC current, and AC voltage. Figure 6 shows five days from the considered data.

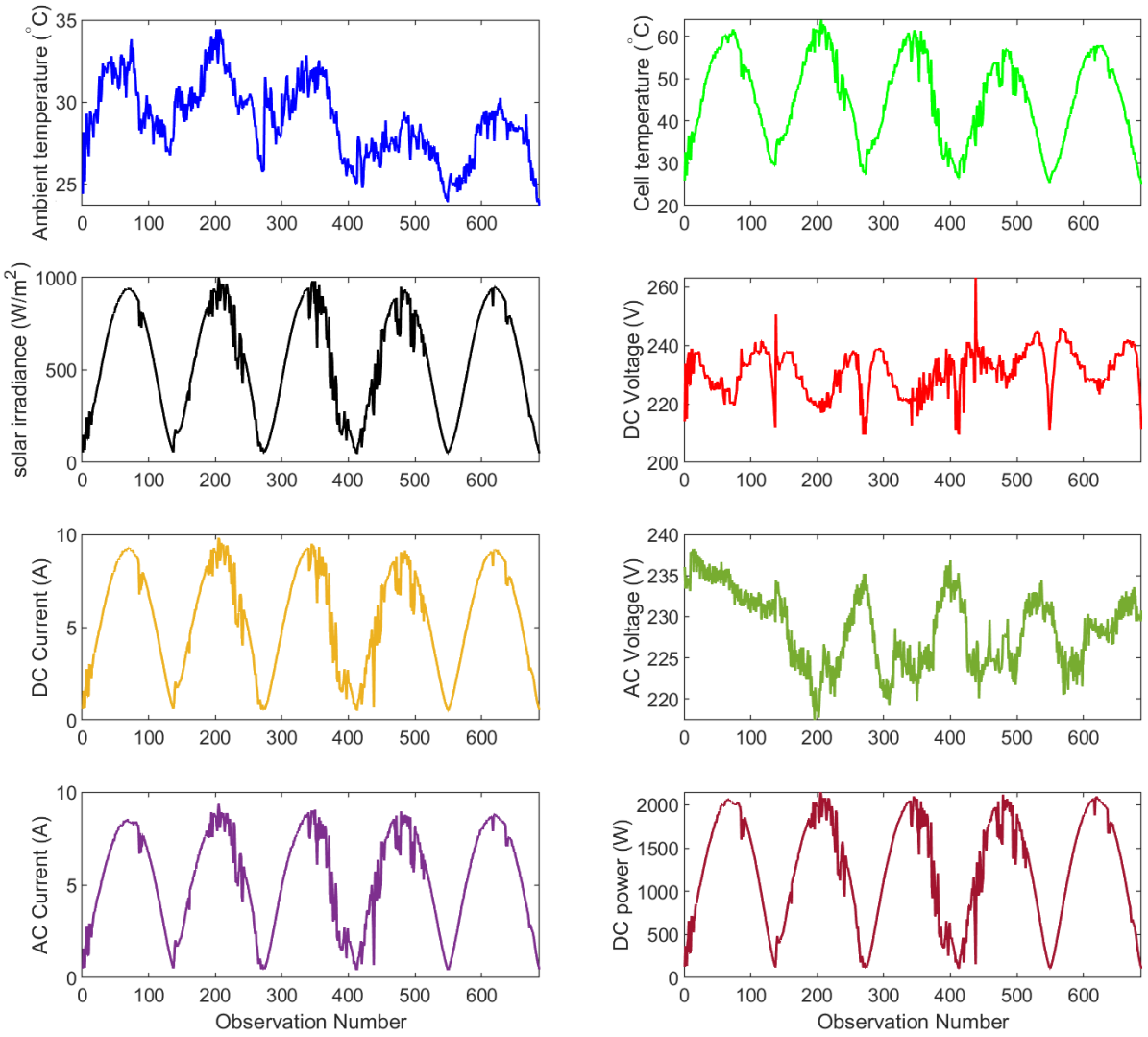


Figure 6. Five days of the studied PV data.

Figure 7 depicts the estimated (PDF) of training data based on the KDE approach. We observe that the considered variables have non-Gaussian distribution. Thus, applying the conventional monitoring charts (e.g., TEWMA) is not appropriate as the assumption of normality is violated. Hence, using an appropriate model to capture the dynamic in this data and generate uncorrelated residuals is needed.

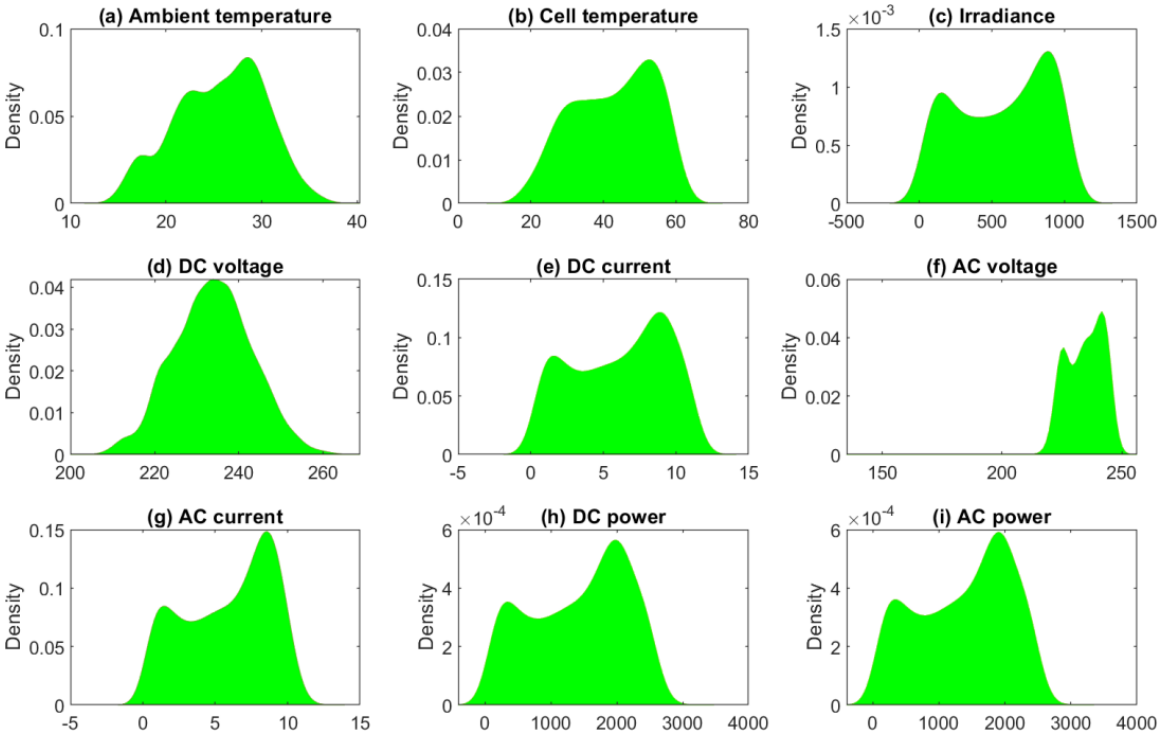


Figure 7. Distribution of the considered variables.

Here, the correlation between the nine variables is quantified using the Pearson correlation heatmap, as illustrated in Figure 8. We can conclude from Figure 8 that DC current, DC power, AC current, and AC power are highly correlated. Moreover, the cell temperature is strongly related to DC current, DC power, AC current, and AC power. Negative correlation between DC voltage and cell temperature. The AC voltage is totally uncorrelated with the rest of the variables. The irradiance, cell temperature, current DC, AC current, power DC, and power AC present a high correlation between them. Finally, a low correlation is reported between the DC voltage and the ambient cell temperature, the DC current, AC current, DC power, and AC power.

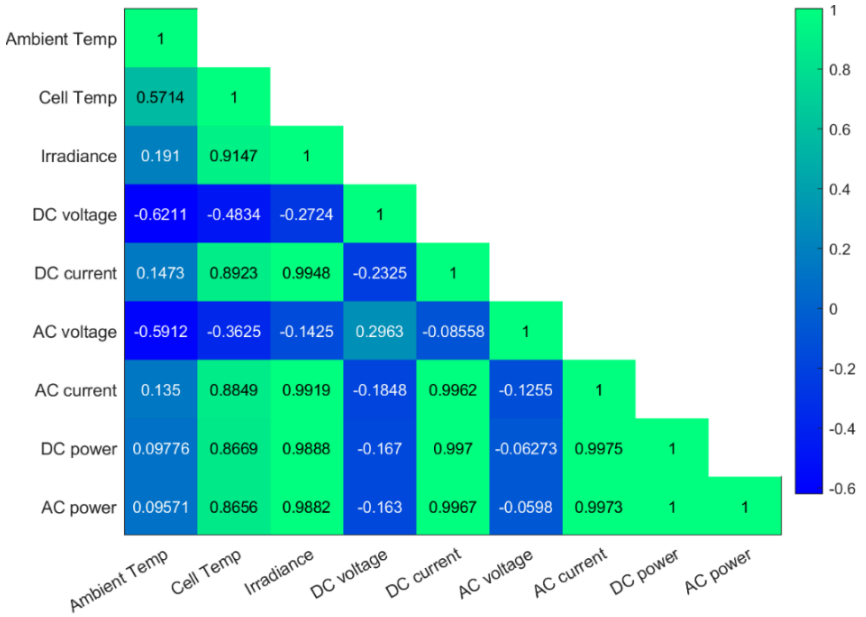


Figure 8. Correlation matrix of training PV data.

In PCR and PLS models, the number of selected **PCs** can strongly affect the quality of the constructed prediction models. We have applied the cumulative percentage variance (CPV) method to determine the suitable number of CPs. As depicted in Figure 9, the obtained results show that five and four CPs are needed to describe 99.87% of the variability in X for both PLS (Figure 9a) and 99.98 for PCR (Figure 9b) models, respectively.

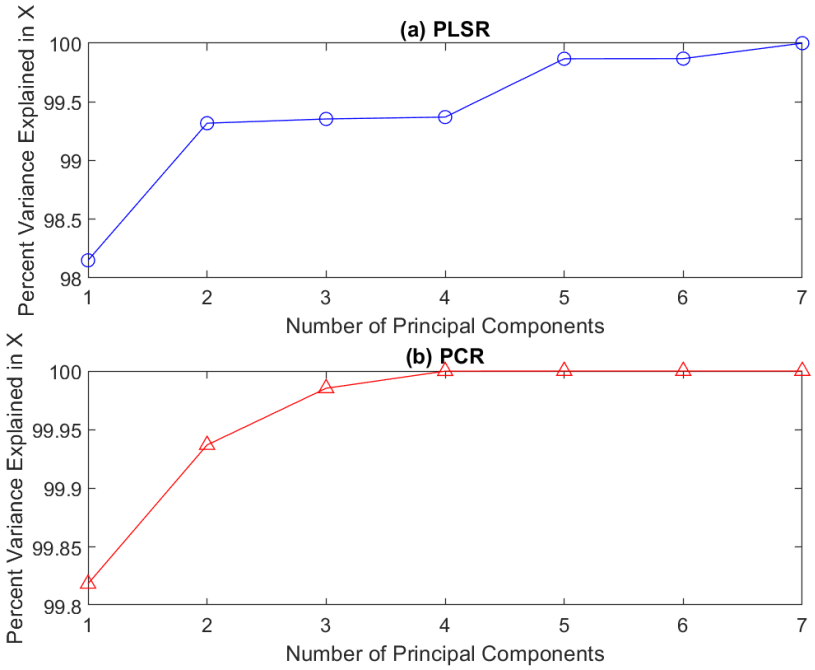


Figure 9. Correlation matrix of training data.

Figure 10 presents the scatter plot of the measured and predicted DC power using PLS and PCR models based on testing data. Overall, results indicate that both models have allowed good prediction, and relatively PLS provided highest prediction quality.

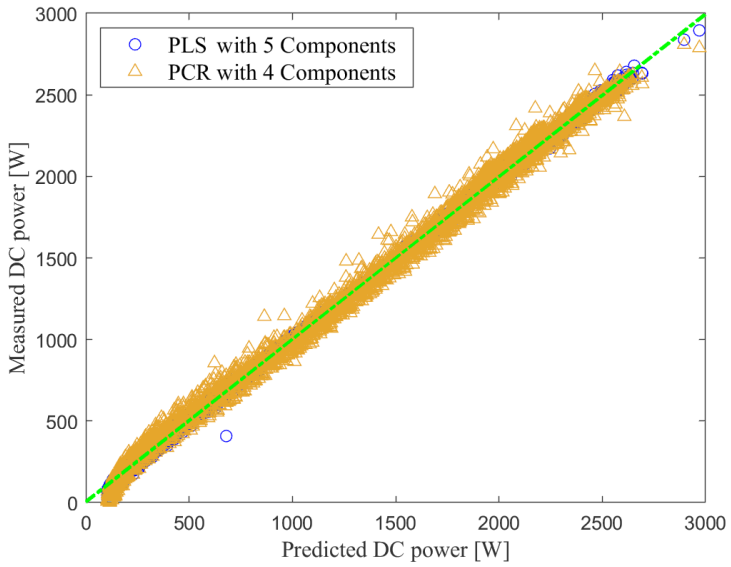


Figure 10. Scatter plot of the measured and predicted DC power using PLS and PCR models.

Figure 11 shows the boxplots of residual errors of PLS and PCR models obtained with training data. As illustrated, best prediction results are obtained with PLS model.

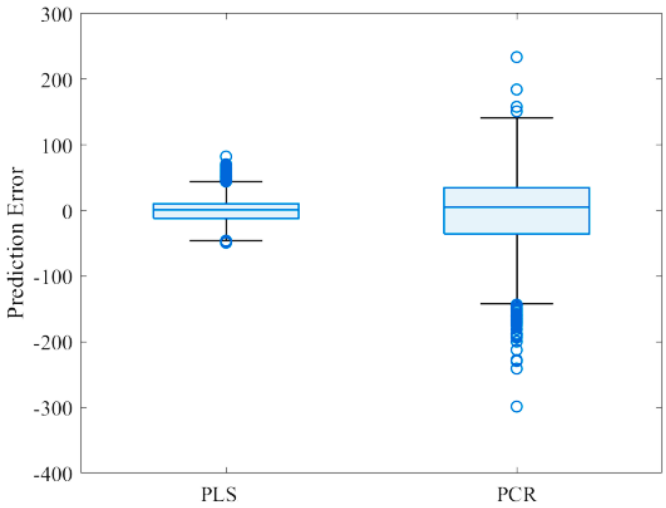


Figure 11. Box plot of residual errors of PLS and PCR models.

In fact, the PLS model returned the smallest residual prediction errors, where most values were close to zero, resulting in the compacter box compared to PCR.

4. The LVR-TEWMA-Based Fault Detection in PV Systems

This section presents the key concept of the proposed fault detection approach for PV systems monitoring. In this work, at first, latent variable regression methods (i.e., PCR and PLS) are calibrated using normal operating data from the inspected PV system. Then, the calibrated models are employed to sense any anomalies in new data via the TEWMA statistical monitoring charts. The deviation (called residuals) between the measured data and the prediction from the constructed LVR models (i.e., PCR and PLS) delivers pertinent information and can be evaluated to uncover anomalies in the PV system. Reasonably, when the inspected PV system is operating customarily, the residuals fluctuate around zero because of noise measurements. On the other hand, if an atypical event or an anomaly occurs in the PV system, the residuals diverge from zero to reveal that the PV system requires attention. Figure 12 schematically recapitulates the proposed LVR-TEWMA anomaly detection framework.

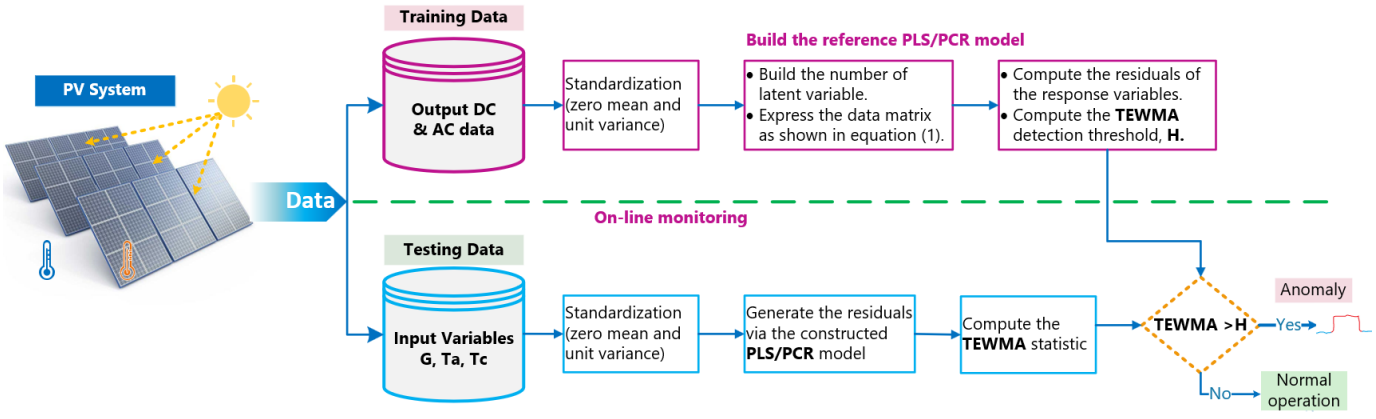


Figure 12. General fault detection procedure in PV system using LVR-based KDE triple exponentially smoothing driven monitoring schemes.

Succinctly, the PLS residuals are expressed as follows:

$$R = Y - \hat{Y} \tag{23}$$

After that, we applied the TEWMA chart to evaluate the residuals for anomaly detection purposes. The LVR-TEWMA anomaly detection strategy can be mainly split into two phases. (1) During the model construction phase, we calibrate the LVR models and compute the TEWMA decision threshold based on normal operating data. (2) In the monitoring phase, we generate the residuals via the constructed LVR models and apply the already computed TEWMA threshold for monitoring new data. Here, we applied the KDE approach to determine the reference threshold. To do so, we estimate the PDF of the TEWMA measurements through KDE. Next, the detection threshold of the LVR-TEWMA approach is obtained as the $(1-\alpha)$-th quantile of the estimated PDF in respect of targeted false alarm probability α.

In summary, after building the LVR models and computing the TEWMA decision threshold, the new online data can be evaluated to sense potential anomalies. Specifically, the TEWMA charting statistic is compared with the detection threshold, H, to decide about the presence of an anomaly as follows:

$$\delta = \begin{cases} No\ anomaly & if\ TEWMA < H, \\ Anomaly & if\ TEWMA > H. \end{cases} \tag{24}$$

5. Results

This section verifies the performance of the proposed LVR-TWEMA approaches in detecting several types of anomalies in PV systems that are recorded during real plants' lifetime. Here, we focus our study on the DC side of PV systems, and we consider here specifically the following scenarios: String faults (SF), Inverter Disconnections (ID), Scenarios with Circuit Breaker Faults, Short-Circuit Faults (SCF), and Sensor Faults in Pyranometer (SFP). Different scenarios are carried out using real measurements gathered using the CDER 9.54 kWp PV system described in Section 2. The normal and faulty data were selected in the summer when most days are characterized by clear skies and medium ambient air temperature (which ranges from 20 to 35 °C). We employed five statistical metrics for the evaluation: true positive rate (TPR), false positive rate (FPR), accuracy, the area under curve (AUC), and EER (equal error rate) [62].

5.1. Scenarios with String Faults

The aim of this experiment is to assess the capacity of LVR-TWMA approaches to uncover the string's fault (i.e., open circuits). Broadly speaking, Open circuits are generally related to the degradation of DC protection or detachment between PV modules in series. Here, we deactivate the circuit breaker to voluntarily create the desired open circuit faults. Figure 13 presents the corresponding detection results using the PLS- and PCR-based TEWMA methods. Both methods detect this fault with large magnitude. Table 3 quantitatively confirms the obtained results in terms of the five scores. Both methods reach a high detection rate, with no false alarms and an accuracy of 0.9942.

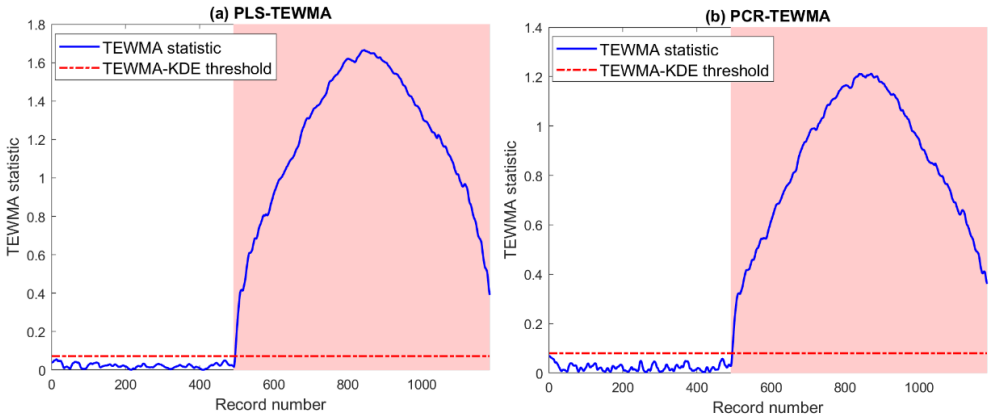


Figure 13. Detection results of (**a**) PLS and (**b**) PCR-based TEWMA charts in the presence of string fault.

Table 3. Results of PLS-based and PCR-based TEWMA charts under string fault.

Method	TPR	FPR	Accuracy	AUC	EER
PLS-TEWMA	0.98	0	0.9942	0.99	0.0034
PCR-TEWMA	0.98	0	0.9942	0.99	0.0034

5.2. Scenarios with Inverter Disconnections

We now investigate the monitoring efficiency of the LVR-TEWMA methods in the presence of inverter disconnections in the inspected PV system. In this case, the PV system will still shut down till the inverter is reconnected. Here, we consider inverter disconnections resulted due to grid instability. More specifically, inverter disconnections refer to a situation where the voltage and frequency of the grid exceed the inverter operating limits. Inverter disconnections are short in time and have large amplitude. In this case study, we used real data with inverter disconnections to study the performance of the proposed monitoring techniques.

Figure 14 depicts the results of the PLS-TWEMA and PCR-TEWMA schemes. We observe that the two schemes successfully flagged the occurred inverter disconnections; these anomalies are with large amplitude. We also notice that these anomalies are within short periods, which can be very helpful in distinguishing them from other abnormal conditions, such as temporary shading and string faults. Table 4 summarizes the results of this comparison using the five statistical metrics. Results show that the PLS-TEWMA approach outperformed the PCR-TEWMA approach in detecting all occurred faults (i.e., TPR = 100%).

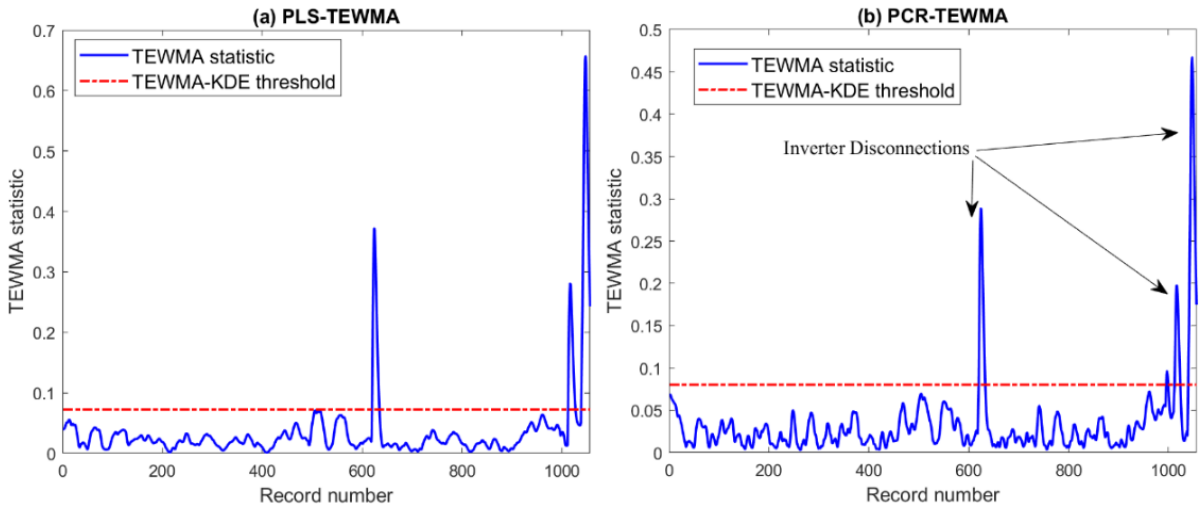


Figure 14. Detection results of (**a**) PLS- and (**b**) PCR-based TEWMA charts in the presence of inverter disconnections.

Table 4. Results of PLS- and PCR-based TEWMA charts under inverter disconnections.

Method	TPR	FPR	Accuracy	AUC	EER
PLS-TEWMA	1	0.0418	0.9583	0.9791	0.0417
PCR-TEWMA	0.75	0.0399	0.9593	0.8550	0.0407

5.3. Scenario with Circuit Breaker Faults

Now, we assess the performances of the proposed methodologies under circuit breaker faults. These components are principally designed to protect PV systems and users from eventual shocks, short circuits, and potential overloads. Figure 15 and Table 5 illustrate the detection results of the PLS-TEWMA and PCR-TEWMA methods in the presence of RCCB

faults. In overall, it is clear that both PLS and PCR maintain high performance with this type of fault. They reach a high detection rate with a TPR of around 0.98.

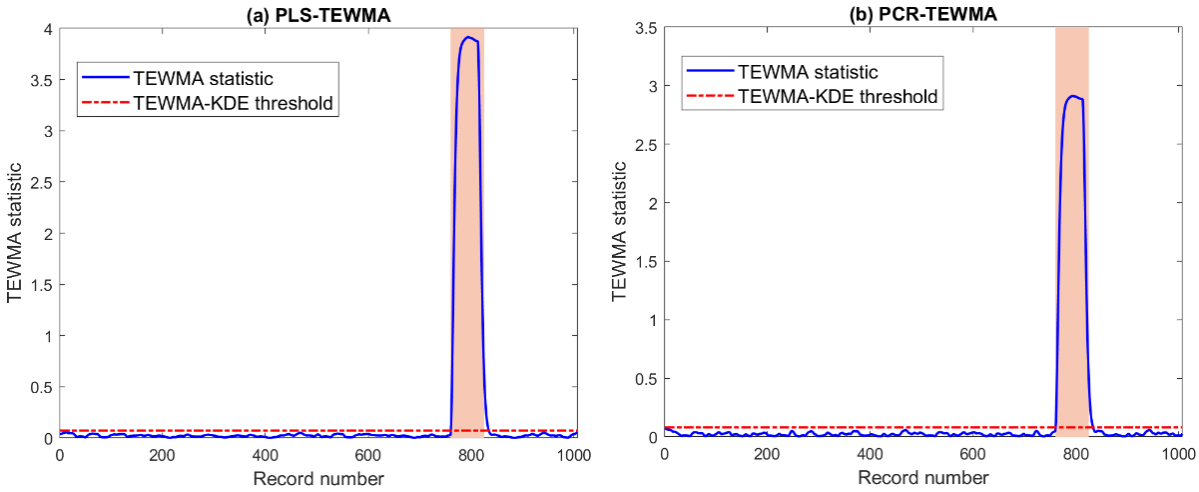


Figure 15. Detection results of (**a**) PLS- and (**b**) PCR-based TEWMA charts in the presence of RCCB faults.

Table 5. Results of PLS- and PCR-based TEWMA charts under RCCB faults.

Method	TPR	FPR	Accuracy	AUC	EER
PLS-TEWMA	0.9815	0.0220	0.9782	0.9797	0.0218
PCR-TEWMA	0.9815	0.0210	0.9791	0.9802	0.0209

5.4. Short-Circuit Fault

In this scenario, the proposed methods are evaluated to detect short-circuit faults. Figure 16 displays the results of the two approaches when two PV modules are short-circuited. We can see that the two approaches can uncover this fault. Table 6 indicates that the PLS-TEWMA has the best performance, with a TPR of 0.8649. On the contrary, the PCR-TEWMA provides unsatisfactory results with high missed detection (TPR = 0.1351). This could be attributed to the prediction accuracy of the PLS model and the sensitivity of the TEWMA chart to the small changes.

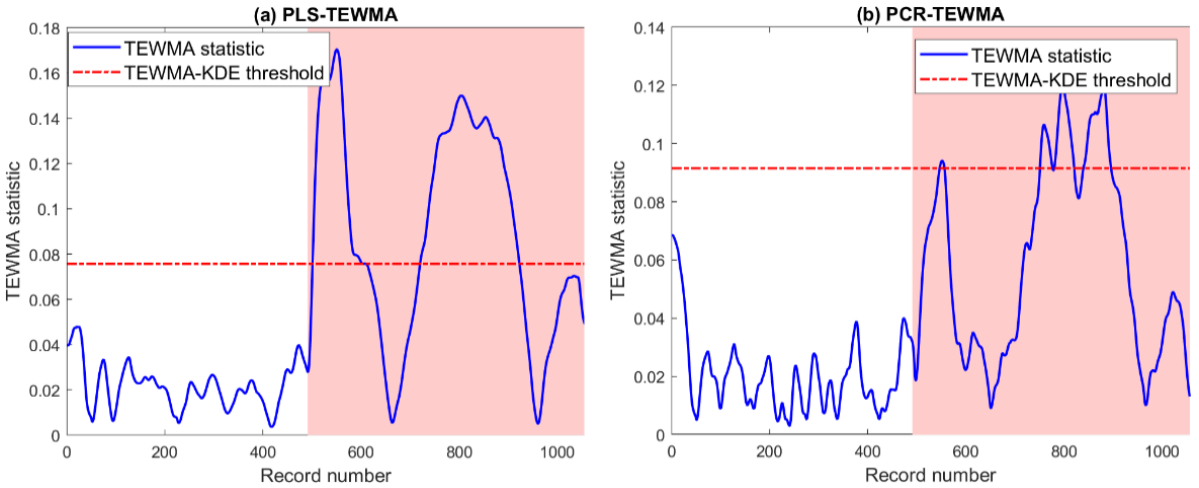


Figure 16. Detection results of (**a**) PLS- and (**b**) PCR-based TEWMA charts in the presence of two short-circuited modules.

Table 6. Results of PLS- and PCR-based TEWMA charts under two modules short-circuited.

Method	TPR	FPR	Accuracy	AUC	EER
PLS-TEWMA	0.8649	0	0.9823	0.9324	0.0095
PCR-TEWMA	0.1351	0	0.8867	0.5676	0.0606

5.5. Sensor Bias Faults in the Pyranometer

We now consider sensor faults in pyranometers that could be due to degradation, aging, or other factors. Importantly, the pyranometer adapts the PV modules' direction to optimally track the sun's movements and be continuously highly exposed to sunlight. Consequently, faults affecting this precious component can prevent the PV system from being in the right orientation, which can significantly decrease the produced energy. Here, we simulate bias faults with different magnitudes to assess the sensitivity of the two proposed approaches. The simulated bias, b, begins at sample 200. We used the following model to simulate bias faults in the solar irradiance measurements.

$$y_i = x_i + b, \tag{25}$$

where y_i and x_i are the faulty and fault-free measurements, respectively. Here, as part of the total variation in solar irradiance measurements, we consider the following magnitudes: 5%, 10%, 20%, 30%, 40%, and 50%.

Detection results of the PLS-TEWMA and PCR-TEWMA charts in the case of bias faults (with 10%) in the pyranometer measurements are depicted in Figure 17. Visually, Figure 15 shows that both charts can recognize this bias fault. A comparison between the two charts under the presence of bias faults with different magnitudes is given in Table 7.

Detection results of the PLS-TEWMA and PCR-TEWMA charts in the case of bias faults (with 10%) in the pyranometer measurements are depicted in Figure 17. Visually, Figure 17 shows that both charts can recognize this bias fault. A comparison between the two charts under the presence of bias faults with different magnitudes is given in Table 8. Results in Table 8 confirm the superiority of the PLS-TEWMA scheme compared to the PCR-TEWMA, particularly for faults with small levels. For instance, for bias faults with 5%, the PLS-TEMWA scheme detects this sensor fault with a TPR of 0.9231, while the PCR-TEWMA scheme obtains a TPR of 0.7802.

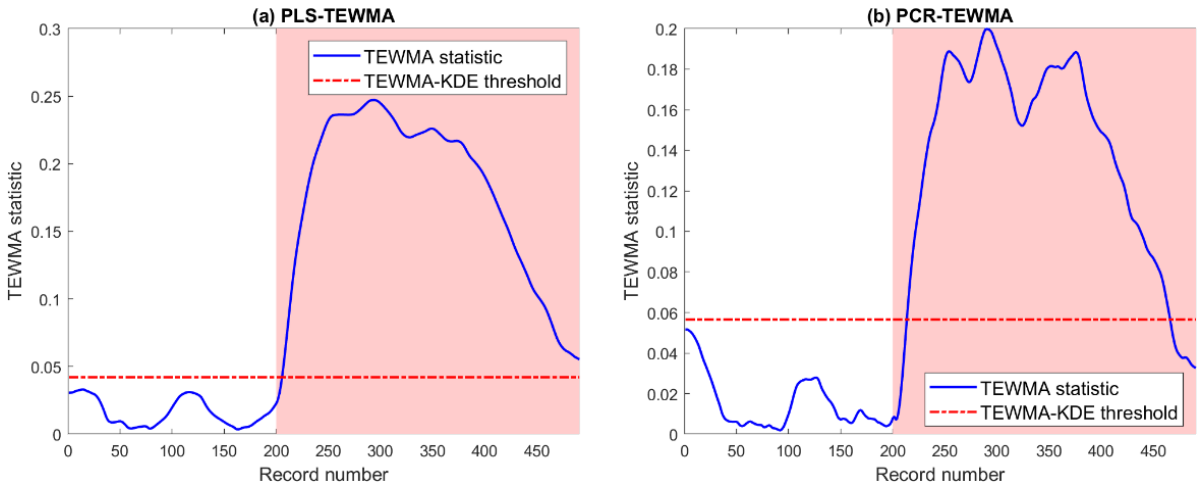


Figure 17. Detection results of (**a**) PLS- and (**b**) PCR-based TEWMA charts in the presence of sensor bias fault in the pyranometer measurements (a bias of 10% of the total variation in solar irradiance measurements).

Table 7. Results of PLS- and PCR-based TEWMA charts under bias sensors fault in pyranometer.

Bias Sensor (B)	TEWMA Method	TPR	FPR	Accuracy	AUC	EER
50%	PLS	0.9780	0	0.9931	0.9890	0.0041
	PCR	0.9451	0	0.9828	0.9725	0.0102
40%	PLS	0.9670	0	0.9897	0.9835	0.0061
	PCR	0.9341	0	0.9794	0.9670	0.0122
30%	PLS	0.9670	0	0.9897	0.9835	0.0061
	PCR	0.9231	0	0.9759	0.9615	0.0143
20%	PLS	0.9560	0	0.9863	0.9780	0.0081
	PCR	0.9011	0	0.9691	0.9505	0.0183
10%	PLS	0.9451	0	0.9828	0.9725	0.0102
	PCR	0.8571	0	0.9553	0.9286	0.0265
5%	PLS	0.9231	0	0.9759	0.9615	0.0143
	PCR	0.7802	0	0.9313	0.8901	0.0407

Table 8. Results of PLS- and PCR-based DEWMA charts under bias sensors fault in pyranometer.

Bias Sensor (B)	DEWMA Method	TPR	FPR	Accuracy	AUC	EER
50%	PLS	0.9622	0	0.9776	0.9811	0.0224
	PCR	0.9553	0	0.9735	0.9777	0.0265
40%	PLS	0.9588	0	0.9756	0.9794	0.0244
	PCR	0.9313	0	0.9593	0.9656	0.0407
30%	PLS	0.9313	0	0.9593	0.9656	0.0407
	PCR	0.9038	0	0.9430	0.9519	0.0570
20%	PLS	0.9003	0	0.9409	0.9502	0.0591
	PCR	0.8729	0	0.9246	0.9364	0.0754
10%	PLS	0.7388	0	0.8452	0.8694	0.1548
	PCR	0.7457	0	0.8493	0.8729	0.1507
5%	PLS	0.7216	0	0.8350	0.8608	0.1650
	PCR	0.6976	0	0.8208	0.8488	0.1792

Figure 18 depicts the AUC values of the two charts for the considered faults. The main conclusion from Figure 18 is that when the amplitude of the occurred fault is small, the PLS-TEWMA scheme is better than the PCR-TEWMA scheme. On the other hand, for a large fault, the two schemes are comparable.

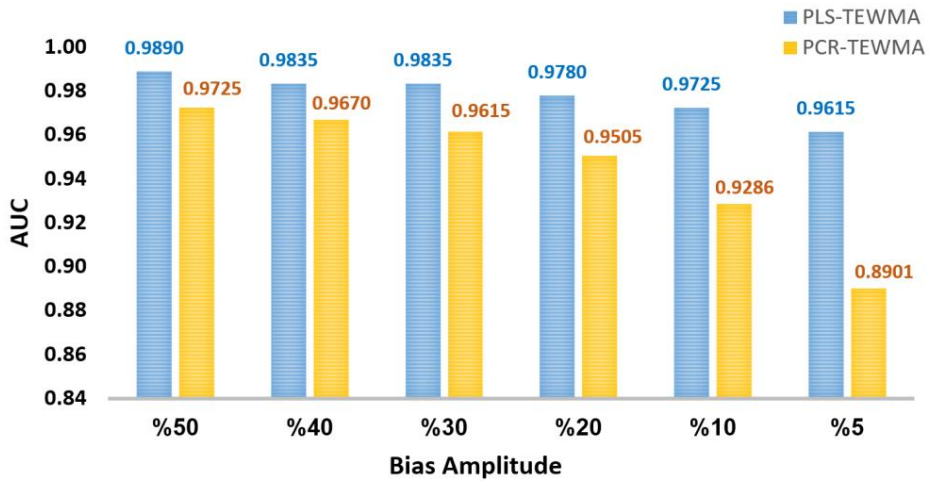


Figure 18. AUC values by PLS- and PCR-based TEWMA charts for different bias magnitudes in the pyranometer.

For comparison purposes, we applied PLS- and PCR-based double EWMA (DEWMA) charts to detect bias sensors fault in the pyranometer (Table 8). Here, we also used the KDE approach to estimate the decision thresholds for the two charts, which enables more detection flexibility. From Table 8, we observe that for abnormal changes with large magnitudes, the two approaches (i.e., PLS-DEWMA and PCR-DEWMA) achieved relatively comparable performance, while for small changes, the PLS-DEWM chart outperformed the PCR-DEMWA approach.

From Tables 7 and 8, we observe that the PLS- and PCR-based TEWMA charts showed improved detection ability compared to the PLS- and PCR-based DEWMA chart, particularly for small faults. At the same time, they are relatively comparable for moderate and large faults. Overall, the proposed PLS-TEWMA approach is more effective in detecting small faults compared to the other charts. This is mainly due to the capacity of the TEWMA chart to sense small changes in data. However, the results revealed that the PLS-TEWMA approach can sense small changes but with some missed detection (TPR = 0.9231) (Table 7). This could be attributed to the used PLS model, which is a linear model, that is not suited to capture process nonlinearity. Thus, we recommend adopting nonlinear models, such as nonlinear PLS, to enhance the detection capability of the proposed approach.

6. Conclusions

This study is within a data-driven anomaly detection framework for PV systems monitoring. An efficient strategy combining latent variable regression methods (i.e., PCR and PLS) with the TEWMA monitoring chart has been developed to detect anomalies in the DC sides of PV systems. Specifically, LVR methods have been built using training data and then used to generate residuals, which are used as anomaly indicators. Here, the TEWMA chart is applied to residuals for fault detection. The decision threshold in the TEWMA chart is determined using a KDE-based approach to allow more flexibility and improve detection accuracy. Several real anomalies and simulated sensor faults have been considered to study the performance of the proposed approaches. Results revealed that the PLS-TEWMA chart provided superior detection performance compared to that of the PCR-TEWMA chart and PLS- and PCR-based DEWMA charts, particularly for small faults.

In future work, we plan to improve the proposed technique by incorporating machine learning-based classification methods to discriminate between the detected anomalies. Another direction of improvement consists of amalgamating the benefit of deep-learning models [45,63] and the detection capacity of the TEWMA to sense small changes. In addition, it will be interesting to investigate the detection capability of this approach in uncovering other types of sensor anomalies, such as sensor degradation anomalies, sensor-freezing anomalies, and multiple sensor anomalies that affect different sensors (e.g., pyranometers and thermocouples) simultaneously. Data from PV plants is tainted with noise, which can degrade the detection performance by increasing the number of false alarms rate in the case of a low signal-to-noise ratio. We also plan in future work to construct a robust detection approach that combines wavelet multiscale presentation with the TEWMA chart.

Author Contributions: B.B.: Conceptualization, formal analysis, investigation, methodology, software, writing—original draft, and writing—review and editing. F.H.: Conceptualization, formal analysis, investigation, methodology, software, supervision, writing—original draft, and writing—review and editing. B.T.: Formal analysis, software, funding acquisition, writing—original draft, and writing—review and editing. Y.S.: Investigation, conceptualization, formal analysis, methodology, writing—review and editing, and supervision. A.H.A.: project administration, funding acquisition, and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by funding from the King Abdullah University of Science and Technology (KAUST), Office of Sponsored Research (OSR), under Award No: OSR-2019-CRG7-3800, and from the Centre de Développement des Energies Renouvelables (CDER), Direction Générale de la Recherche Scientifique et du Développement Technologique (DGRSDT).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Ahmed, W.; Sheikh, J.A.; Farjana, S.H.; Mahmud, M.A.P. Defects Impact on PV System GHG Mitigation Potential and Climate Change. *Sustainability* **2021**, *13*, 7793. [CrossRef]
2. IEA PVPS. Trends in Photovoltaic Applications. 2021. Available online: https://tecsol.blogs.com/files/iea-pvps-trends-report-2021-1.pdf (accessed on 23 October 2022).
3. Garoudja, E.; Harrou, F.; Sun, Y.; Kara, K.; Chouder, A.; Silvestre, S. Statistical fault detection in photovoltaic systems. *Sol. Energy* **2017**, *150*, 485–499. [CrossRef]
4. Alam, M.K.; Khan, F.; Johnson, J.; Flicker, J. A comprehensive review of catastrophic faults in PV arrays: Types, detection, and mitigation techniques. *IEEE J. Photovolt.* **2015**, *5*, 982–997. [CrossRef]
5. Hare, J.; Shi, X.; Gupta, S.; Bazzi, A. Fault diagnostics in smart micro-grids: A survey. *Renew. Sustain. Energy Rev.* **2016**, *60*, 1114–1124. [CrossRef]
6. Ye, J.; Giani, A.; Elasser, A.; Mazumder, S.K.; Farnell, C.; Mantooth, H.A.; Kim, T.; Liu, J.; Chen, B.; Seo, G.S.; et al. A Review of Cyber–Physical Security for Photovoltaic Systems. *IEEE J. Emerg. Sel. Top. Power Electron.* **2021**, *10*, 4879–4901. [CrossRef]
7. Wang, W.; Harrou, F.; Bouyeddou, B.; Senouci, S.M.; Sun, Y. A stacked deep learning approach to cyber-attacks detection in industrial systems: Application to power system and gas pipeline systems. *Clust. Comput.* **2022**, *25*, 561–578. [CrossRef]
8. Wang, W.; Harrou, F.; Bouyeddou, B.; Senouci, S.M.; Sun, Y. Cyber-attacks detection in industrial systems using artificial intelligence-driven methods. *Int. J. Crit. Infrastruct. Prot.* **2022**, *38*, 100542. [CrossRef]
9. Le, M.; Nguyen, D.K.; Dao, V.D.; Vu, N.H.; Vu, H.H.T. Remote anomaly detection and classification of solar photovoltaic modules based on deep neural network. *Sustain. Energy Technol. Assess.* **2021**, *48*, 101545. [CrossRef]
10. Janarthanan, R.; Maheshwari, R.U.; Shukla, P.K.; Shukla, P.K.; Mirjalili, S.; Kumar, M. Intelligent detection of the PV faults based on artificial neural network and type 2 fuzzy systems. *Energies* **2021**, *14*, 6584. [CrossRef]
11. Harrou, F.; Saidi, A.; Sun, Y.; Khadraoui, S. Monitoring of photovoltaic systems using improved kernel-based learning schemes. *IEEE J. Photovolt.* **2021**, *11*, 806–818. [CrossRef]
12. Mellit, A.; Tina, G.M.; Kalogirou, S.A. Fault detection and diagnosis methods for photovoltaic systems: A review. *Renew. Sustain. Energy Rev.* **2018**, *91*, 1–17. [CrossRef]
13. Madeti, S.R.; Singh, S.N. Modeling of PV system based on experimental data for fault detection using kNN method. *Sol. Energy* **2018**, *173*, 139–151. [CrossRef]
14. Badr, M.M.; Hamad, M.S.; Abdel-Khalik, A.S.; Hamdy, R.A.; Ahmed, S.; Hamdan, E. Fault Identification of Photovoltaic Array Based on Machine Learning Classifiers. *IEEE Access* **2021**, *9*, 159113–159132. [CrossRef]
15. Benkercha, R.; Moulahoum, S. Fault detection and diagnosis based on C4. 5 decision tree algorithm for grid connected PV system. *Sol. Energy* **2018**, *173*, 610–634. [CrossRef]
16. Harrou, F.; Taghezouit, B.; Sun, Y. Robust and flexible strategy for fault detection in grid-connected photovoltaic systems. *Energy Convers. Manag.* **2019**, *180*, 1153–1166. [CrossRef]
17. 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]
18. Harrou, F.; Sun, Y.; Taghezouit, B.; Saidi, A.; Hamlati, M.E. Reliable fault detection and diagnosis of photovoltaic systems based on statistical monitoring approaches. *Renew. Energy* **2018**, *116*, 22–37. [CrossRef]
19. Huang, C.M.; Chen, S.J.; Yang, S.P. A Parameter Estimation Method for a Photovoltaic Power Generation System Based on a Two-Diode Model. *Energies* **2022**, *15*, 1460. [CrossRef]
20. Kang, B.K.; Kim, S.T.; Bae, S.H.; Park, J.W. Diagnosis of output power lowering in a PV array by using the Kalman-filter algorithm. *IEEE Trans. Energy Convers.* **2012**, *27*, 885–894. [CrossRef]
21. Skomedal, Å.F.; Øgaard, M.B.; Haug, H.; Marstein, E.S. Robust and fast detection of small power losses in large-scale PV systems. *IEEE J. Photovolt.* **2021**, *11*, 819–826. [CrossRef]
22. Spataru, S.; Sera, D.; Kerekes, T.; Teodorescu, R. Diagnostic method for photovoltaic systems based on light I–V measurements. *Sol. Energy* **2015**, *119*, 29–44. [CrossRef]
23. Chouder, A.; Silvestre, S. Automatic supervision and fault detection of PV systems based on power losses analysis. *Energy Convers. Manag.* **2010**, *51*, 1929–1937. [CrossRef]
24. Hou, Z.S.; Wang, Z. From model-based control to data-driven control: Survey, classification and perspective. *Inf. Sci.* **2013**, *235*, 3–35. [CrossRef]
25. Taghezouit, B.; Harrou, F.; Sun, Y.; Arab, A.H.; Larbes, C. Multivariate statistical monitoring of photovoltaic plant operation. *Energy Convers. Manag.* **2020**, *205*, 112317. [CrossRef]

26. Fadhel, S.; Delpha, C.; Diallo, D.; Bahri, I.; Migan, A.; Trabelsi, M.; Mimouni, M.F. PV shading fault detection and classification based on IV curve using principal component analysis: Application to isolated PV system. *Sol. Energy* **2019**, *179*, 1–10. [CrossRef]
27. Amaral, T.G.; Pires, V.F.; Pires, A.J. Fault detection in PV tracking systems using an image processing algorithm based on PCA. *Energies* **2021**, *14*, 7278. [CrossRef]
28. Deceglie, M.G.; Micheli, L.; Muller, M. Quantifying soiling loss directly from PV yield. *IEEE J. Photovolt.* **2018**, *8*, 547–551. [CrossRef]
29. Kiliç, H.; Gumus, B.; Khaki, B.; Yilmaz, M.; Palensky, P.; Authority, P. A Robust Data-Driven Approach for Fault Detection in Photovoltaic Arrays. In Proceedings of the 10th IEEE PES Innovative Smart Grid Technologies Europe, ISGT-Europe 2020, Virtual, 26–28 October 2020.
30. Xiong, Q.; Liu, X.; Feng, X.; Gattozzi, A.L.; Shi, Y.; Zhu, L.; Ji, S.; Hebner, R.E. Arc fault detection and localization in photovoltaic systems using feature distribution maps of parallel capacitor currents. *IEEE J. Photovolt.* **2018**, *8*, 1090–1097. [CrossRef]
31. Kumar, B.P.; Ilango, G.S.; Reddy, M.J.B.; Chilakapati, N. Online fault detection and diagnosis in photovoltaic systems using wavelet packets. *IEEE J. Photovolt.* **2017**, *8*, 257–265. [CrossRef]
32. Edun, A.S.; Kingston, S.; LaFlamme, C.; Benoit, E.; Scarpulla, M.A.; Furse, C.M.; Harley, J.B. Detection and localization of disconnections in a large-scale string of photovoltaics using SSTDR. *IEEE J. Photovolt.* **2021**, *11*, 1097–1104. [CrossRef]
33. Wang, M.H.; Lin, Z.H.; Lu, S.D. A Fault Detection Method Based on CNN and Symmetrized Dot Pattern for PV Modules. *Energies* **2022**, *15*, 6449. [CrossRef]
34. Cui, F.; Tu, Y.; Gao, W. A Photovoltaic System Fault Identification Method Based on Improved Deep Residual Shrinkage Networks. *Energies* **2022**, *15*, 3961. [CrossRef]
35. Harrou, F.; Taghezouit, B.; Sun, Y. Improved kNN-based monitoring schemes for detecting faults in PV systems. *IEEE J. Photovolt.* **2019**, *9*, 811–821. [CrossRef]
36. Harrou, F.; Taghezouit, B.; Khadraoui, S.; Dairi, A.; Sun, Y.; Hadj Arab, A. Ensemble Learning Techniques-Based Monitoring Charts for Fault Detection in Photovoltaic Systems. *Energies* **2022**, *15*, 6716. [CrossRef]
37. Karatepe, E.; Hiyama, T. Controlling of artificial neural network for fault diagnosis of photovoltaic array. In Proceedings of the 2011 16th International Conference on Intelligent System Applications to Power Systems, Hersonissos, Greece, 25–28 September 2011; pp. 1–6.
38. Chen, S.Q.; Yang, G.J.; Gao, W.; Guo, M.F. Photovoltaic fault diagnosis via semisupervised ladder network with string voltage and current measures. *IEEE J. Photovolt.* **2020**, *11*, 219–231. [CrossRef]
39. Gaggero, G.B.; Rossi, M.; Girdinio, P.; Marchese, M. Detecting System Fault/Cyberattack within a Photovoltaic System Connected to the Grid: A Neural Network-Based Solution. *J. Sens. Actuator Netw.* **2020**, *9*, 20. [CrossRef]
40. Parvez, I.; Aghili, M.; Sarwat, A.I.; Rahman, S.; Alam, F. Online power quality disturbance detection by support vector machine in smart meter. *J. Mod. Power Syst. Clean Energy* **2019**, *7*, 1328–1339. [CrossRef]
41. Madakyaru, M.; Harrou, F.; Sun, Y. Monitoring distillation column systems using improved nonlinear partial least squares-based strategies. *IEEE Sens. J.* **2019**, *19*, 11697–11705. [CrossRef]
42. Kourti, T.; MacGregor, J.F. Process analysis, monitoring and diagnosis, using multivariate projection methods. *Chemom. Intell. Lab. Syst.* **1995**, *28*, 3–21. [CrossRef]
43. Geladi, P.; Kowalski, B.R. Partial least-squares regression: A tutorial. *Anal. Chim. Acta* **1986**, *185*, 1–17. [CrossRef]
44. Harrou, F.; Sun, Y.; Madakyaru, M. Kullback-leibler distance-based enhanced detection of incipient anomalies. *J. Loss Prev. Process Ind.* **2016**, *44*, 73–87. [CrossRef]
45. Harrou, F.; Sun, Y.; Hering, A.S.; Madakyaru, M.; Dairi, A. *Linear Latent Variable Regression (LVR)-Based Process Monitoring*; Elsevier BV: Amsterdam, The Netherlands, 2021.
46. Li, B.; Morris, J.; Martin, E.B. Model selection for partial least squares regression. *Chemom. Intell. Lab. Syst.* **2002**, *64*, 79–89. [CrossRef]
47. MacGregor, J.F.; Kourti, T. Statistical process control of multivariate processes. *Control Eng. Pract.* **1995**, *3*, 403–414. [CrossRef]
48. Madakyaru, M.; Harrou, F.; Sun, Y. Improved data-based fault detection strategy and application to distillation columns. *Process Saf. Environ. Prot.* **2017**, *107*, 22–34. [CrossRef]
49. Wang, X.; Kruger, U.; Lennox, B. Recursive partial least squares algorithms for monitoring complex industrial processes. *Control. Eng. Pract.* **2003**, *11*, 613–632. [CrossRef]
50. Ahn, S.J.; Lee, C.J.; Jung, Y.; Han, C.; Yoon, E.S.; Lee, G. Fault diagnosis of the multi-stage flash desalination process based on signed digraph and dynamic partial least square. *Desalination* **2008**, *228*, 68–83. [CrossRef]
51. Bouyeddou, B.; Harrou, F.; Saidi, A.; Sun, Y. An Effective Wind Power Prediction using Latent Regression Models. In Proceedings of the 2021 International Conference on ICT for Smart Society (ICISS), Bandung, Indonesia, 2–4 August 2021; pp. 1–6.
52. Frank, L.E.; Friedman, J.H. A statistical view of some chemometrics regression tools. *Technometrics* **1993**, *35*, 109–135. [CrossRef]
53. Harrou, F.; Nounou, M.N.; Nounou, H.N. Detecting abnormal ozone levels using PCA-based GLR hypothesis testing. In Proceedings of the 2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), Singapore, 16–19 April 2013; pp. 95–102.
54. Alevizakos, V.; Chatterjee, K.; Koukouvinos, C. The triple exponentially weighted moving average control chart. *Qual. Technol. Quant. Manag.* **2021**, *18*, 326–354. [CrossRef]

55. Alevizakos, V.; Chatterjee, K.; Koukouvinos, C. A nonparametric triple exponentially weighted moving average sign control chart. *Qual. Reliab. Eng. Int.* **2021**, *37*, 1504–1523. [CrossRef]
56. Mahmoud, M.A.; Woodall, W.H. An evaluation of the double exponentially weighted moving average control chart. *Commun. Stat.—Simul. Comput.* **2010**, *39*, 933–949. [CrossRef]
57. Zhang, L.; Chen, G. An extended EWMA mean chart. *Qual. Technol. Quant. Manag.* **2005**, *2*, 39–52. [CrossRef]
58. Zhang, L.; Bebbington, M.S.; Govindaraju, K.; Lai, C.D. Composite EWMA control charts. *Commun. Stat.-Simul. Comput.* **2004**, *33*, 1133–1158. [CrossRef]
59. Martin, E.; Morris, A. Non-parametric confidence bounds for process performance monitoring charts. *J. Proc. Control* **1996**, *6*, 349–358. [CrossRef]
60. Chen, Y.C. A tutorial on kernel density estimation and recent advances. *Biostat. Epidemiol.* **2017**, *1*, 161–187. [CrossRef]
61. Mugdadi, A.R.; Ahmad, I.A. A bandwidth selection for kernel density estimation of functions of random variables. *Comput. Stat. Data Anal.* **2004**, *47*, 49–62. [CrossRef]
62. Harrou, F.; Khaldi, B.; Sun, Y.; Cherif, F. An efficient statistical strategy to monitor a robot swarm. *IEEE Sens. J.* **2019**, *20*, 2214–2223. [CrossRef]
63. Dairi, A.; Harrou, F.; Sun, Y.; Khadraoui, S. Short-term forecasting of photovoltaic solar power production using variational auto-encoder driven deep learning approach. *Appl. Sci.* **2020**, *10*, 8400. [CrossRef]