

University of Wollongong

Research Online

Faculty of Engineering and Information
Sciences - Papers: Part A

Faculty of Engineering and Information
Sciences

1-1-2013

A SAX-based advanced computational tool for assessment of clustered rooftop solar PV impacts on LV and MV networks in smart grid

M J. E Alam

University of Wollongong, mjea982@uowmail.edu.au

K M. Muttaqi

University of Wollongong, kashem@uow.edu.au

D Sutanto

University of Wollongong, soetanto@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/eispapers>



Part of the [Engineering Commons](#), and the [Science and Technology Studies Commons](#)

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

A SAX-based advanced computational tool for assessment of clustered rooftop solar PV impacts on LV and MV networks in smart grid

Abstract

Future distribution networks with increasing level of solar PV penetration will be managed using smart grid technologies capable of producing appropriate and timely response during normal and abnormal operational events. Distribution feeder loads vary throughout the day according to the trend of consumption of the customers. Solar PV outputs fluctuate in proportion to irradiance level of sun. Simultaneous occurrence of both of these variations would result in various operating conditions that may lead to unexpected events, and would require a large amount of network data to be processed and analyzed for decision making. It is envisaged that such data will be available in the future grids with the availability of smart technologies and advanced communication in residential dwellings, commercial buildings and industrial complexes. In this paper, an advanced intelligent computational tool is developed to characterize and analyze the large amount of data associated with wide variations in network behavior using SAX (Symbolic Aggregate Approximation) and pattern recognition. The proposed tool is capable of dealing with network asymmetry, load unbalance, single-phase solar PV integration and their impacts on upstream networks and will assist in making right and timely decision to mitigate adverse impacts of solar PV. The proposed tool has been tested with a practical three-phase distribution system in Australia and can provide an extensive assessment with less computational efforts and time. © 2010-2012 IEEE.

Keywords

networks, mv, lv, impacts, pv, advanced, solar, sax, rooftop, clustered, assessment, tool, computational, grid, smart

Disciplines

Engineering | Science and Technology Studies

Publication Details

M. J. E. Alam, K. M. Muttaqi & D. Sutanto, "A SAX-based advanced computational tool for assessment of clustered rooftop solar PV impacts on LV and MV networks in smart grid," IEEE Transactions on Smart Grid, vol. 4, (1) pp. 577-585, 2013.

A SAX-Based Advanced Computational Tool for Assessment of Clustered Rooftop Solar PV Impacts on LV and MV Networks in Smart Grid

M. J. E. Alam, *Student Member, IEEE*, K. M. Muttaqi, *Senior Member, IEEE*, and D. Sutanto, *Senior Member, IEEE*

Abstract-- Future distribution networks with increasing level of solar PV penetration will be managed using smart grid technologies capable of producing appropriate and timely response during normal and abnormal operational events. Distribution feeder loads vary throughout the day according to the trend of consumption of the customers. Solar PV outputs fluctuate in proportion to irradiance level of sun. Simultaneous occurrence of both of these variations would result in various operating conditions that may lead to unexpected events, and would require a large amount of network data to be processed and analyzed for decision making. It is envisaged that such data will be available in the future grids with the availability of smart technologies and advanced communication in residential dwellings, commercial buildings and industrial complexes. In this paper, an advanced intelligent computational tool is developed to characterize and analyze the large amount of data associated with wide variations in network behavior using SAX (Symbolic Aggregate Approximation) and pattern recognition. The proposed tool is capable of dealing with network asymmetry, load unbalance, single-phase solar PV integration and their impacts on upstream networks and will assist in making right and timely decision to mitigate adverse impacts of solar PV. The proposed tool has been tested with a practical three-phase distribution system in Australia and can provide an extensive assessment with less computational efforts and time.

Index Terms—Solar PV Impacts, SAX, Pattern Clustering, Anomaly detection, Smart Grid.

I. INTRODUCTION

A large number of single phase solar photovoltaic (PV) units are expected to be integrated with future distribution grids in clustered form [1]. Conventional distribution networks are typically designed to supply electric power from upstream high or medium voltage networks to consumers through low voltage feeders. With integration of multiple numbers of solar PV units, normally expected network behavior may be changed [2]. For example, voltage profile may rise along the feeder with high PV penetration, whereas traditionally voltage drop along the feeder is observed. Similarly, power could flow from load (consumers) to source (substation), whereas in

traditional distribution grids power flow from source (substation) to load (consumers) is typically found. The degree of change in network behavior will vary throughout the day depending on the variations in load demand and PV output. The highest amount of changes could typically be expected at noon time when solar PV units would generate at the peak level of the available capacity. With high PV penetration, the risk of violating system operation limits caused by changes in network behavior may therefore arise at this period of the day. Considering the possibility of operation limit violation, it is important to investigate the potential impacts of PV clusters on distribution network to take corrective actions so that solar PV penetration can be increased to the desired level.

At present, the assessment of solar PV impacts is typically performed by applying a number of assumptions on network data, especially the load demand data on LV feeders. Simplification of network is also performed by aggregating all the feeder loads and PV generation at the LV substation secondary side [3]. In future, under smart grid environment, numerous smart meters with integrated communication facilities are anticipated to be installed to collect distribution feeder operational data throughout the day that would be available for analysis of normal and abnormal events. Impact analysis of solar PV cluster then could be performed with realistic data on the original network topology retaining detailed model of LV networks. This would produce results with higher accuracy and would aid in more effective decision making for implementing corrective actions. However, a massive amount of high resolution time series data will be collected by the smart meters and need to be analyzed. Task of impact assessment would be easier and would incur less computational time if the assessment tool is capable of data mining to detect hidden patterns and anomalies from the stored database and identify clusters that are of interest for PV impact assessment. These clusters will contain smaller dataset and can be studied in more details with less computational efforts and time. Intelligent clustering and pattern recognition techniques may be used to achieve this.

Further, performing data mining tasks on raw data may require extensive computational efforts. Approximation of the high resolution time series data can be performed to decrease the size of the database so that the computational performance of the tool can be improved. A number of time series approximation techniques have been proposed in the data

This work is supported by the Australian Research Council (ARC) and Essential Energy Linkage Grant, LP100100618.

M. J. E. Alam, K. M. Muttaqi and D. Sutanto is with Endeavour Energy Power Quality and Reliability Center, School of Electrical Computer and Telecommunications Engineering, University of Wollongong, NSW 2522, Australia (e-mail: mjea982@uowmail.edu.au, kashem@uow.edu.au, soetanto@uow.edu.au).

mining literature [4]. Recently, a symbolic approximation technique, namely SAX (Symbolic Aggregate Approximation) method, for time series data representation has been proposed [4], that features data dimensionality reduction and lower bounding capability [5] which are essential for data mining tasks. Only a few applications of this method in power systems [6] have been observed so far. Authors in this paper have reduced a number of electricity price time series data into SAX form prior to clustering resulting in improved computational performance. The SAX method, however, has been applied extensively in other disciplines with reported excellent results [7]. Based on the beneficial features of this method discussed in literature, the purpose of this paper is to explore the usefulness of this method for the assessment of solar PV impacts using large amount of time series data available from smart meters and other smart appliances in future distribution grids.

The major contribution of this paper is to propose a comprehensive rooftop solar PV impact assessment tool to analyze a large database of future smart grid measurements. The novelty of the contribution lies in the application of SAX algorithm combined with data mining and pattern recognition techniques for PV impact analysis by extracting the behavior of a particular PV impact of interest (such as voltage rise, reverse power flow, cloud passing etc.) at a given location and time of occurrence from a large database of smart grid measurements. This offers a new way of data analysis, impact characterization and feature extraction from a huge volume of high-resolution recorded data collected from a distribution system containing solar PV. The tool will be capable of PV impact analysis based on a large set of realistic time series data available from numerous smart meters installed in the networks. SAX approximation of time series data will be performed to reduce the dimensionality of the database. Time series clustering and anomaly detection techniques will then be applied to identify only the data that are important for PV impact analysis. With a large number of single phase PV units integrated into distribution feeders, the tool will also be capable to evaluate the solar PV impacts on individual phases in LV systems, as well as on the upstream MV networks.

II. CLUSTERED ROOFTOP SOLAR PV IMPACTS ON DISTRIBUTION NETWORKS

Solar PV impacts on distribution networks are experienced in terms of changes in operational behavior of the networks, such as, changes in voltage, power flow, power loss, etc. The impacts of the integration of solar PV in clustered form may propagate to the upstream networks. Hence, the changes produced by the solar PV units in the LV networks may impact the operation of the voltage control devices in the MV networks. For example, tap operation of voltage regulators may be influenced by the change in voltage profile caused by solar PV units. Several impacts of clustered solar PV on distribution networks are briefly described below.

A. Voltage Rise

In the traditional distribution networks, voltage along a

distribution feeder typically reduces from the voltage at the substation bus, along the feeder, due to voltage drop across the line segments. If solar PV resources are integrated, loads are served locally resulting to less current flows through the feeders. This reduces the amount of voltage drop and power losses and as a result voltage profile is improved. However, if the generations from PV resources are high enough to offset the amount of loads in the feeder, the surplus amount of power will create voltage rise. This effect may be observed at midday when total PV generation in the feeder is higher than the feeder load demand. With cluster based installation of PV, the voltage rise impact may propagate to upstream MV networks. Voltage profile throughout the day at a given bus of a hypothetical distribution feeder, shown in Fig. 1(a), would look like the one given in Fig. 1(b); the voltage profiles with and without solar PV are identified with arrows showing the voltage rise at midday caused by solar PV output. Fig. 1(b) clearly indicates that two different types of time series patterns are produced with and without PV inclusion.

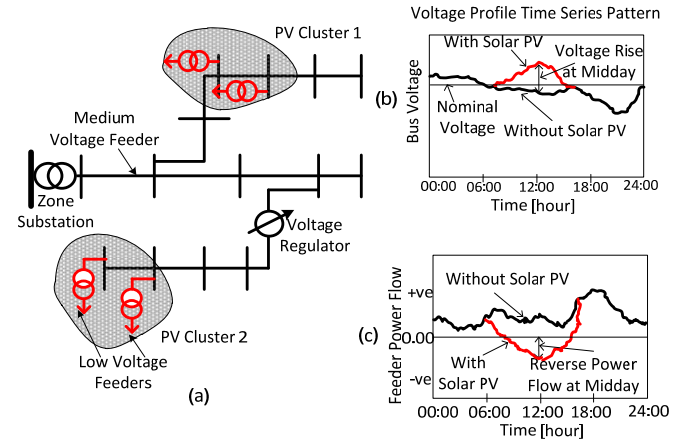


Fig. 1. Time series patterns of solar PV impacts; (a) simple distribution feeder (b) voltage profile pattern (c) power flow pattern.

B. Reverse Power Flow

LV distribution feeders are typically radial and the direction of power flow is, therefore, from substation towards the loads. Similar to the voltage rise impact, power flow direction would be changed when PV generation exceeds the load demand. Active power flow profile throughout the day at the secondary side of an LV distribution feeder is shown in Fig. 1(c). Reverse power flow is observed at midday due to excess output from solar PV. The difference in power flow time series patterns due to PV integration is obvious from Fig 1(c).

C. Variation of Feeder Power Loss

Power losses would typically decrease in distribution feeder with solar PV resources integration as less amount of power is imported from the substation. However, with a high penetration of PV cluster, if the reverse power flow is higher than the power flow without PV, an increase in feeder power loss may be observed [8]. Power loss may vary due to the variation of the PV output at the different time of the day.

D. Voltage Unbalance

Allocation of PV inverters among different phases of a distribution feeder is mainly controlled by the choice of the customers. The distribution of PV generation in a PV cluster therefore may not be equal at all the phases. This may deteriorate the existing voltage unbalance factor of the network [9]. The unbalance factor may vary from time to time due to the variations of solar irradiance and PV output.

E. Change in Tap Operations

Tapping operation of voltage regulators in the MV distribution feeders is performed in response to changes in voltage at the load centre. Voltage regulators may tap up or down their tap positions to keep voltage at the load centre within a bandwidth of voltage upper and lower limits. Voltage rise caused by PV clusters may require the regulators to operate during midday to keep the voltage profile below the upper limit.

III. PROPOSED COMPUTATIONAL TOOL FOR SOLAR PV IMPACT ASSESSMENT IN SMART GRID ENVIRONMENT

As stated earlier, huge computational effort will be necessary to perform PV impact assessment using the entire volume of data stored by numerous meters installed in a future smart grid environment. For example, the length of a voltage profile time series data measured at a 5-minute interval for a week at some arbitrary bus in a three phase feeder will be 2016, for each phase. If the feeder contains 100 buses, then the weekly voltage profile database will contain at least 100×2016 pieces of data for each phase. However, the analysis of the entire volume of the stored data may not be of significance for the investigation of solar PV impacts. This is due to the fact that the level of PV impact on the network behavior will depend on many factors, such as, the location of the PV cluster, penetration level, clearness of sky, sun insolation level, temperature, etc. Therefore, solar PV impacts, such as voltage rise or reverse power flow, may not be significantly visible in all the stored time series data. If the times and the PV locations can be found when the PV impacts are more intense, then detailed analysis can be carried out for those scenarios only.

To select the time series data of particular importance for PV impact assessment from a large database, data mining using clustering and pattern recognition techniques would be necessary. Such a method can identify hidden patterns (such as incidents of voltage rise) and anomalies in the time series database. However, performing this sort of data mining tasks on the original time series data will impose significant computational burden on the tool. A time series approximation technique that can reduce the dimensionality of the data while retaining its features at the same time can be useful in this purpose. Recently, a symbolization based time series approximation technique termed as Symbolic Aggregate Approximation (SAX) has attracted much attention of the data mining research community due to its capability to reduce the computational effort in data mining applications. The proposed tool in this paper will use SAX method to reduce

data dimensionality prior to applying data mining techniques. Once the patterns of interest and times of anomalies have been identified, these smaller set of database can be analyzed in greater details to understand why they occur, to help in predicting such future events, and to design corrective measures to reduce the impacts of the rooftop solar PV.

Usually, the detailed analysis is to be carried out using load flow analysis to further investigate the daylong behavior of distribution networks with time varying load and PV generation data. Adoption of a load flow calculation algorithm that needs less computational time and effort would be necessary in this part to make the analysis tool practically useful. Current-mismatch based Newton-Raphson algorithm [10] will be used in the proposed tool that will speed up the computational process.

The advantage of using SAX method in data mining through pattern recognition and the current-mismatch based Newton-Raphson load flow method will be described in the following subsections.

A. A Brief Review of SAX Method

Time series data consists of patterns of an event sampled over a time period at some specified interval. The length of the data would typically be huge, particularly if high resolution in measurement is used. Working with a huge amount of time series database for data mining applications would impose a high computational burden on the analysis tool. To reduce this burden, SAX method can be applied to symbolize the time series data to reduce its dimensionality. Moreover, it can ensure the lower bounding of the data. Both of these features will be explained below briefly [4].

Consider a time series data S of length n that has the elements s_1, s_2, \dots, s_n . The first step in SAX approximation is to normalize the original n -dimensional time series data and then to convert it into a w -dimensional time series T with elements t_1, t_2, \dots, t_w , where, typically, $w \ll n$. This can be performed using a Piecewise Aggregate Approximation (PAA) [11] of the original time series S using the following expression.

$$t_i = \frac{w}{n} \sum_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i} s_j \quad (1)$$

Once the PAA representation of the data is obtained using (1), the lower dimensional time series T is then transformed into discrete form through a number of equiprobable symbols. This is performed using Gaussian distribution, as normalized time series data usually show a highly Gaussian distribution characteristic [4]. To obtain the symbols having equal probability, breakpoints need to be defined that will divide the Gaussian curve into equal sized areas.

For example, if the PAA representation of a time series is desired to be symbolized using “ a ” number of alphabets, then $a-1$ number of break points have to be obtained that will divide the Gaussian curve into “ a ” number of equal sized areas. This method is demonstrated in Fig. 2 for a time series that is symbolized using 3 alphabets, that is, $a = 3$.

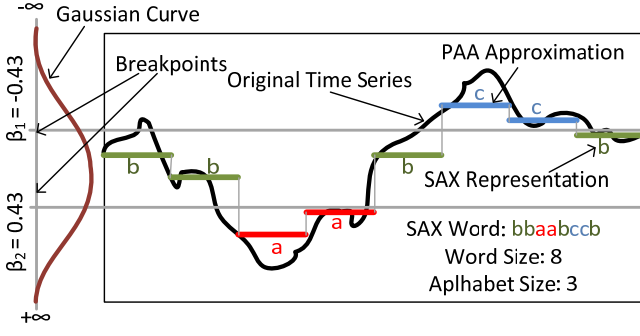


Fig. 2. SAX representation of time series data

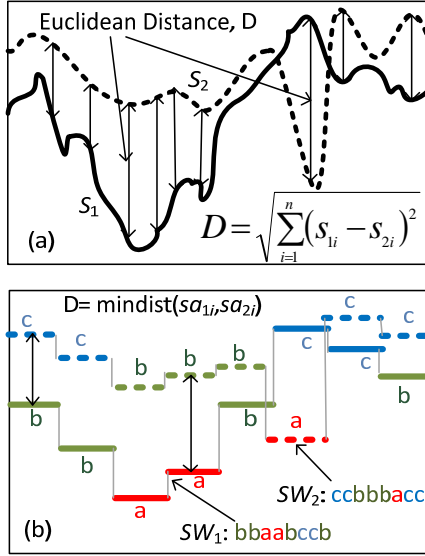


Fig. 3. Distance measure of time series data; (a) Euclidean distance for original data (b) Gaussian curve breakpoint distance for SAX word.

To perform this, the Gaussian curve $N(0,1)$ is divided into 3 areas using the breakpoints at β_1 and β_2 . Values of the breakpoints are obtained from standard normal distribution tables [4]. In general, if the symbolization is to be performed using “a” number of alphabets, the breakpoints of the Gaussian curve will be located at $\beta_1, \beta_2, \dots, \beta_{a-1}$.

Symbolization of the time series is performed in ascending order of breakpoints. For example, the PAA coefficients below the smallest breakpoint are symbolized with alphabet ‘a’; similarly, the PAA coefficients greater than or equal to the smallest breakpoint and less than the second smallest breakpoint is symbolized with alphabet ‘b’, and so on. The concatenation of all the symbols representing the PAA version of the original time series is defined as *word* and the total number of alphabet is called *word size* [12]. The total number of different alphabets used in the word is defined as *alphabet size* [12]. The word size represents the total number of segments into which the original time series will be divided and the alphabet size indicates the granularity of the data. Fig. 2 shows the SAX representation of an original time series using an 8-word and 3-alphabet SAX data.

Dimensionality reduction of original time series data using an approximation technique needs to be performed in such a

way that any essential feature of the data is not lost [5]. In time series data mining, distances among multiple time series are usually obtained using different types of distance measures. The most commonly used distance measure is the Euclidean distance [4, 12] as shown in Fig. 3(a), for two time series S_1 and S_2 . Essential features for data mining in the approximated data are retained if the distance measures of the approximated data are less than or equal to the distance measures of the original data [5]. This is defined as *Lower Bounding* of the distance. For the SAX representations of SW_1 and SW_2 in Fig. 3(b) of the original time series in Fig. 3(a), the lower bounding will ensure that,

$$D(SW_1, SW_2) \leq D(S_1, S_2) \quad (2)$$

where, $D(SW_1, SW_2)$ is the lower bounding distance measure of the SAX data and $D(S_1, S_2)$ is the distance between original time series data.

For the SAX representation, the lower bounding distance measure is obtained from the breakpoints of the Gaussian curve using a function named ‘mindist’ [4] defined as follows,

$$\text{mindist}(SW_1, SW_2) = \sqrt{\frac{n}{w} \sum_{i=1}^w [\text{dist}(sa_{1i}, sa_{2i})]^2} \quad (3)$$

where, sa_{1i} and sa_{2i} are the elements (alphabets) of the SAX words SW_1 and SW_2 , and ‘dist’ is the distance function whose values can be obtained by table-lookup method [4] from a table containing the breakpoints of the Gaussian curve used for symbolization of the original time series.

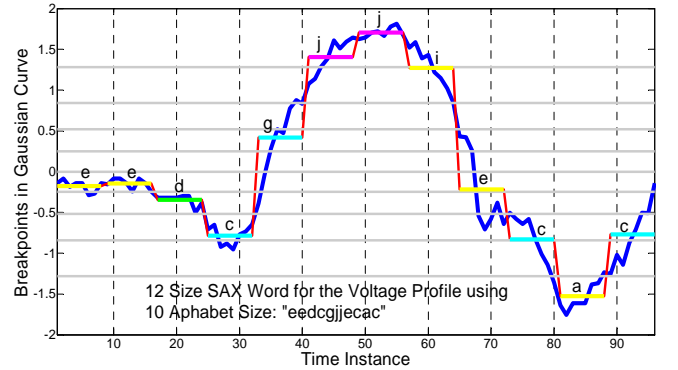


Fig. 4. SAX approximation of a time series voltage profile data.

B. SAX Representation of Network Behavior Data

The selection of the word and alphabet size for SAX representation is highly data dependent and may be infeasible to determine analytically [4]. Determination of the suitable word and alphabet size is therefore performed experimentally. It is anticipated that the smart meters will store network behavior data (e.g., voltage, current) in the normally used units (e.g., Volt, Ampere). To perform SAX conversion, this data has to be normalized to have values between *zero* and *one*. Then a suitable size of word and alphabet has to be determined experimentally to obtain SAX representations that reduce the dimensionality yet retains the main features of the data. An

example of SAX conversion of a daylong voltage profile of a distribution feeder node with solar PV is presented in Fig. 4.

C. Clustering of Patterns from the Time Series Datasets

Clustering of the time series data containing network behavioral patterns will be needed to find out which group of data is of interest for detailed impact assessment of solar PV.

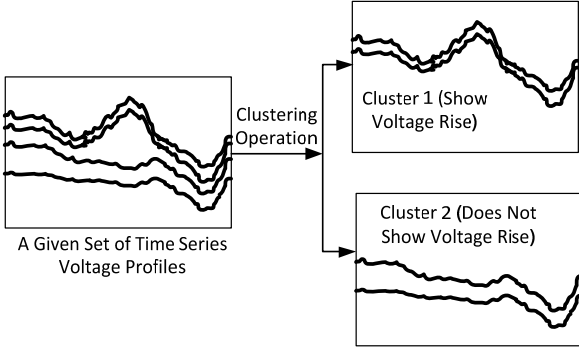


Fig. 5. Identification of voltage rise patterns using clustering technique

For example, Fig. 5 shows that a given set of time series patterns containing voltage profiles with and without rising trend can be separated into two different clusters. For solar PV impact assessment, the cluster containing voltage rise profiles will only be used for further analysis.

Several methods of time series clustering have been reported in literature [13]. The proposed tool will use the partitional clustering technique using k-means algorithm. The k-means technique is an iterative clustering algorithm that first determines the cluster centroids and then minimizes the sums of the data-point to cluster centroid distances, summed over all the clusters, iteratively [13]. For example, S_k , $k=1, 2, \dots, n$ represents n -number of time series patterns that needs to be grouped into c number clusters. Using k-means algorithm, c number of cluster centroids are first defined. An objective function, J , is then defined using the sums of data-point to cluster centroid distances over all clusters, as shown below [13].

$$J = \sum_{i=1}^c \sum_{k=1}^n \|S_k - v_i\|^2 \quad (4)$$

where, v_i , $i=1, 2 \dots c$ are c -number of cluster centroids. These cluster centroids are updated iteratively until J is minimized.

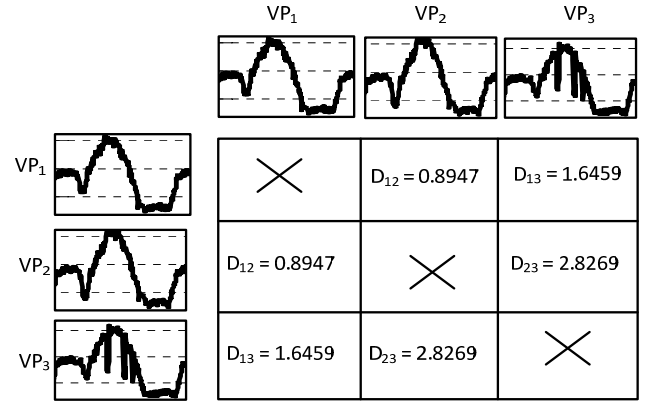
The number of points in each of the cluster centroids is equal to the length to the time series data. Therefore, the computational effort to calculate the cluster centroids and to perform the iterative adjustment is reduced when lower dimensional approximation of the time series is used. Application of k-means on SAX representation of data also produces better results as the initialization of the cluster centroids with a lower dimensional data improves the quality of the results [4].

D. Detection of Anomalies in Datasets

In some cases, time series data collected and stored by the

smart meters may contain unusual or exceptional network behavior. For example, passing clouds may introduce fluctuations in the smooth rising trend of the voltage profile time series data. Detection of the unusual patterns is necessary for performing further analysis. Standard algorithms for unusual pattern recognition can be used for this purpose. For simplicity, Brute Force [12] algorithm is used in this paper that measures the non-self match distances of each of the time series from each other and the time series that possesses the largest distance from its nearest match is detected as the most unusual pattern. This algorithm is explained below for finding the most unusual pattern from 3 voltage profiles.

The voltage profiles are arranged in a matrix form as shown in Fig. 6. For the 3 voltage profiles, size of the matrix is 3x3. The diagonal cells are left crossed as only the non-self match distances are of use for anomaly detection. The off-diagonal elements in each column give the non-self distances of the voltage profiles from each other. The smallest distance in each column provides the nearest match distance of each voltage profiles. The largest of the three smallest distances from the three columns in the matrix thus indicates the most anomalous voltage profile, which is, VP_3 . Similarly, the second most anomalous profile can be found using the second largest of the nearest match distance. In this way, a ranking of the anomalous time series profiles can be made.



VP = Voltage Profile, D = Distance between two voltage profiles

Fig. 6. Algorithm for finding unusual voltage profile

E. Time Series Power Flow Analysis with Solar PV

Once the time series with the patterns of interest and anomalies have been identified, power flow calculations have to be conducted. The formulation of current mismatch based load flow equation with solar PV output at k -th time instant, where $k = 1, 2, \dots, n$ (n is the length of time series data), is shown below. An arbitrary bus i of a b bus distribution feeder is considered, as shown in Fig. 7, where the load demands and PV generation at the i -th bus and k -th time instant is specified from the available system data. The specified current at the i -th bus in the k -th instant can be obtained using (5).

$$I_i^{spec}(k) = \frac{[P_i^{PV}(k) - P_i^L(k)] + j[Q_i^{PV}(k) - Q_i^L(k)]}{V_i^*(k)} \quad (5)$$

where,

$I_i^{spec}(k)$ = Specified current vector

$P_i^L(k)$ = Active load vector

$Q_i^L(k)$ = Reactive load vector

$P_i^{PV}(k)$ = PV active power output vector

$Q_i^{PV}(k)$ = PV reactive power output vector

$V_i(k)$ = Node voltage vector

The vectors mentioned here are corresponding to the i -th bus at the k -th time instant, and * indicates the complex conjugate operator.

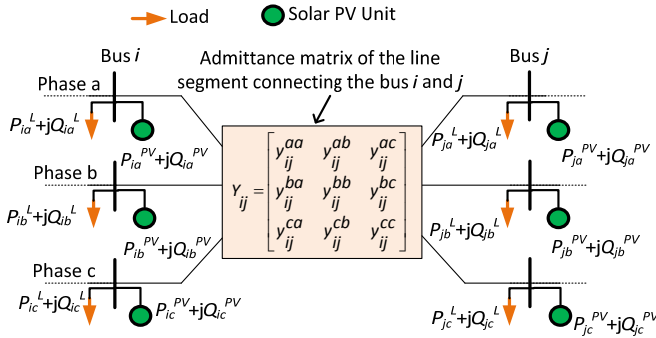


Fig. 7. Part of a three phase distribution feeder connected by two arbitrary buses i and j

Current at k -th time instant can also be calculated from network admittance matrix using,

$$I_i^{calc}(k) = \sum_{j=1}^b Y_{ij}(k) V_j(k) \quad (6)$$

where,

$I_i^{calc}(k)$ = Calculated current vector

$Y_{ij}(k)$ = Admittance matrix between i -th and j -th bus

$V_j(k)$ = Voltage vector at the j -th bus

If node voltage values used in (5) and (6) are correct, then the mismatch between specified and calculated current would be nearly zero, that is,

$$\Delta I_i(k) = I_i^{spec}(k) - I_i^{calc}(k) \approx 0 \quad (7)$$

Newton-Raphson method is applied at each time instant k to obtain the node voltages that satisfies the condition in (2), by relating the incremental changes in node current with incremental changes in node voltage with system Jacobian, as given below.

$$\Delta I_i(k) = J(k) \Delta V_i(k) \quad (8)$$

where,

$\Delta I_i(k)$ = incremental change in current resulting from mismatch between specified and calculated current at k -th time instant

$J(k)$ = system Jacobian matrix k -th time instant

$\Delta V_i(k)$ = incremental change in voltage used as the

update in voltage for power flow iterations corresponding to k -th time instant

The details of power flow iterations are not the main focus in this paper, rather it emphasizes only the computational advantage of current-mismatch based formulation of load flow equations. Details of this algorithm can be obtained in [10]. The elements of the Jacobian matrix shown in (8) can be obtained from the system admittance matrix with modifications of the diagonal blocks by adding contributions from loads [10]. Therefore, once the system admittance matrix is formed, it can be used for all the instances in the time series data, unless the network topology is changed. This reduces the computational burden for time series load flow analysis with a given set of load and PV generation patterns.

F. Flowchart of the Proposed Advanced Computational Tool

It is obvious that analysis of a large amount of measured data to assess the presence of network issues and concerns, and also the severity of the solar PV impacts would be highly computational intensive and time consuming. The analysis tool, therefore, need to have a “sense making” capability [14] to find out which part of the database is to be used for in-depth analysis. Voltage and power flow variations can be considered as the primary impacts that are directly caused by solar PV. Variations of power loss, voltage unbalance factor and tap changer operation are rather secondary impacts that are generally caused by the primary impacts. Detection of the changes in voltage and power flow patterns will therefore be used as the main criteria to identify the time series data that can be used for detailed analysis and extraction of the features.

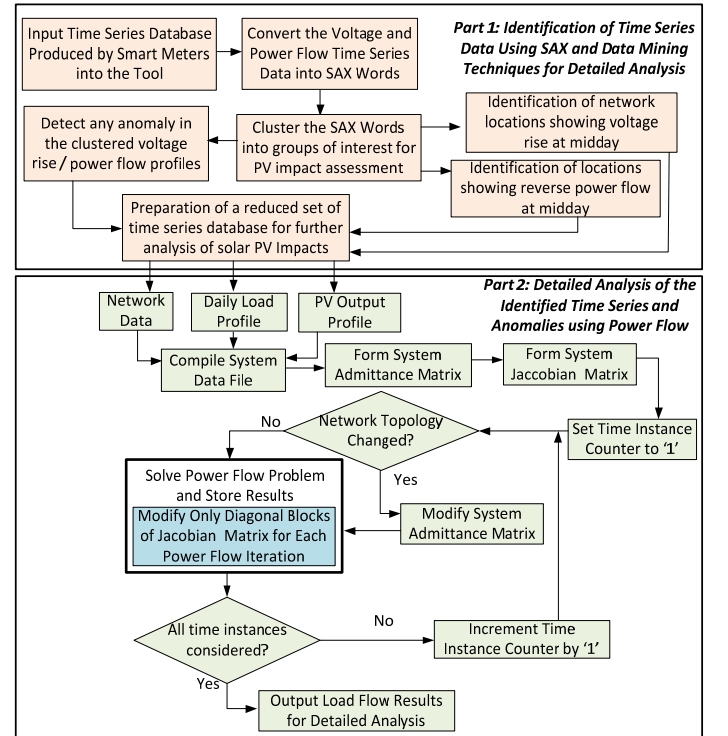


Fig. 8. Flowchart of the proposed analytical tool.

Functionality of the proposed tool consists of two main parts as shown in Fig. 8. In the first part, the large volume of voltage profile and time series data archived by smart meters will be taken as input. The data will be then be converted into SAX words prior to performing clustering and anomaly detection. Once the network locations and times of interest for PV impact analysis are obtained from the discovered patterns and anomalies, the corresponding load demand, PV generation and network data will be extracted from the large database to perform load flow analysis in the second part of the tool.

IV. APPLICATION EXAMPLE

The proposed tool has been tested on a practical distribution network in New South Wales, Australia. Results will be presented in the following subsections.

A. Test Network

The test distribution network consists of an 80 km 11 kV backbone with 18 spurs of different lengths in the range of 5 to 30 km. This rural network serves several sparsely populated areas. Three series voltage regulators are installed along the main feeder to control the voltage at remote ends. A simplified schematic diagram of the test network is shown in Fig. 9. The transformer symbols shown inside the PV cluster indicate the downstream LV feeder contains solar PV units. LV feeders are modeled using four-wire configuration and neutral wire is considered as solidly grounded at the customer service drop.

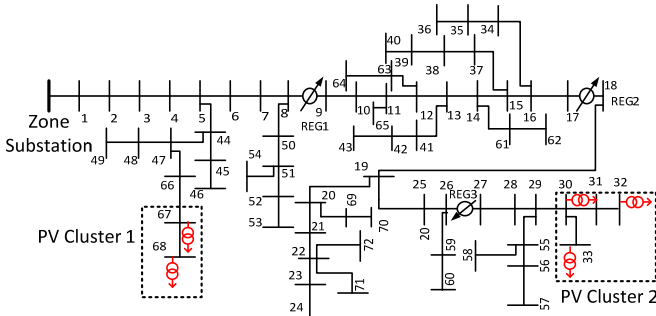


Fig. 9. Test distribution network with locations of PV clusters.

B. Dimensionality Reduction of Results Database using SAX

For a 24 hours voltage profile with 5-min resolution data, the length of the time series is 288. Using SAX, the voltage profile time series can be reduced to a lower dimension, w , where w is the size of the SAX word. The compression ratio achieved using SAX can be defined as,

$$\text{compression ratio} = \frac{n}{w} \quad (9)$$

where, n is the length of original data and w is the SAX word size. With a 20 word SAX representation, the compression ratio achieved for 1 day's voltage profile data is 14.4. Similarly, line flow data can be converted to SAX word and considerable reduction in data dimensionality can be achieved. Reduced data dimension will incur less computational effort in clustering and pattern recognition tasks to be performed on the data collected by smart meters.

C. Identification of Network Location Affected by Solar PV Clusters

Voltage profile time series data was clustered into 4 groups using k-means algorithm and the results are presented in Fig. 10. Clustering results of the original raw data are shown in Fig. 10(a) and that from SAX representations are shown in Fig. 10(b). It is observed in Fig. 10(a) that all of the four clusters created from the original data include voltage rise behavior, and cluster 1 contains a mix (low and high) of voltage rise patterns. However, Fig. 10(b) shows that based on SAX representation of the data, the profiles with significant voltage rise are now grouped into cluster 4 only. The node locations corresponding to these voltage profiles can be identified using the cluster indices and further analysis of voltage rise now can be performed.

Similarly, daylong power flow profiles from the system nodes have been clustered using the original data as shown in Fig 11(a) and using the approximated time series data shown in Fig 11(b). Cluster 3 in Fig. 11(a) shows that the reverse power flow profiles due to solar PV have been clustered together with the ones that have not been affected by PV operation. On the other hand, when clustered using time series data obtained from SAX approximation, Fig. 11(b) shows that all the time series dataset with reverse power flow profiles at midday have been grouped into cluster 2 only.

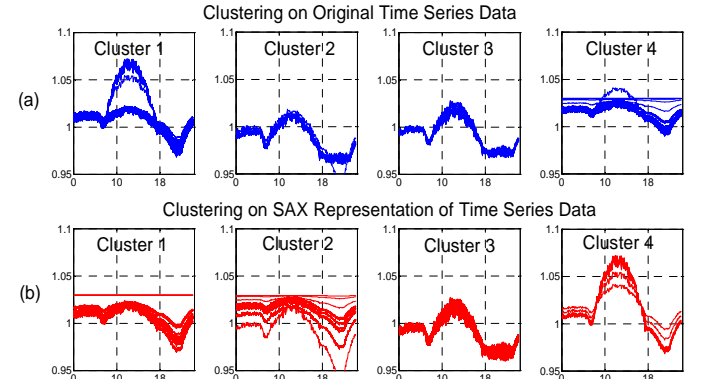


Fig. 10. Clustering of voltage profiles for identification of network locations showing voltage rise: (a) using original time series data (b) using SAX approximation

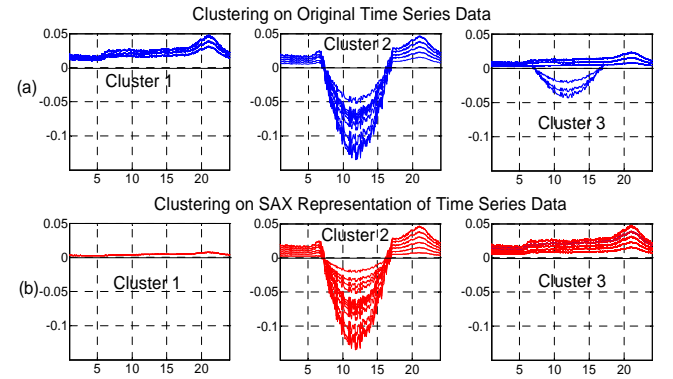


Fig. 11. Clustering of power flow profiles for identification of network locations showing reverse power flow: (a) using original time series data (b) using SAX approximation

The main advantage of applying SAX was found in reduction of computational time. A comparison was made on average runtime of k-means routine for 10 consecutive runs with original and SAX data. Reduction in elapsed time was about 16 times and 10 times for voltage and power flow profiles, respectively.

D. Detection of Abnormal Voltage Profile

Consecutive 14 days voltage profile time series data of one of the network buses affected by solar PV were analysed as 14 time series patterns. The original time series data of each day's voltage profile was converted into SAX words. The distances of each of the SAX time series from each other were calculated to determine the time series that possesses the largest distance from its nearest non-self match. For simulation tests, cloud passing effects have been embedded in the time series data for two of the 14 days considered. Detection of the voltage profile anomaly due to distortion by cloud passing effect was performed successfully, as shown in Fig. 12 below.

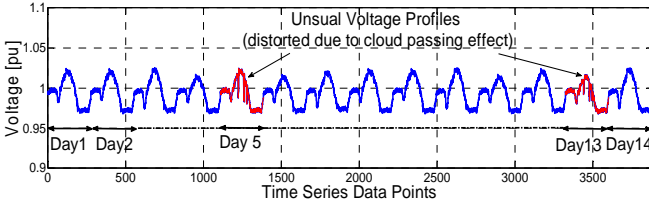


Fig. 12. Detection of unusual voltage behaviour.

E. Detailed Analysis of Identified Dataset Related to Solar PV Impacts

Once the patterns and anomalies of interest for PV impact assessment have been obtained from a large set of time series database, more detailed analysis can be carried out. The application of such detailed analysis may include determination of better network operation strategy with high PV penetration, or, design of corrective actions to mitigate PV impacts, etc. The type and methodology of the next stage of analysis will depend on its application. Load flow solution will be commonly required to perform such detailed analysis. For brevity of space, this paper will only show an example related to the PV impact on upstream voltage regulator operation.

Clustering of voltage profiles of the test system has identified that the LV feeders downstream to the third voltage regulator is experiencing voltage rise impact. Based on this information, it is important to investigate how the load center voltage of this regulator is being affected by the voltage rise in the downstream LV feeders and how is this affecting the tap changer operation. The voltage profile at the load centre of this regulator was studied and presented in Fig. 13(a) that shows the voltage is marginally equal to the upper limit set in the controller. This makes an intuitive suggestion that tap operation may have been performed by the regulator to keep the voltage within upper limit and this is verified by the tap operation profile shown in Fig. 13 (b). With the load and PV generation data corresponding to this scenario available at hand, load flow analysis was performed to investigate how the

tap operation would be without solar PV. The load center voltage without PV is shown in Fig. 13(a) that shows the voltage is well below the threshold and no tap operation would have been performed without PV.

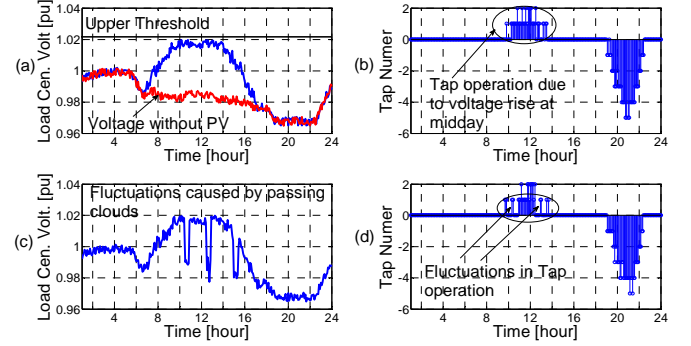


Fig. 13. Impact of midday voltage rise on tap changer operation: (a) Voltage profile at the load center of the third voltage regulator (b) Tap operation corresponding to the voltage profile in Fig. 13(a) (c) Fluctuations introduced by passing clouds in the load center voltage (d) Tap operation corresponding to the voltage profile in Fig. 13(c).

Passing clouds introduce fluctuations in the distribution feeder voltage with solar PV and may also produce fluctuating impact in the tap changer operation. The load and PV generation data corresponding to the most anomalous day identified in Fig. 12 by the anomaly detection technique were used in the load flow analysis. Tap operation of the third voltage regulator with this data was examined. Passing cloud induced fluctuations were observed at the load center voltage, as shown in Fig. 13(c), and fluctuation in tap operations were observed at the same time periods, as shown in Fig. 13(d). It is to be noted that the voltage fluctuation at about 15:00 hours did not impact the regulator operation as voltage rise at this moment was not so high to cause tap changer to operate.

Since archived real-time data are used in the PV impact analysis, more accurate results should be obtained when compared to those using assumption based synthetic data. This will help utility planning engineers to understand the PV impacts on the distribution systems better. The proposed approach can be used to track and visualize the identified PV impacts of interest, and also identify unique signatures within the utility system. This can be achieved using the three-phase load flow analysis, capable of dealing with network asymmetry, load unbalance, single-phase solar PV integration, and their impacts on upstream network, described in Section III.E. Using the unique signatures of the identified PV impacts, the proposed approach can assist in providing distribution system infrastructure security via early warning systems of solar PV impacts in real-time.

V. CONCLUSION

An advanced analysis tool for the assessment of clustered rooftop solar PV impacts in the future distribution networks under smart grid environment is proposed. With a large volume of real time network data collected from monitoring systems or smart meters, this tool will be able to intelligently

decide which part of the data would be effective for analysis of solar PV impacts on distribution networks. To identify the hidden patterns and anomalies in the archived time series database, this tool will apply data mining techniques. However, to reduce the computational efforts while applying the pattern discovery algorithms, the proposed tool will use SAX representation to reduce the dimensionality of the database. Further, the tool will also deploy a load flow algorithm demanding less computational effort while performing in-depth analysis of the identified data of interest. A practical distribution network in Australia was used in this paper to verify the applicability of the proposed computational tool. It was observed that the developed tool is able to identify the useful portions of the large database for further detailed analysis and feature extraction to characterize the response. It is envisaged that the proposed tool can identify the unique signatures associated with different types of PV impacts which can be used in real-time for warning of the PV impacts in the networks.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support and cooperation of Essential Energy and Endeavour Energy personnel for providing practical system information and data.

REFERENCES

- [1] S. Morozumi, N. Inoue, Y. Arashiro, Y. Chiba, and T. Iwasaki, "Strategies and status of grid-connection Technology Development in NEDO," in *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, Pittsburgh, PA, USA, 20-24 Jul., 2008, pp. 1-6.
- [2] R. A. Walling, R. Saint, R. C. Dugan, J. Burke, and L. A. Kojovic, "Summary of Distributed Resources Impact on Power Delivery Systems," *Power Delivery, IEEE Transactions on*, vol. 23, pp. 1636-1644, 2008.
- [3] Y. Liu, J. Bebic, B. Kroposki, J. de Bedout, and W. Ren, "Distribution System Voltage Performance Analysis for High-Penetration PV," in *Energy 2030 Conference*, Atlanta, GA, USA, 17-18 Nov., 2008, pp. 1-8.
- [4] J. Lin, E. Keogh, L. Wei, and S. Lonardi, "Experiencing SAX: a novel symbolic representation of time series," *Data Mining and Knowledge Discovery*, vol. 15, pp. 107-144, 2007.
- [5] E. Keogh, "Mining shape and time series databases with symbolic representations," in *13th International Conference in Knowledge Discovery and Data Mining*, San Jose, CA, USA, 12-15 Aug., 2007.
- [6] H. Mori and Y. Umezawa, "A SAX-based method for extracting features of electricity price in power markets," in *Transmission & Distribution Conference & Exposition: Asia and Pacific*, Seoul, Korea, 26-30 Oct., 2009, pp. 1-4.
- [7] E. Keogh, J. Lin, F. Ada Waichee, and H. Van Herle, "Finding Unusual Medical Time-Series Subsequences: Algorithms and Applications," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 10, pp. 429-439, 2006.
- [8] V. H. M. Quezada, J. R. Abbad, and T. G. S. Roman, "Assessment of energy distribution losses for increasing penetration of distributed generation," *Power Systems, IEEE Transactions on*, vol. 21, pp. 533-540, 2006.
- [9] F. Shahnia, R. Majumder, A. Ghosh, G. Ledwich, and F. Zare, "Sensitivity analysis of voltage imbalance in distribution networks with rooftop PVs," in *Power and Energy Society General Meeting*, Minneapolis, MN, USA, 25-29 Jul., 2010, pp. 1-8.
- [10] P. A. N. Garcia, J. L. R. Pereira, S. Carneiro, Jr., V. M. da Costa, and N. Martins, "Three-phase power flow calculations using the current injection method," *Power Systems, IEEE Transactions on*, vol. 15, pp. 508-514, 2000.
- [11] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Locally adaptive dimensionality reduction for indexing large time series databases," in *ACM SIGMOD Conference on Management of Data*, Santa Barbara, CA, USA, May. 2001, pp. 151-162.
- [12] L. Wei, E. Keogh, X. Xi, and M. Yoder, "Efficiently finding unusual shapes in large image databases," *Data Mining and Knowledge Discovery*, vol. 17, pp. 343-376, 2008.
- [13] T. Warren Liao, "Clustering of time series data--a survey," *Pattern Recognition*, vol. 38, pp. 1857-1874, 2005.
- [14] G. Venayagamoorthy, "Dynamic, Stochastic, Computational, and Scalable Technologies for Smart Grids," *Computational Intelligence Magazine, IEEE*, vol. 6, pp. 22-35, 2011.

BIOGRAPHIES



M J E Alam (Std. Member'10) received B.Sc. and M.Sc. Degree in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, in 2005 and 2009, respectively, with focus on electrical energy and power systems.

At present he is conducting PhD research at the University of Wollongong, New South Wales, Australia. Prior to starting PhD studies, he has been involved with the power industry in Bangladesh for 4.5 years, where he worked in the field of power generation, transmission and distribution.

His research interest includes modeling and analysis of power systems considering the impacts of distributed and renewable energy resources.



K M Muttaqi (M'01, SM'05) received the Ph.D. degree from Multimedia University, Malaysia, in 2001. Currently, he is an Associate Professor at the School of Electrical, Computer, and Telecommunications Engineering, University of Wollongong, Wollongong, Australia. He was associated with the University of Tasmania, Australia as a Research Fellow/Lecturer/Senior Lecturer from 2002 to 2007, and with the Queensland University of Technology, Australia as a Research Fellow from 2000 to 2002. Previously, he

also worked for Multimedia University as a Lecturer for three years. His special fields of interests include distributed generation, renewable energy, power system planning, intelligent grid, and power system reliability.



D Sutanto (SM'89) obtained his BEng. (Hons) and PhD from the University of Western Australia. He is presently the Professor of Power Engineering at the University of Wollongong, Australia. His research interests include power system planning, analysis and harmonics, FACTS and Battery Energy Storage systems. He is a Senior Member of IEEE. He is currently the IEEE IAS Area Chair for Region 10 (Asia Pacific).