Research Article

Multivariate Statistics and Supervised Learning for Predictive Detection of Unintentional Islanding in Grid-Tied Solar PV Systems

Shashank Vyas,¹ Rajesh Kumar,¹ and Rajesh Kavasseri²

¹Centre for Energy and Environment, Malaviya National Institute of Technology, Jawaharlal Nehru Marg, Jaipur 302017, India

²Department of Electrical and Computer Engineering, North Dakota State University, 1340 Administration Avenue, Fargo, ND 58102, USA

Correspondence should be addressed to Rajesh Kumar; rkumar.ee@gmail.com

Received 28 October 2015; Accepted 6 March 2016

Academic Editor: Francesco Carlo Morabito

Copyright © 2016 Shashank Vyas et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Integration of solar photovoltaic (PV) generation with power distribution networks leads to many operational challenges and complexities. Unintentional islanding is one of them which is of rising concern given the steady increase in grid-connected PV power. This paper builds up on an exploratory study of unintentional islanding on a modeled radial feeder having large PV penetration. Dynamic simulations, also run in real time, resulted in exploration of unique potential causes of creation of accidental islands. The resulting voltage and current data underwent dimensionality reduction using principal component analysis (PCA) which formed the basis for the application of Q statistic control charts for detecting the anomalous currents that could island the system. For reducing the false alarm rate of anomaly detection, Kullback-Leibler (K-L) divergence was applied on the principal component projections which concluded that Q statistic based approach alone is not reliable for detection of the symptoms liable to cause unintentional islanding. The obtained data was labeled and a K-nearest neighbor (K-NN) binomial classifier was then trained for identification and classification of potential islanding precursors from other power system transients. The three-phase short-circuit fault case was successfully identified as statistically different from islanding symptoms.

1. Introduction

The distribution segment of the electricity supply network is always under stress given the variable consumption patterns in complex geographical spread. Since the power grid was not designed for the inclusion of sources in its distribution pathway, the integration of distributed energy resources is a technically complex issue. When the sources are renewable-energy based, their stochastic nature and the variability in availability of generation induce extra randomness in the system operation. The distributed generators (DGs) challenge the conventional functioning of the distribution feeders and lead to operational issues affecting the power quality, stability, and protection aspects [1].

Since most of the dispersed solar PV integration is taking place on the distribution side [2], especially in the form of rooftop systems, the vulnerability of such a system to supply

interruptions and shutdowns becomes an important concern. Interconnecting a solar PV system thus becomes challenging; however, the continued feeding of loads in the vicinity of the PV when the mains suddenly go off is a situation that must be avoided.

Most of the DGs including PV inverters operate in constant PQ control mode or constant power (active and reactive) control mode. This means they are commanded to give output power in synchronism with the grid based on the provided power set points. PV is generally operated at unity power factor [3] because this is a strategy of maximizing the energy yield from the array through maximum power point tracking (MPPT). The inverter thus cannot adjust its active or reactive power output accordingly to regulate the grid frequency and voltage.

Islanding is said to occur when the DG continues providing power (with utility level voltage and frequency) to

FIGURE 1: Creation of a power island.

a segment containing certain loads even after that portion of the network, including the point of common coupling (PCC), gets disconnected from the main power system. This is clear from a schematic diagram given in Figure 1. Based on the reasons for disconnection, islanding can be categorized into intentional or unintentional. When the distribution system operator knowingly separates certain sections from the main grid with an intention to secure critical loads, this practice is called intentional islanding. The move is preplanned and is generally required in situations of network congestion or a large power system blackout. On the other hand, accidental or unplanned disconnection of a loads-DG portion from the mains and continuation of the DG in grid-connected mode of operation, maintaining grid level voltage and frequency, are known as unintentional islanding. The creation and maintenance of an unplanned island are harmful to system health. Basically, such an islanded network operates as an independent autonomous entity without the regulation from the grid. Since the DG continues to operate in constant PQ control mode, it cannot adjust its supply to maintain voltage and frequency according to the loads hence leading to poor power quality. Sudden resumption of grid due to action of automatic reclosers can cause circulating currents to flow if the utility and the island are out of phase. Loss of effective grounding in the islanded portion and transient overvoltages are other issues. There is also a constant safety threat to utility repair personnel due to a live portion existing on a dead power network.

Although internationally documented experiences of unintentional islanding events are limited, some real events noted in [5, 6] indicate the potential threat expected to intensify in the scenario of rising DG penetration. A recently published survey in [7] reaffirms the same concern among distribution utilities that feel that unintentional islanding will impact their network the most after DG interconnection.

Prevalent anti-islanding methods include classical methods involving passive, active, and hybrid techniques. These local techniques either monitor the parameters around the DG and sense any threshold-exceeding changes in them to detect islanding (passive) or force the parameters out of the safe range (active) or combine both strategies (hybrid). The problems of threshold selection causing large nondetection zones (NDZs) in passive and power quality disturbance from active techniques impact their use in high DG penetration. Many computational intelligence (CI) based techniques have also been reported in the literature [8] but they are fundamentally based on classical techniques and seem to reinforce the reactive strategy of detecting the island formation and

then disconnecting the DG. This practice is not expected to remain in the future smart grid that will accommodate a large share of renewable-energy based DG power which cannot be wasted even for a few cycles. Hence, a predictive approach to islanding detection can prove to be a robust solution.

Early works like [9] contemplated predicting PV inverter unintentional islanding in distribution grids considering the predictable utility supply and the load and PV generation profiles and utilizing analytical modeling of the early self-commutated inverters [10, 11]. More recently, data mining techniques were used on real [12] and simulated [13] phasor measurement unit (PMU) data to predict islanding in bulk transmission networks. Other works predicted the parameters at which a possible island could form. Traditional concepts of limit cycle behavior and small signal stability and describing function methods [14] and analysis of various inverter control functions [15] were used.

This paper describes the application of anomaly detection techniques: multivariate statistical and supervised learning based for predictive detection of an unintentional islanding event. The discovered anomalous currents occur before the islanding event on a modeled distribution feeder with a solar PV interconnection. This is a highlight of the paper which begins with a description of the system model in Section 2. Section 3 describes the anomalous precursors obtained from the dynamic simulations as part of the exploratory study. Section 4 describes data reduction using PCA and application of Q statistics for detecting the islanding precursors from other signals. An improved detection accuracy obtained using K-L divergence is detailed in Section 5. Section 6 describes labeling of the data points for training a K-NN classifier and reports its performance for the same test data sets as used in the previous two sections. Section 7 concludes the paper.

2. Power System Model

The distribution network modeled in this study is based on the benchmark IEEE 13 node test radial feeder [16] and was modeled in MATLAB-Simulink. Some modifications were made to the original feeder model in order to carry out the islanding studies as required. First of all, the substation automatic voltage regulator (AVR) at node 650 was not included in the model. This was done so as to avoid any possible interaction of PV with the tap-changing controls of the AVR [17] that could mask any signature related to islanding on the system. A 100.7 kWp solar PV array was integrated at node 692 through a three-phase inverter, thus making this node the PCC. These changes are visible in the modified feeder shown in Figure 2.

Apart from the changes mentioned previously, the constant current load at node 675 was removed and the active and reactive power demands P and Q , respectively, of the constant current load at 692 were modified so as to make section 671-692 the islandable feeder section. The PV inverter operates on unity power factor and thus the loads on lateral 671-675 had to be scaled according to the fixed PV penetration and feeder capacitor bank size for attaining P - Q balance required for

FIGURE 2: The modified IEEE 13 node feeder.

the exploratory islanding study described in the next section. The details of the system component modeling are given in our earlier work [18] from which the same model is taken here. This section only produces the results that describe the system model functioning and verification.

To verify that the system performs according to theory, it was required to test the voltage and frequency at any point on section 671-692 when it is islanded with the PV inverter. The voltage and frequency in an unintentional island depend on the mismatch of P and Q between the loads and the source(s) in that network. Accordingly, an island was forced to form by opening the islanding switch (circuit breaker in Figure 2) from $t = 0.45$ s to $t = 0.48$ s. The solar irradiance in this case and for all cases discussed in this paper was kept fixed at 1000 W/m^2 in order to operate the PV array at standard test conditions (STC) for a fixed, rated output. The MPPT is switched on at $t = 0.40$ s from the start of the simulation and reaches the MPP at 0.42 s after a few transients.

The situation of P - Q mismatch between the island loads and sources was created for which the values were set as follows: $P_{\text{load}} = 90 \text{ kW}$, $Q_{\text{load}} = 151 \text{ kVAr}$, $P_{\text{PV}} = 100 \text{ kW}$ (effective three-phase AC inverter output), and $Q_{\text{PV}} = 0$. Q was supplied by the feeder capacitor bank and the inverter's filter circuit, coming in the islanded network; thus, $Q_{\text{supply}} = Q_{\text{cap.bank}} + Q_{\text{inv.filterckt.}} = 600 \text{ kVAr} + 10 \text{ kVAr} = 610 \text{ kVAr}$. The island load values are for a single-phase load while the supply values of P and Q come from three phase sources. The voltage magnitude and frequency in the resulting island, measured at the PCC, are plotted along with many other quantities in Figure 3. The PV array continues operating in the preprogrammed constant P - Q control mode and hence the undervoltage and underfrequency are evident in the figure. The low value of voltage magnitude and frequency is consistent with the given amount of P - Q mismatch inside the island [19].

The harmonics observable in the displayed voltage and current are obvious when a PV system is integrated. However, the current harmonics shown are those in the grid-side current and not the inverter output. The simulation runs the model using a discrete solver which samples the voltages and currents at the rate of 1 MHz . Similar kinds of wave shapes were obtained from field tests carried out on a medium voltage distribution feeder section in Spain [4]. The resulting

FIGURE 3: Island parameters in P - Q mismatch case.

FIGURE 4: Field-test results as in [4].

island three-phase voltages and currents are reproduced with permission in Figure 4. Here, overvoltages are observed in the island. The time scale of observations in the field-test results is 10 ms/division . Some kinds of similarities in these two figures provide some confidence about the modeling approach although comparing simulation results with field-test results might not be completely appropriate.

3. Exploration of Anomalous Precursors to Unintentional Islanding

The level of P - Q mismatch between the loads and the PV inverter has a close association with island formation and sustenance as was seen in the previous section. Many practical studies documented in [20, 21] have characterized inverters'

islanding behavior for different levels of P - Q mismatch. However, in the expected scenario of rising PV penetration levels on distribution feeders, complete P - Q match is not a remote possibility. Internationally accepted reports like [22] have acknowledged that P - Q balance between PV inverter and loads is a quantifiable possibility. Field tests in [5] and laboratory tests in [23] have studied the inverters' anti-islanding capabilities for complete match. Also, a case study done for India [24] has estimated the risk of unintentional islanding in a spot distribution network based on the number of hours for which such a condition occurs. These documented practices have highlighted the significance of power mismatch to islanding, but after the occurrence of the event. This study explores the impact of complete P - Q match case on the possibility of building up of an imminent islanding condition.

Such dynamic load-PV interactions on a high PV penetration radial feeder for different grid conditions can throw up interesting results related to system islanding. In an attempt to explore such patterns, two types of disturbances were programmed to occur from the substation, undervoltage and overvoltage, in concurrence with exact P - Q balance. These two types of disturbances are commonly used in practice for islanding related studies [25]. In either case, the following values were set to implement P - Q match between the 1-phase load and the PV and Q sources on the islandable feeder section: $P_{\text{load}} = 23.33 \text{ kW}$ and $Q_{\text{load}} = 203.33 \text{ kVAr}$. A 10 kW resistor in parallel with this load model forces the exact P match as P and Q demands of the inverter filter circuit are negligible. The values of P_{PV} and Q_{supply} remain the same as before.

An undervoltage disturbance of 0.7 per unit (pu) of the nominal voltage amplitude was forced from the grid side for a period of 30 ms from $t = 0.45 \text{ s}$ to $t = 0.48 \text{ s}$. The grid-side current flowing in phase C of section 671-692 has many anomalous peaks during the voltage disturbance and after the disturbance ends as shown in Figure 5. The 30 ms window during and after the disturbance is important for this study from the perspective of data extraction. The circled peak is an anomaly whose severity to cause islanding is explained and verified in [26] which also verifies the occurrence of a similar peak for the same set of conditions on a single-phase, single bus system implemented in emulators and related hardware. An overvoltage disturbance, symmetrically 1.3 pu of the nominal voltage amplitude, was programmed to occur from the grid side for 30 ms from $t = 0.45 \text{ s}$ to 0.48 s . The resulting grid-side anomalous current shown in Figure 6 was not proved to be severe enough to lead to section islanding. The complete system was remodeled and simulated for 0.5 s in real time on a real-time digital simulator (RTDS) for both disturbance cases. This was to validate the model robustness and the real-time results similar to Simulink results are shown in Figures 7 and 8 for both disturbance cases, respectively. The large peaks are not noteworthy since they are due to initial PV integration transients. The 30 ms window of each disturbance case contains data points sampled at 1 MHz. The voltage and current samples of phase C taken around the PCC collected in this window are used as one of the data sets in each of

FIGURE 5: Anomalous peak as islanding precursor.

the different statistical and CI techniques discussed in the next three sections.

A 30 ms window after each disturbance ends is also captured during the simulation run. For the case of Figure 7 previously, this window contains data points corresponding to the anomalous current liable to cause islanding and hence is used as a part of the composite training data set used for the K -NN classifier training. More details about data sets and applications follow in the next section.

4. Data Handling and Anomaly Detection Using PCA

The simulation for each of the two grid disturbances was run for 0.5 seconds at a sampling rate of 1 MHz as discussed above. The data of the voltage and current samples mentioned above was collected and bifurcated into different data sets according to three conditions found in each of the two cases, namely, normal, during disturbance, and after disturbance. The normal condition was common for both the disturbance case simulations and contained 30484 samples. The normal condition corresponds to that where no extra disturbances, apart from the PV induced harmonics, are present in the system.

PCA was used to reduce the dimensionality of the data set corresponding to the normal system operation case as voltage and current are correlated quantities. The standard singular value decomposition (SVD) approach was used in MATLAB to find the principal component matrix. The two-dimensional

FIGURE 6: Overvoltage with P - Q match. FIGURE 7: RTDS result for undervoltage with P - Q match.

data resulted in two principal components (PCs). Based on the variance of the projections onto the two PCs, the 1st PC was retained for all analyses. The latent matrix containing the values of variances onto the 2 PCs called the scores is shown in Table 1 and makes the choice of PC selection clear.

The PCA model created was used for detecting any abnormal occurrence using statistical process control strategy for anomaly detection. The data belonging to different cases simulated was preprocessed and projected onto the 1st PC of the reference PCA model. The aim of this strategy is

FIGURE 8: RTDS result for overvoltage with P - Q match.

TABLE 1: Latent values.

On PC 1	On PC 2
5.37×10^6	486.6

to differentiate a condition that can cause unintentional islanding on the modeled feeder from conditions like faults and other transients which appear close to islanding and thus are tricky to detect and identify correctly. The existing literature describes techniques that detect an islanding condition among other transients like surges, load and capacitor switching, and faults after the island has been formed. This study initiates efforts towards exploring possible practical causes of the event and detecting such conditions from the ones that appear close enough to fool the inverter. Hence, only the four cases resulting from the two grid-side disturbances as described previously and a 3-phase short-circuit fault case have been simulated.

Case 1 is the normal system operation case which has been described previously. Each of the grid-side undervoltage and overvoltage disturbance conditions gives rise to two cases. A three-phase line-to-line-to-line-to-ground (L-L-L-G) fault at the PCC is designated as case 4. All the event based cases simulated are summarized in Table 2.

For n number of data sample vectors $x \in R^m$ stacked above one another to form a data matrix $X^{n \times m}$, application of PCA on X leads to a $p \times p$ coefficient matrix P . If an $r \leq m$ number of PCs are retained based on latent values, then X can be resolved as a PCA model and a residual model as $X = X_{pca} + X_{res}$. The projection onto the PC or loading matrix leads to formation of a score matrix $T^{n \times m}$. The original data matrix X can be reconstructed using score matrices T_{pca} and T_{res} and loadings P_{pca} and P_{res} as $X = T_{pca} P_{pca}^T + T_{res} P_{res}^T$, where P_{res} and T_{res} are of $n \times m - r$ dimension [27].

The statistical process control method is widely used in industrial engineering for quality control purposes. It has found other applications in many domains for outlier detection by checking whether the process variables are in

TABLE 2: Cases simulated.

Case	Number of samples	Event
Case 1	30484	Normal
Case 2	30584	UV + P-Q match (during disturbance)
Case 3	30622	UV + P-Q match (after disturbance)
Case 4	30585	L-L-L-G fault at PCC
Case 5	30656	OV + P-Q match (during disturbance)
Case 6	30644	OV + P-Q match (after disturbance)

control or not. Any process variable is indicated as out of control when a certain statistic associated with it crosses its upper limit. PCA has two multivariate statistics associated with it: Hotelling's T^2 statistic and Q statistic. Both have an upper control limit (UCL) defined and when both of them are crossed by the corresponding statistics of a data point or data set, this indicates an anomalous and abnormal behavior.

Hotelling's T^2 statistic is a multivariate distance for a set of data points from a target value indicating variance inside the PCA model. If f is a mean-centered (scaled) sample data vector, then $t_{pca} = fP_{pca}$ is a score vector. The T^2 statistic for f is defined as $T^2 = t' \Lambda t$, where Λ is a diagonal matrix having r eigenvalues of data matrix $F^{n \times m}$ for $r \leq m$ number of retained PCs. The UCL for the statistic is defined as T_{α}^2 . If all data points are linear and normally distributed, T_{α}^2 follows an F distribution and is given as $T_{\alpha}^2 = r((n^2 - 1)/n(n - r))F_{r, n-r}$ at a given level of confidence α .

The Q statistic is a measure of deviation of the original data points from the projection onto the PC axes. Hence, it measures variance among data points inside the residual subspace. The Q statistic is calculated using residuals and, for a residual vector e of a scaled sample vector f , Q statistic is given as $Q = e^T e = f^T(I - P_{pca}P_{pca}^T)$, where I is an identity matrix. For normally distributed linear data points, the Q statistic follows a central χ^2 distribution and its UCL is given as $Q_{\alpha} = (\sigma^2/2\mu) \times \chi^2(2\mu^2/\sigma^2)$ at a given level of confidence α . Here, μ and σ^2 are the mean and variance of the Q statistic.

Recently, PCA based process control strategy has been applied for detecting the occurrence of islanding and distinguishing it from several nonislanding events. PMU recordings of frequency measurements on 6 different sites in the UK power grid were used as reference data for implementing T^2 and Q statistic based islanding detection in [28]. The occurrence of an islanding situation was evident only when Q_{α} was crossed in addition to the crossing of T_{α}^2 by the corresponding multivariate statistics for a test event data set. Since the power system is a dynamically changing system, the system variables used for creating the reference PCA model change dynamically causing it to change with time also. To tackle this issue, a recursive PCA algorithm was developed in [29] for the same UK power system case. The reference PCA model was updated in every iteration and the detection

TABLE 3: Event detection results using Q statistic.

Case	Q	$Q > Q_{\alpha}$	Islanding precursor
2	1.57×10^7	No	No
3	4.05×10^7	Yes	Yes
4	1.23×10^9	Yes	No
5	8.57×10^7	Yes	No
6	4.32×10^7	Yes	No

results for abnormal transients verified its effectiveness over the simple PCA approach. This study has made use of the usual SVD for creating the reference PCA model since the reference data does not change from one event to another as the simulation has been performed for fixed settings to observe some unique changes that occur in fixed windows as described previously.

Each new test data set $X^{\text{samples} \times 2}$ underwent scaling to make the mean along the columns zero. The mean-centered data set X_{mc} was projected onto the 1st PC of the reference PCA model by $X_{mc} \times P_{pca}$. Correspondingly, the T^2 and Q statistics were calculated. Each of the remaining 5 cases was used as the test case. Since crossing of the T_{α}^2 limit for the reference case by the T^2 statistic of any test case indicates only a faulty or out-of-control event, the Q statistic was used as the only parameter for detection. Q statistic measures deviation inside the residual subspace and hence is a strong indicator of any abnormal or anomalous condition.

Following the same, the 5 test cases were subjected to mean-centering as before and were projected onto the 1st PC of the reference PCA model. The Q statistics for each case projected data matrix were found and compared with the UCL Q_{α} of the reference case score. Q_{α} at 98% confidence level was calculated to be $= 3.846 \times 10^7$. The results of this multivariate statistics based detection are given in Table 3.

As seen in Table 3, this approach identifies the anomalous case correctly. It also identifies the disturbance event in case 2 correctly as not an anomaly that can island the system. However, cases 4, 5, and 6 are incorrectly identified. This shows that the Q statistic based statistical process control approach is not completely reliable for detecting the anomalous currents that can lead to islanding on the system. To improve upon the false detection rate, the Kullback-Leibler (K-L) divergence based approach using the PCA model is presented in the next section.

5. K-L Divergence Based Detection

K-L divergence, also known as relative entropy, is an important statistical measure coming from information theory. It has shown a great potential for application in fault detection and diagnosis (FDD). It has been aptly used for incipient fault detection in mechanical and electrical systems in [30] and has also been widely used in multimedia security and neuroscience. However, the application of K-L divergence in islanding detection related studies could not be confirmed in the literature. This section details the use of K-L divergence

involving the PCA model for improved accuracy of anomaly detection.

K-L divergence is basically a measure of dissimilarity between two probability distributions. If two data samples are drawn from two populations having the same distribution, their K-L divergence will be zero. For two continuous probability density functions (PDFs) $f(x)$ and $g(x)$ of a random variable x , the K-L Information (KLI) is defined as $I(f \parallel g) = \int f(x) \log(f(x)/g(x))dx$. The K-L divergence is then given as $K-LD(f, g) = I(f \parallel g) + I(g \parallel f)$, a symmetric operation of KLI. For discrete distributions, K-LD is defined as the mean value of the log-likelihood ratio of the two distributions.

For an anomalous behavior or a sudden change in a process, the PDF of the corresponding data set changes from the reference case and if it goes beyond the safe threshold ϵ , it can be statistically detected. For two normal (Gaussian) probability densities f and g having means and variances as μ_1, μ_2 and σ_1^2, σ_2^2 , respectively, the K-L divergence between them can be given by a simple expression:

$$K-LD = \frac{1}{2} \left[\frac{\sigma_2^2}{\sigma_1^2} + \frac{\sigma_1^2}{\sigma_2^2} + (\mu_1 - \mu_2)^2 \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) - 2 \right]. \quad (1)$$

In our case, we find the divergence between two distributions: projection of different event (test) cases onto the 1st PC and the reference PCA score or projection of case 1 on the 1st PC. Nonparametric kernel-density estimation has been used to approximate each of these two distributions as normal distributions graphically. Since mean-centering of data samples is a part of PCA, the means of the projections for both test cases and reference case are zero. Since the PC scores are linear combinations of the original data samples, they are assumed to be fairly normally distributed [31]. Taking this assumption into consideration, we have used the following formula to calculate K-LD between a test case and the reference case:

$$K-LD = \frac{1}{2} \left[\frac{\sigma_{\text{test case}}^2}{\sigma_{\text{ref.}}^2} + \frac{\sigma_{\text{ref.}}^2}{\sigma_{\text{test case}}^2} - 2 \right]. \quad (2)$$

Here, $(\mu_{\text{ref}} - \mu_{\text{test}})^2 = 0$ and $\sigma_{\text{ref.}}^2$ is nothing but the variance of projection on the 1st PC which is in the first column of Table 1. The other variances are those of the projections of the different test cases onto the 1st PC.

Using (2), test cases 2 to 6 were used as the second distribution and case 1 was taken as the reference distribution. The values of K-L divergence calculated for different cases are given in Table 4. The results from Table 4 throw an important picture. All those cases which had $Q > Q_\alpha$ and were wrongly detected seem to have been differentiated by their K-L divergence values. It can be seen clearly that cases 4, 5, and 6 do not fall in the same category as had been previously clubbed by the Q statistic approach. The extremely large and small values of cases 4 and 6, respectively, segregate them into different category of events; however, the similar orders of values for case 3 and case 6 do not give a clear boundary.

The kernel-density estimated normal PDFs for cases 3, 4, and 6 and their divergence from that of case 1 are shown

TABLE 4: Event detection results using K-LD.

Case	σ^2	K-LD	Islanding precursor
2	1.95×10^6	0.554	No
3	4.92×10^6	0.004	Yes
4	5.53×10^3	485.55	No
5	8.42×10^6	0.1023	No
6	5.12×10^6	0.0012	No

FIGURE 9: Divergence: case 3 versus case 1.

in Figures 9, 10, and 11, respectively. The more the K-LD, the more the gap between the densities. The L-L-L-G fault case has the least variance and hence it has the largest K-L divergence among all cases. Physically, this event creates such low voltages for a given short-circuit capacity of the feeder that the PV inverter itself trips thus avoiding islanding and this fact is brought out by its large divergence from the reference case PDF. However, after looking at the divergence values of case 3 and case 6, setting the correct ϵ for an event to be identified as the anomalous islanding precursor seems to be the problem with this approach although the false alarm detection rate has reduced to 1/5 from 3/5 in the previous section. To tackle this issue of threshold selection, a machine-learning based approach to detect anomalous events correctly has been presented in the next section.

6. K-NN Classifier Based Approach

Moving from the statistical techniques presented in the previous sections, this section describes application of a supervised learning technique called K -nearest neighbors (K -NN) classification. K -NN methods are instance based learning methods that classify a new test instance based on its similarity to the training data points stored. Such techniques are also called lazy-learning techniques because they do not create a model for classification of test data but rather they

FIGURE 10: Divergence: case 4 versus case 1.

FIGURE 11: Divergence: case 6 versus case 1.

carry out classification only when a new instance comes up. As a result, they look into the database of training data points similar to the test point based on some distance measure and assign a class label to the new point usually based on a majority vote among the training points in its neighborhood [32]. The algorithm used here is described as follows.

Training Algorithm. (i) Add each training example $(x, f(x))$ to the list of training examples. Here, $f(x)$ is a binary encoded indicator response variable storing the class label for x .

Classification Algorithm. Given a query point x^q to be classified,

- (i) let $x_1, x_2, \dots, x_k$ be the k training instances nearest to x_q based on a distance measure;
- (ii) return:

$$\widehat{f(x_q)} \leftarrow \frac{\sum_{i=1}^k f(x_i)}{k}. \quad (3)$$

For this study, this algorithm was selected because the anomalous instances liable to cause islanding in case 3 were all consecutively located in the data set. Two class labels were used for binary classification of the test data points. Class label 0 = all data points $\in$ set of cases that *cannot cause islanding* while class label 1 = all data points $\in$ set of cases that *can cause islanding*. Physically, this corresponds to all currents $I > 0.1$ kA as decided from the static fault study explained in [26]. Three training data sets were used as a composite training data set.

First 20000 Points from Case 1, All Data Points from Case 2, and 555 Anomalous Data Points from Case 3. The number of neighbors k was set = 5 and Euclidean distance measure was used. The classifier was trained and tested with three test data sets:

Test set I: last 10,484 points of case 1.

Test set II: data points $\in$ case 4.

Test set III: data points $\in$ case 6.

The average cross-validated classification error or 10-fold loss on the training data was 0.0070 indicating high training accuracy. The classifier performance was tested for each of the test cases using the majority vote with nearest point tie-break rule. For test set II, the classifier identified all data points to $\in$ class 0. This pertains to a 100% accuracy in this case. The variance of data points in case of the 3- ϕ fault is the least among all cases. Fault also causes very low voltage which itself can trip a PV inverter and thus it is detected naturally and cannot be labeled as an anomalous precursor. This confirms the correctness of the classifier in assigning label 0 to this case. The classifier accuracy for test sets I and III was found to be 97.42% and 90.12%, respectively, after multiple runs. The confusion matrices for test sets I and II are shown in Tables 5 and 6, respectively. Case 6 comes very close to the case of actual islanding precursors discovered in case 3 and hence a large number of data points were assigned label 1. The average classifier accuracy can be reported as 95.75%. The classifier takes an average time of 294 ms in classifying a new test data point. As k was reduced till 1, the time taken remained the same but the accuracy improved even for testing on the third training data set 3. Clearly, in this approach also, the three-phase short-circuit fault case is identified to be different from all other cases with 100% accuracy.

A comparison of the performances of the three methods discussed in this paper is presented in Table 7.

7. Conclusions

This paper has contributed an exploratory study towards understanding, discovering, and analyzing the possible reasons that can unintentionally island a modified IEEE 13 bus

TABLE 5: Confusion Matrix I.

	Class 0	Class 1
Class 0	10214	270
Class 1	0	0

TABLE 6: Confusion Matrix III.

	Class 0	Class 1
Class 0	27618	3026
Class 1	0	0

TABLE 7: Performance comparison of the three methods.

Method	Accuracy	Issues
Q statistic based	40%	High false alarm rate
K-L divergence	80%	Setting correct ϵ
K-NN classifier	95.75%	k and time trade-off

system with large PV penetration on a segment. The application of multivariate statistical techniques and a supervised learning technique to identify anomalous signatures liable to cause islanding from other transients has been detailed. The three-phase short-circuit fault is clearly identified as not an islanding precursor case by both the K-L divergence and K-NN classification methods. The paper also shows that using PCA based process control strategy alone is not sufficient for predictive islanding detection. The classifier gives the best accuracy among all and it can be concluded that its implementation should detect the precursor and trip the PV inverter before the utility PCC relay. The classification time can be equal to the relay delay if not less than it.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The authors acknowledge the Ministry of New and Renewable Energy of the Government of India for supporting this work under the National Renewable Energy Fellowship Scheme. The support provided by Dr. Deepak Fulwani and Mr. Vinod Kumar of the Indian Institute of Technology Jodhpur for helping with the RTDS study is equally acknowledged.

References

- [1] R. C. Dugan and T. E. McDermott, "Distributed generation," *IEEE Industry Applications Magazine*, vol. 8, no. 2, pp. 19–25, 2002.
- [2] L. Mihalache, S. Suresh, Y. Xue, and M. Manjrekar, "Modeling of a small distribution grid with intermittent energy resources using MATLAB/SIMULINK," in *Proceedings of the IEEE Power and Energy Society General Meeting*, pp. 1–8, San Diego, Calif, USA, July 2011.
- [3] C. Greacen, R. Engel, and T. Quetchenbach, "A guidebook on interconnection and islanded operation of mini grid power systems up to 200 kW," Tech. Rep., Lawrence Berkeley National Laboratory, Berkeley, Calif, USA, 2013.
- [4] F. J. Pazos, "Power frequency overvoltages generated by solar plants," in *Proceedings of the 20th International Conference on Electricity Distribution (CIRED '09)*, pp. 159–159, Prague, Czech Republic, June 2009.
- [5] F. J. Pazos, "Operational experience and field tests on islanding events caused by large photovoltaic plants," in *Proceedings of the 21st International Conference on Electricity Distribution (CIRED '11)*, Frankfurt, Germany, June 2011.
- [6] C. Li, J. Savulak, and R. Reinmuller, "Unintentional islanding of distributed generation-Operating experiences from naturally occurred events," *IEEE Transactions on Power Delivery*, vol. 29, no. 1, pp. 269–274, 2014.
- [7] R. F. Arritt and R. C. Dugan, "Review of the impacts of distributed generation on distribution protection," in *Proceedings of the 59th IEEE Annual Conference on Rural Electric Power Conference (REPC '15)*, pp. 69–74, Asheville, NC, USA, April 2015.
- [8] J. A. Laghari, H. Mokhlis, M. Karimi, A. H. A. Bakar, and H. Mohamad, "Computational intelligence based techniques for islanding detection of distributed generation in distribution network: a review," *Energy Conversion and Management*, vol. 88, pp. 139–152, 2014.
- [9] G. J. Vachtsevanos and H. Kang, "Simulation studies of islanded behavior of grid-connected photovoltaic systems," *IEEE Transactions on Energy Conversion*, vol. 4, no. 2, pp. 177–183, 1989.
- [10] J. Stevens, R. Bonn, J. Ginn, and S. Gonzalez, "Development and testing of an approach to anti-islanding in utility-interconnected photovoltaic systems," Tech. Rep. SAND 2000-1939, Sandia National Laboratories, 2000.
- [11] M. A. Eltawil and Z. Zhao, "Grid-connected photovoltaic power systems: technical and potential problems—a review," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 1, pp. 112–129, 2010.
- [12] V. Vittal, T. Merho, M. Kezunovic et al., "Data mining to characterize signatures of impending system events or performance from PMU measurements," Tech. Rep., Arizona State University and Texas A&M University, 2013.
- [13] Z. Pakdel, *Intelligent instability detection for islanding prediction [Ph.D. dissertation]*, Virginia Polytechnic Institute and State University, 2011.
- [14] S. J. Ranade, N. R. Prasad, S. Omick, and L. F. Kazda, "A study of islanding in utility-connected residential photovoltaic systems part I-models and analytical methods," *IEEE Transactions on Energy Conversion*, vol. 4, no. 3, pp. 436–445, 1989.
- [15] G. A. Smith, P. A. Onions, and D. G. Infield, "Predicting islanding operation of grid connected PV inverters," *IEE Proceedings: Electric Power Applications*, vol. 147, no. 1, pp. 1–6, 2000.
- [16] W. Kersting, "Radial distribution test feeders," in *Proceedings of the IEEE Power Engineering Society Winter Meeting*, vol. 2, pp. 908–912, February 2001.
- [17] M. S. Elnozahy and M. M. A. Salama, "Technical impacts of grid-connected photovoltaic systems on electrical networks—a review," *Journal of Renewable and Sustainable Energy*, vol. 5, no. 3, Article ID 032702, 2013.
- [18] S. Vyas, R. Kumar, and R. Kavasseri, "Observations from study of pre-islanding behaviour in a solar PV system connected to a distribution network," in *Proceedings of 4th International*

- Conference on Advances in Computing, Communications and Informatics (ICACCI '15)*, pp. 645–650, Kochi, India, August 2015.
- [19] T. Tran-Quoc, T. M. C. Le, C. Kieny, and S. Bacha, “Behaviour of grid-connected photovoltaic inverters in islanding operation,” in *Proceedings of the IEEE PES Trondheim PowerTech: The Power of Technology for a Sustainable Society (POWERTECH '11)*, pp. 1–8, Trondheim, Norway, June 2011.
 - [20] M. Dymond, “Practical results from islanding tests on the 20 kW Kalbarri PV system,” in *Proceedings of the Solar '97: “Sustainable Energy”, 35th Annual Conference of the Australian and New Zealand Solar Energy Society*, December 1997.
 - [21] M. Ross, C. Abbey, Y. Brissette, and G. Joos, “Photovoltaic inverter characterization testing on a physical distribution system,” in *Proceedings of the IEEE Power and Energy Society General Meeting (PES '12)*, pp. 1–7, San Diego, Calif, USA, July 2012.
 - [22] B. Verhoeven, “Probability of islanding in utility networks due to grid connected photovoltaic power systems,” IEA, Tech. Rep IEA-PVPS T5-07:2002, 2002.
 - [23] R. J. Bravo, S. A. Robles, and E. Muljadi, “Assessing solar PV inverters’ anti-islanding protection,” in *Proceedings of the 40th IEEE Photovoltaic Specialist Conference (PVSC '14)*, pp. 2668–2671, IEEE, Denver, Colo, USA, June 2014.
 - [24] K. A. Joshi and N. M. Pindoriya, “Risk assessment of unintentional islanding in a spot network with roof-top photovoltaic system—a case study in India,” in *Proceedings of the IEEE Innovative Smart Grid Technologies—Asia (ISGT Asia '13)*, pp. 1–6, Bangalore, India, November 2013.
 - [25] S. Srivastava, K. S. Meera, R. Aradhya, and S. Verma, “Performance evaluation of composite islanding and load management system using real time digital simulator,” in *Proceedings of 15th National Power Systems Conference*, pp. 242–247, Mumbai, India, 2008.
 - [26] S. Vyas, R. Kumar, and R. Kavasseri, “Exploration and investigation of potential precursors to unintentional islanding in grid-interfaced solar photo voltaic systems,” in *Proceedings of the IEEE International Conference on Energy Systems and Applications (ICESA '15)*, pp. 579–584, Pune, India, October 2015.
 - [27] J. Tang, W. Yu, T. Chai, and L. Zhao, “On-line principal component analysis with application to process modeling,” *Neurocomputing*, vol. 82, no. 2012, pp. 167–178, 2012.
 - [28] X. Liu, D. M. Laverty, R. J. Best, K. Li, D. J. Morrow, and S. McLoone, “Principal component analysis of wide-area phasor measurements for islanding detection—a geometric view,” *IEEE Transactions on Power Delivery*, vol. 30, no. 2, pp. 976–985, 2015.
 - [29] Y. Guo, K. Li, D. M. Laverty, and Y. Xue, “Synchrophasor-based islanding detection for distributed generation systems using systematic principal component analysis approaches,” *IEEE Transactions on Power Delivery*, vol. 30, no. 6, pp. 2544–2552, 2015.
 - [30] J. Harmouche, *Statistical incipient fault detection and diagnosis with kullback-leibler divergence: from theory to applications [Ph.D. thesis]*, Supélec, 2014.
 - [31] A. Youssef, C. Delpha, and D. Diallo, “Performances theoretical model-based optimization for incipient fault detection with KL Divergence,” in *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO '14)*, pp. 466–470, September 2014.
 - [32] T. M. Mitchell, *Machine Learning*, McGraw Hill, New York, NY, USA, 1997.