Off-Grid Parameters Analysis Method Based on Dimensionality Reduction and Self-organizing Map

Tomas Burianek, Tomas Vantuch, Jindrich Stuchly and Stanislav Misak

Abstract Off-Grid systems are energetic objects independent of an external power supply. It uses primarily renewable sources what causes low and variable short-circuit power that must be controlled. Also in the Off-Grid system is a need to keep power quality parameters in requested limits. This requires a power quality parameters forecast and also a forecast of electric energy production from renewable sources. For these forecasts a complex analysis of Off-Grid parameters is an important task. This paper proposes method that processes data set from the Off-Grid system using Dimensionality Reduction and the Self-organizing Map to obtain relations and dependencies between power quality parameters, electric power parameters and meteorological parameters.

Keywords Self-organizing map · Principal component analysis · Time-series · Clustering · Off-grid system

1 Introduction

The Off-Grid System is a system independent of the power supply from external grids. Its specific characteristic is a low and variable short-circuit power [8] in comparison with the On-Grid systems. This is given mainly by a character of a source part because the Off-Grid systems use primarily renewable sources (RES). These

T. Burianek (✉) · T. Vantuch · J. Stuchly · S. Misak
Department of Computer Science, Department of Electrical Power Engineering,
VSB - Technical University of Ostrava, FEECS,
17. listopadu 15, 708 33 Ostrava-Poruba, Czech Republic
e-mail: tomas.burianek.st1@vsb.cz

T. Vantuch
e-mail: tomas.vantuch@vsb.cz

J. Stuchly
e-mail: jindrich.stuchly@vsb.cz

S. Misak
e-mail: stanislav.misak@vsb.cz

235

RES have a stochastic character and in a conjunction with storage devices are only sources of a short-circuit power. This problem with a stochastic character of RES power supply is possible to solve by a storage device, whereas the power management in the Off-Grid is controlled by using sophisticated artificial methods for example by Demand Side Management (DSM) [17]. However, another problem related to variable and a low short-circuit power is keeping power quality parameters (PQP) in requested limits. PQP are defined by the national and international standards, norms [1, 9]. (i) Level of supplied voltage, (ii) supplied voltage frequency, (iii) periodical fluctuation of voltage as well as (iv) total harmonic distortion (THD) is possible to mention from a primary PQP. These above mentioned PQP is needful to keep in defined limits for a restraint of a reliable and safe running all of the appliances in the system. PQP out of desired limits may cause reduction of a service life of the appliances and in the worst case may cause damage. This brings a need of PQP controlling implementation to the DSM conception. This is possible to realize only with understanding of a PQP progress in a time interval (PQP forecasting) and relations between PQP and other Off-Grid parameters such as electrical parameters and meteorological conditions. The developing of tools for PQP forecasting (predictors) is very important for restraint of reliable and safe Off-Grid system running whereas the developing of predictors is possible to realize only in a case of detailed analysis of relations between PQP and other Off-Grid parameters, how was mentioned.

An analysis of a time-series data is an important field of a data mining [5]. Each time-series that consists of many data points could be seen as a single object. In retrieval of relations between those complex objects a clustering methods are important tools for analysis. There are many algorithms used for a time-series clustering. The basic k-means clustering algorithm is one of the most used for this task. The k-means assigns objects to k clusters, where k is a user defined parameter [2]. Another is the Hierarchical clustering algorithm that creates a nested hierarchy of corresponding groups based on the Euclidean distance between pairs of objects [19]. The Self-organizing Map is a clustering technique that provides a mapping from a high dimensional input space to a two dimensional grid [12].

This paper proposes method for the Off-Grid parameters analysis based on the clustering that employs the Self-organizing Map (SOM). Each parameter time-series results to high-dimensional data with redundant information. Therefore a dimensionality reduction method is applied on the data to extract only important features before presenting to the Self-organizing Map. The Self-organizing Map then provides mapping from the reduced data set to the two-dimensional grid. To obtain relevant clusters from map the k-means clustering algorithm adapted to the SOM is then involved.

Paper is organized as follows: Sect. 2 describes the dimensionality reduction with the Principal Component Analysis method in more details, Sect. 3 describes the Self-organizing Map method and in Sect. 4 the k-means clustering of the Self-organizing Map is explained. Method for the Off-Grid parameters analysis is proposed in Sect. 5 and the experiments and results obtained by proposed method are discussed in Sect. 6. The conclusion about results and future work are given in Sect. 7.

2 Dimensionality Reduction

The problem of dimensionality reduction for a set of variables $X = \{x_1, \dots, x_n\}$ where $x_i \in \mathbb{R}^D$ is to find a lower dimensional representation of this set $Y = \{y_1, \dots, y_n\}$ where $y_i \in \mathbb{R}^d$ and where $d < D$ (often $d \ll D$) in such a way to preserve content of the original data as much as possible [3]. For dimensionality reduction task several methods may be used such as Principal Component Analysis (PCA), Stochastic Neighbor Embedding (SNE), Factor Analysis (FA), Diffusion Maps, Sammon Mapping, Autoencoder, Neighborhood Preserving Embedding (NPE) [15, 21]. In the next subsection the Principal Component Analysis is described in more details.

2.1 Principal Component Analysis

Principal component analysis (PCA) is a statistical linear method that provides the dimensionality reduction. Dimensionality reduction is done by embedding the original data into linear subspace with lower dimension in a way to preserve as much of a relevant information as possible [10, 18]. It finds a mapping M to lower dimension data with maximal variance. This is done by solving equation of a eigenproblem [7].

$$cov(X)M = \lambda M, \tag{1}$$

where $cov(X)$ is the covariance matrix of a input data X. The mapping matrix M is orthogonal and is formed by principal components - eigenvectors of the covariance matrix $cov(M)$. The matrix λ contains eigenvalues of the covariance matrix on diagonal. To provide the dimensionality reduction a columns of the mapping matrix M are sorted according to the decreasing eigenvalues in the matrix λ. The mapping matrix is then truncated to keep only first d principal components. A new data set Y reduced into the dimension d is then computed by Eq. 2.

$$Y = XM_d. \tag{2}$$

Principal Component Analysis was successfully applied in the area of feature extraction and signal processing [20].

3 Self Organizing Map

The Self-organizing Map (SOM) was introduced in 1982 by Kohonen [12]. This method performs a nonlinear projection mapping from a higher dimensional input space into a lower dimensional output grid of prototype nodes called map [11]. It is related to the classical Vector Quantization (VQ) [13]. The output map is usually two-dimensional that is easy for a graphical visualization and analysis.
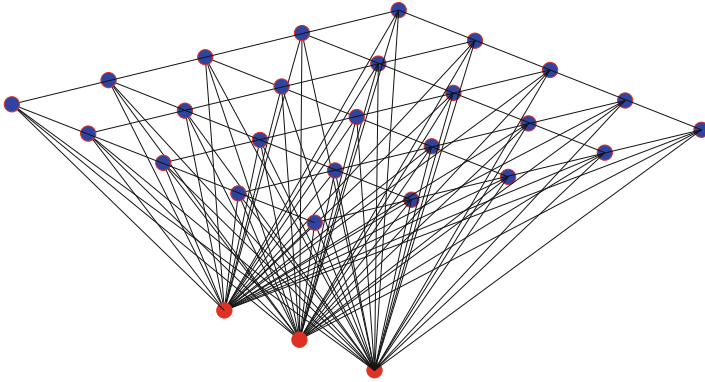
Fig. 1 Self-organizing Map structure with 5×5 map grid and 3 input nodes connected with map by weighted connections

The idea of method is that a similar data items corresponds to the prototype nodes closely related in the grid and a less similar data models are farther to each other. The Self-organizing map is based on an artificial neural network trained by an unsupervised competitive learning process. The model structure consists of map of prototype nodes, where each node has its position in the map grid and has a weigh vector of the same size as the input data space. The weight vectors preserves mapping from the high dimensional input space into the lower dimensional grid space. The map grid is usually organized into a hexagonal or a rectangular shape. Whole structure of the SOM could be presented by Fig. 1 as a two-layer neural network, where an output layer forms the grid of nodes and an input layer are input nodes, where each grid node is connected with an input node by the weighted connection.

The primary aim of the SOM learning is to optimize the weights of nodes. Thereby the map structure is organized in a way to respond to the similar input data items in the close related map prototype nodes.

Process of the training starts with initialization of all weights. It could be proceed by initialization to small random values or by initialization by random samples from the input data set. After initialization there is a repeated process of adaptation. From the input data set with the size n an input vector x_i where $0 \leq i < n$ is presented to the map by computing Euclidean distance to each node's weight vector. The input item i could be presented in ordered manner or by another method, for example could be randomly picked from the input data set. The node b with a least distance to the input vector is considered as a best matching unit (BMU). Then weights of the BMU node and nodes close to it on the map grid are adapted according to the input vector. The weight adaptation process of each node j in map with m nodes is depicted by Eq. 3.

$$w_j(t + 1) = w_j(t) + \alpha(t)\eta(t, j, b)(x_i - w_j(t)). \tag{3}$$

The new weight vector for the next step $t+1$ is computed from the previous weight in the step t that is adjusted towards the input vector x_i. A magnitude of this adaptation depends on a monotonically decreasing learning coefficient $\alpha(t)$ and a neighborhood function $\eta(t, j, b)$. The neighborhood function $\eta(t, j, b)$ captures influence of the BMU node b to the current node j in the step t based on a distance between nodes on the output map grid. For the neighborhood function could be used several functions but the most common is the Gaussian function. The neighborhood influence is decreasing during the adaptation process. The adaptation process stops after selected number of steps.

4 SOM Clustering Using K-Means Algorithm

Clustering of the Self-organzing Map could be done by a hierarchical or partitive clustering algorithms [22]. In this work the partitive clustering algorithm called k-means is involved.

The k-means algorithm that was first introduced in [16] is well-known clustering method. This algorithm minimizes an error function:

$$E = \sum_{i=1}^{k} \sum_{x \in S_i} \|x - c_i\|^2 , \tag{4}$$

where k is a number of clusters and c_i is the center of cluster S_i. Computation of error is done by sum of squared distances of each data point x in each cluster S. The algorithm of k-means iteratively computes partitioning for the data and updates cluster centers based on the error function.

In Self-organizing map the prototype nodes are used for clustering instead of all input data set. In this approach, the number of clusters k is unknown. Therefore k-means algorithm is run multiple times for each $k \in \langle 2, \sqrt{N} \rangle$ where N is number of samples and the best k settings is selected based on the Davies-Boulding index [4] calculated for each k clustering.

5 Proposed Off-Grid Parameters Analysis Method

To analyze relations and dependencies of parameters in the Off-Grid system, an algorithm consisting of several steps is proposed. A block diagram in Fig. 2 describes four main steps.

In the first step a data set is constructed from parameters, where each row is a time-series of each parameter. Each parameter has different range of values. Then each row is normalized to the same range in $\langle 0, 1 \rangle$.
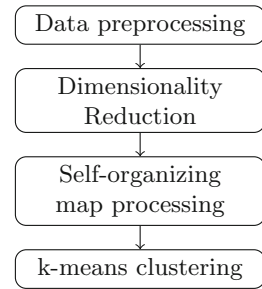
The second step consist of dimensionality reduction of the data set to a specific dimension. In this step only important features are extracted from the original high-dimensional data and speeds up the following SOM adaptation. The Principal Component Analysis is used for task of dimensionality reduction. The specific dimension could be obtained by computing intrinsic dimension of the data set. The Maximum Likelihood Estimation [14], the eigenvalues of PCA [6] or other method may be used for this task.

In the next step the reduced data set is processed by the Self-organizing Map. Each map prototype node is labeled by parameter names, where each node may have 0, 1 or more labels. Labeled map could be used for visual analysis with the idea that parameters of the same or near nodes in the map are related to each other. This visual analysis of groups of parameters could be supported by visualization of the U-matrix, which shows distances between map nodes in the input space by coloring associated neighboring cells.

The last step performs identification of parameter groups - clusters. For this task the k-means clustering algorithm adapted to the Self-organizing Map from Sect. 4 is involved. Obtained clusters are then analyzed to reveal important relations in the Off-Grid parameter data set.

6 Experiments and Results

The experiments of the Off-Grid Parameters Analysis method were performed on the data set obtained from the Off-Grid system. This data set consists of 117 power quality, meteorological and electric power parameter time-series. Each time-series is formed by 14400 one-minute ticks.

Input settings of proposed method are showed in Table 1. Selected dimensionality reduction method was PCA described in Sect. 2.1. Setting of other parameters of the SOM are based on designed map structure and performed experiments. The results obtained by the proposed method correspond with the original idea of relations between meteorological parameters, power parameters and basic power quality parameters. Figure 3 shows on the right side the Self-organizing Map labeled by

Table 1 Table of settings

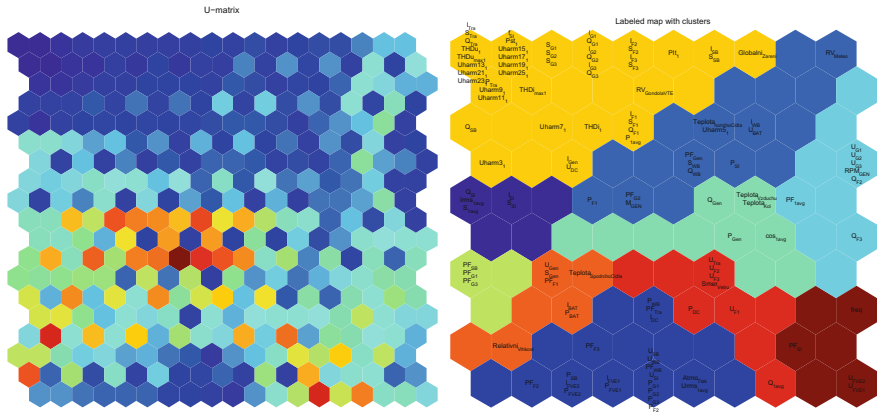| Dimensionality Reduction Method	PCA
Intrinsic Dimensionality Method	MLE
SOM Map Size	10×10
SOM Lattice	Hexa
SOM neighborhood	Gaussian
SOM learning coefficient α	0.5
SOM epochs	1000

Fig. 3 Left figure represents U-matrix visualization, right figure shows clusters and labels in map

names of all parameters and on the left side the visualized U-matrix that gives notion about distances between prototype nodes in the input space and its distribution. Clusters computed by the k-means algorithm are highlighted in the right figure by distinct colors of cells. The resulting groups of parameters are shown in Table 2.

Cluster 1 associates voltage on each part of Photovoltaic power lant (PV) string U_{FVE_1} and U_{FVE_2} which are in relation with frequency *freq* in the Off-Grid System and subsequently affect PF_{SI} - a power factor of Off-Grid inverter. Management of Off-Grid Inverter (SI) has ability to limit output power from the renewable sources. If parameter *freq* rises to defined bound 50.5 Hz(50 Hz is nominal), Photovoltaic power plant Inverter (PVI) overpasses out of an ideal point of Maximum power point tracking (MPPT) and increases voltage on its DC inputs to value U_{OC} what is an open circuit voltage of PV arrays. In this case when consumption of electric energy in the Off-Grid system is minimal an appliances are in the stand-by mode and the parameter PF_{SI} converges to zero, because connected appliances that have almost pure capacitive character caused by switching sources. Results in this cluster correspond with this fact.

Cluster 2 collects important power quality parameters together with power and meteorological variables. These parameters associate internal linkages of the system

Table 2 Table of clusters

Cluster 1	$PF_{SI}, freq, U_{FVE2}, U_{FVE1}$
Cluster 2	$I_{Tra}, S_{Tra}, Q_{Tra}, THDu_1, THDu_{max1}, Uharm13_1, Uharm21_1, Uharm23_1, P_{Tra}, Uharm9_1, Uharm11_1, Q_{SB}, Uharm3_1, f_{SI}, Pst_1, Uharm15_1, Uharm17_1, Uharm19_1, Uharm25_1, THDi_{max1}, S_{G1}, S_{G2}, S_{G3}, Uharm7_1, I_{Gen}, U_{DC}, I_{G1}, Q_{G1}, I_{G2}, Q_{G2}, I_{G3}, Q_{G3}, THDi_1, I_{F2}, S_{F2}, I_{F3}, S_{F3}, RV_{GondolaVTE}, I_{F1}, S_{F1}, Q_{F1}, P_{1avg}, PIt_1, I_{SB}, S_{SB}, Globalni_{Zareni}$
Cluster 3	$Q_{SI}, Irms_{1avg}, S_{1avg}, I_{SI}, S_{SI}$
cluster 4	$PF_{F2}, P_{SB}, I_{FVE2}, P_{FVE2}, PF_{F3}, I_{FVE1}, P_{FVE1}, P_{WB}, PF_{Tra}, I_{DC}, U_{SB}, U_{WB}, PF_{WB}, U_{SI}, P_{G1}, P_{G2}, P_{G3}, P_{F2}, P_{F3}, Atmo_{Tlak}, Urms_{1avg}$
Cluster 5	$P_{F1}, PF_{G2}, M_{GEN}, PF_{Gen}, S_{WB}, Q_{WB}, Teplota_{hornihoCidla}, Uharm5_1, P_{SI}, I_{WB}, U_{BAT}, RV_{Meteo}$
Cluster 6	$PF_{1avg}, U_{G1}, U_{G2}, U_{G3}, RPM_{GEN}, Q_{F2}, Q_{F3}$
Cluster 7	$Q_{Gen}, P_{Gen}, Teplota_{Vzduch}, Teplota_{Kol}, cos_{1avg}$
Cluster 8	$PF_{SB}, PF_{G1}, PF_{G3}$
Cluster 9	$Rel_{Vlhkost}, U_{Gen}, S_{Gen}, PF_{F1}, I_{BAT}, P_{BAT}, Tep_{SpoCidla}$
Cluster 10	$P_{DC}, U_{Tra}, U_{F2}, U_{F3}, Smer_{Vetru}, U_{F1}, Q_{1avg}$

in a complex plane. The reason of this are relations between meteorological parameters($Globalni_{Zareni}$ and $RV_{Gondola-VTE}$) and power variables such as reactive and apparent power components (Q_{SB} and S_{SB}). The system reacts to the power variables by change of the power quality parameters such as a negative changes of harmonics of higher orders (U_{Harm7} - U_{Harm25}), overall harmonic voltage and current distortion (*Thdu* and *Thdi*) together with short term and long term flicker severity(Pst_1 and Plt_1). Clusters such as Cluster 3, Cluster 4, Cluster 5 and other confirm previously stated ideas about mutual interactions of parameters and even revealed relations that in common system parameters evaluation are not evident. Example for this behavior could be group of parameters in Cluster 4 where PV power output (P_{FVE2} and P_{FVE1}) and current of individual PV strings (I_{FVE2} and I_{FVE1}) and AC power output of PV inverter (P_{SB}) affect atmospheric pressure ($Atmo_{Tlak}$) what makes sense in context with change of weather. Change of the atmospheric pressure leads to the change of energy production from PV. With the high atmospheric pressure a good meteorologic conditions are expected (sun shining) and PV production is near to a maximal possible power.

7 Conclusion and Future Work

In this work the Off-Grid Parameter Analysis Method was proposed. This method has four main steps: preprocessing of the data, dimensionality reduction by Principal Component Analysis, processing by the Self-organizing Map and clustering of map by the k-means algorithm. Results are presented in Fig. 3 that shows the labeled map

with highlighted clusters. List of 10 obtained clusters with named parameters are showed in Table 2.

These results corresponds with the original idea of authors, that measured parameters in the Off-Grid system have an effect on each other. It means that there exist direct links between meteorological parameters, electric power parameters and parameters of power quality. One of the main objectives was analysis and verification of these relations. Results of this analysis will be important base for power quality parameters forecasting, forecasting of electric energy production from the renewable sources and also as a tool for prediction of consumer behavior in given energetic objects. All these tools will be associated in sophisticated dispatching system called the Active Demand Side Management for a complex control of energy flows in the Off-Grid systems.

References

1. IEEE Recommended Practice for Monitoring Electric Power Quality. IEEE Std 1159, c1–81 (2009) (Revision of IEEE Std 1159–1995)
2. Bradley, P.S., Fayyad, U.M.: Refining initial points for k-means clustering. ICML **98**, 91–99 (1998)
3. Cunningham, P.: Dimension reduction. In: Machine Learning Techniques for Multimedia, pp. 91–112. Springer, Berlin (2008)
4. Davies, D.L., Bouldin, D.W.: A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. **1**(2), 224–227 (1979)
5. Esling, P., Agon, C.: Time-series data mining. ACM Comput. Surv. **45**(1), 12:1–12:34 (2012)
6. Fan, M., Gu, N., Qiao, H., Zhang, B.: Intrinsic dimension estimation of data by principal component analysis. CoRR abs/1002.2050 (2010)
7. Francis, J.G.F.: The qr transformation a unitary analogue to the lr transformation part 1. Comput. J. **4**(3), 265–271 (1961)
8. Goksu, O., Teodorescu, R., Bak-Jensen, B., Iov, F., Kjr, P.: An iterative approach for symmetrical and asymmetrical short-circuit calculations with converter-based connected renewable energy sources. Application to wind power. In: Power and Energy Society General Meeting, 2012 IEEE, pp. 1–8 (2012)
9. Ji, N.: Steady-state signal generation compliant with iec61000-4-30: 2008. In: 22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013), pp. 1–4 (2013)
10. Jolliffe, I.: Principal Component Analysis. Springer Series in Statistics, Springer, Berlin (2002)
11. Kohonen, T.: The self-organizing map. Proc. IEEE **78**(9), 1464–1480 (1990)
12. Kohonen, T.: Self-organized formation of topologically correct feature maps. Biol. Cybern. **43**(1), 59–69 (1982)
13. twenty-fifth Anniversay Commemorative Issue Essentials of the self-organizing map. Neural Networks. **37**(0), 52–65 (2013)
14. Levina, E., Bickel, P.J.: Maximum Likelihood Estimation of Intrinsic Dimension. In: NIPS (2004)

15. van der Maaten, L.J., Postma, E.O., van den Herik, H.J.: Dimensionality reduction: A comparative review. Journal of Machine Learning Research **10**(1–41), 66–71 (2009)
16. MacQueen, J.B.: Some methods for classification and analysis of multivariate observations. In: Cam, L.M.L., Neyman, J. (eds.) Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability. vol. 1, pp. 281–297. University of California Press (1967)
17. Matallanas, E., Castillo-Cagigal, M., Gutirrez, A., Monasterio-Huelin, F., Caamao-Martn, E., Masa, D., Jimnez-Leube, J.: Neural network controller for active demand-side management with PV energy in the residential sector. Applied Energy **91**(1), 90–97 (2012)
18. Partridge, M., Calvo, R.A.: Fast dimensionality reduction and simple pca. Intelligent Data Analysis **2**(3), 203–214 (1998)
19. Rodrigues, P., Gama, J., Pedroso, J.: Hierarchical clustering of time-series data streams. Knowledge and Data Engineering, IEEE Transactions on **20**(5), 615–627 (2008)
20. Siuly, S., Li, Y.: Designing a robust feature extraction method based on optimum allocation and principal component analysis for epileptic EEG signal classification. Computer Methods and Programs in Biomedicine **119**(1), 29–42 (2015)
21. Sorzano, C.O.S., Vargas, J., Pascual-Montano, A.D.: A survey of dimensionality reduction techniques. CoRR abs/1403.2877 (2014)
22. Vesanto, J., Alhoniemi, E.: Clustering of the self-organizing map. Neural Networks, IEEE Transactions on **11**(3), 586–600 (2000)