Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Unsupervised learning method for clustering dynamic behavior in the context of power systems ⋆

Georgios Mitrentsis * **Hendrik Lens** *

* *Institute for Combustion and Power Plant Technology (IFK), University of Stuttgart, Germany (e-mail: {georgios.mitrentsis, hendrik.lens}@ifk.uni-stuttgart.de).*

Abstract: Aggregated dynamic equivalent models of active distribution networks (ADNs) are commonly derived using the measurement-based approach. This method exploits acquired data in order to estimate the model parameters using system identification techniques. However, most of the approaches assume that the system maintains the same dynamics for different operating conditions, even though the load mix and the distributed generation (DG) composition are constantly changing. To this end, this paper presents a novel method, which can be used as the first step of the system identification procedure, in order to take into account different system dynamics in ADN modeling. To do so, three unsupervised learning methods for clustering the various dynamic behaviors are introduced, yielding groups of measurements that represent different dynamics. In this context, the proposed methods leverage four clustering algorithms of different notion and complexity, namely k-means++, k-medoids, fuzzy c-means (FCM) and hierarchical clustering. To assess the validity of the proposed approach, real measurements acquired within a year in six real substations in Southern Germany are processed. The results highlight the remarkable difference in system dynamics justifying the necessity of an initial cluster analysis. Finally, the ratio of "Within Cluster sum of squares" to "Between Cluster Variation" (WCBCR) is deployed to compare the effectiveness of the clustering algorithms.

Keywords: active distribution networks, unsupervised learning, clustering, dynamic models, parameter identification, measurement-based approach, multisignal analysis

1. INTRODUCTION

In the last years, distribution networks have started changing drastically containing not only electrical loads but also distributed generation (DG), storage devices, electric vehicle charging infrastructure, and controlled loads. They have a direct impact on system dynamics and thus, system balance and stability need to be investigated further (Zali and Milanović (2013)). To this end, significant research effort was put into upgrading the existing dynamic distribution network models, i.e. load models, which are crucial for power system planning and control (Arif et al. (2017)).

Aggregated dynamic equivalent models are typically viewed as a promising way to model active distribution networks (ADNs) since they do not require a detailed representation of the system and they can accurately capture the various system dynamics. There are two major methods to extract aggregated models based on either system reduction or measurement data. Measurement-based methods have been considered as a more suitable solution for ADN modeling due to their general applicability (Resende et al. (2013)). As a general rule, those models have voltage as input, active or reactive power as output and their parameters are estimated based on data generated by a simulation software or by real measurement units.

However, the majority of those methods mainly derive only one single dynamic model assuming that the system has the same dynamic behavior for all operating conditions (Renmu et al. (2006), Bai et al. (2009), Zali and Milanović (2013), Papadopoulos et al. (2014), and Kontis et al. (2017)). In reality, the load mix and the DG composition are constantly changing due to time and weather variations. To this end, Kontis et al. (2019) and Metallinos et al. (2016) proposed two artificial intelligence methods to build generic dynamic models. However, both methodologies require in-depth knowledge about the grid configuration, e.g. load and generation mix, which in real-world applications is hard to obtain. Therefore, it is essential to develop a methodology which considers the various ADN dynamics in a realistic fashion without relying on a concrete detailed representation of the system.

To tackle the lack of detailed information of the components, which affect the ADN dynamics, while proving a solution that is valid for a wide range of operating conditions, a novel unsupervised learning technique is introduced. The proposed method encourages the development of several models for a single ADN, able to reliably capture the various dynamics. To do so, similar dynamic behaviors are automatically grouped together into clusters. Three

⋆ The research presented in this paper has been supported by TransnetBW GmbH (TSO) and Netze BW GmbH (DSO).

different strategies for clustering dynamic behavior are proposed and compared. The first strategy exploits the raw time domain (TD) data of input and output signals in order to find common patterns within them and to group them accordingly. The second strategy is an enhanced version of the first one. Here, both TD and frequency domain (FD) data of the measured signals are deployed. In the third strategy, a first- or a second-order transfer function is generated for each single measurement and then a step input is applied to each model. In the next step, a cluster analysis is performed based only on the calculated step responses. Furthermore, for each one of the aforementioned strategies, four different clustering algorithms are employed, namely k-means++, k-medoids, fuzzy c-means (FCM) and hierarchical clustering.

Once each single measurement has been assigned to a specific cluster, a different dynamic model can be derived for every cluster based on a multisignal analysis technique of its entries. The derived models can describe either the active or the reactive power response of the system. Importantly, the proposed method can be effortlessly integrated into most of the existing ADN modeling or load modeling approaches in literature, as a pre-processing step in order to take into account different dynamics and thus, enhance their performance.

To assess the validity of the proposed method, measurement units were installed at the lower voltage side of transformers in six substations in Southern Germany. Each transformer connects an ADN of 20 kV to a high voltage grid of 110 kV. Each unit was triggered every time there was a voltage change bigger than a user-defined threshold and high resolution data of voltage, active, and reactive power were recorded. In order to capture a wide range of grid configurations as well as time and weather conditions, the units acquired measurement data over a year.

The clustering results indicate that indeed an array of different dynamics were detected within the examined ADN due to diverse load and generation composition, justifying the imperative need of an initial cluster analysis.

2. PROPOSED METHODOLOGY

The proposed methodology can be described by the flowchart in Fig. 1. The method exploits the data recorded at the point of common coupling (PCC) during a voltage disturbance. Each measurement i yields three vectors containing the rms values of voltage ($\boldsymbol{V}_i$), active ($\boldsymbol{P}_i$) and reactive power ($\boldsymbol{Q}_i$). Moreover, it is crucial that the total number of measurements M are obtained over a sufficient period of time in order to capture the diverse dynamics due to, e.g. time and seasonal variations.

2.1 Data pre-processing

Since the measurements are obtained at PCC and there is DG within the examined ADN, it is clear that the recorded active power does not reflect the total power consumption. Furthermore, it is intended that the proposed method works independently of the initial voltage, active or reactive power conditions. To this end, a pre-processing step is applied to the acquired data. In this phase, the actual differences between the raw TD values ($V_i(t)$, $P_i(t)$, $Q_i(t)$) and the initial pre-disturbance conditions ($V_{0,i}$,$P_{0,i}$, $Q_{0,i}$) of voltage, active and reactive power, respectively, are calculated for each time step t using:

$$\Delta X_i(t) = X_i(t) - X_{0,i}, \quad X_{0,i} = \frac{1}{N_0} \cdot \sum_{t_n \in [t_0-1, t_0)} X_i(t_n) \quad (1)$$

where t_0 is the time of the disturbance and N_0 is the number of samples recorded within 1 s before the disturbance. Furthermore, depending on the context, the variable X represents voltage, active and reactive power, $X_i(t)$ and $X_{0,i}$ indicate the TD and pre-disturbance values of the i-th measurement, respectively, while $\Delta X_i(t)$ is the actual difference between them. It is worth mentioning that the initial values are estimated by averaging the recorded values right before the disturbance due to the measurement noise and the continuous switching of the consumers.

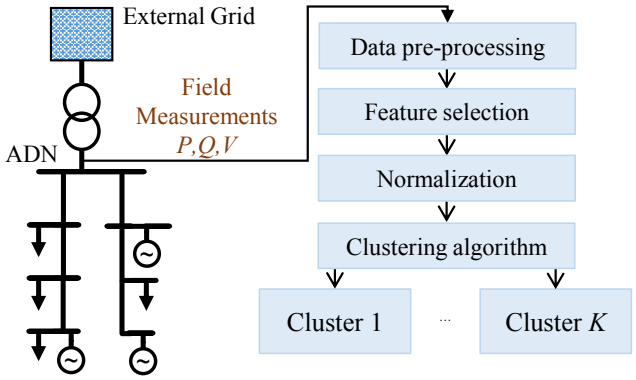


Fig. 1. Flowchart of the proposed methodology.

Once the actual differences of voltage, active and reactive power have been estimated, three different strategies are deployed based on different feature selection in order to find common patterns within the pre-processed data. Therefore, the major difference between the strategies lies in the features that are used by the clustering algorithm.

2.2 Time domain data

The first strategy leverages the pre-processed TD values of voltage and active or reactive power. The choice of which power will be used is determined by the user based on the desired ADN modeling approach. To form the feature matrix, the actual change in voltage and active or reactive power are merged into a vector $\boldsymbol{F}_i = [\Delta V_i(t), \Delta P_i(t)]^\top$ or $\boldsymbol{F}_i = [\Delta V_i(t), \Delta Q_i(t)]^\top$ for every measurement $i = 1, ..., M$.

2.3 Time and frequency domain data

The second strategy is similar to the first one but instead of using only the TD data of the pre-processed signals, it additionally exploits their FD values as derived by the Fast Fourier Transform (FFT) with a sampling frequency of 100 Hz and a signal duration of 2.5 s. Importantly, for each measurement i, the frequency in which the FFT of the examined signal has the maximum amplitude, is appended to every feature vector $\boldsymbol{F}_i$ as:

$$\boldsymbol{F}_i = [\ \Delta V_i(t), \Delta P_i(t), \\ f_{\Delta V,max}, ..., f_{\Delta V,max}, f_{\Delta P,max}, ..., f_{\Delta P,max}]^\top \quad (2)$$

for active power, or in case of reactive power:

$$\boldsymbol{F}_i = [\ \Delta V_i(t), \Delta Q_i(t), \\ f_{\Delta V,max}, ..., f_{\Delta V,max}, f_{\Delta Q,max}, ..., f_{\Delta Q,max}]^\top \quad (3)$$

where $f_{\Delta V,max}$, $f_{\Delta P,max}$ and $f_{\Delta Q,max}$ indicate the frequencies, where the FFT of $\Delta V_i(t)$, $\Delta P_i(t)$ and $\Delta Q_i(t)$ get their maximum amplitude value, respectively. In order to balance both TD and FD influence, the corresponding frequency values are expanded within the feature vector, so that they occupy the same amount of vector positions as their TD signal. An alternative would be to deploy the single-sided amplitude spectrum of the FFT, but the highly noisy data of our application hinder the implementation of this approach in practice.

2.4 Step responses

The third strategy approaches the problem of finding patterns within dynamic behavior in a different way, which can be described by the following steps:

(1) A first- and a second-order transfer function, $G_{1,i}(s)$ and $G_{2,i}(s)$, respectively, are estimated for each single measurement i. The input signal is the change in voltage, $\Delta V_i(t)$, and the output signal is the change in power, $\Delta P_i(t)$ or $\Delta Q_i(t)$. The order of the transfer function was selected to be low in order to avoid overfitting while capturing the general dynamic response. Finally, the transfer function parameters are estimated using the CONtinuous-Time System IDentification (CONTSID) technique (Garnier et al. (2003)).
(2) The two derived models, i.e. $G_{1,i}(s)$ and $G_{2,i}(s)$, are compared for each measurement i. The one that yields the best fit is selected. The best fit is determined in terms of normalized mean square error between the simulated response and the real output value.
(3) A unit step function is applied as an input to each one of the selected transfer function models, yielding the step responses $\Delta Y_i(t)$. Importantly, $\Delta Y_i(t)$ refers either to an active power or reactive power step response based on the desired ADN model.
(4) In this step, the TD values of $\Delta Y_i(t)$ are deployed to form each feature vector $\boldsymbol{F}_i = [\Delta Y_i(t)]^\top$.

It is worth mentioning that the main motivation behind this method is to decrease the noise influence during clustering and reduce the total number of required features.

2.5 Merge and normalization

This step is common for all three strategies. In this phase, each individual feature vector $\boldsymbol{F}_i$ is put into a set $\boldsymbol{F} = [\boldsymbol{F}_1, \boldsymbol{F}_2, ..., \boldsymbol{F}_M]$, which forms the feature matrix. Then, each row of $\boldsymbol{F}$ is normalized and the new matrix is directed as input to the clustering algorithm. Since clustering raw dynamic responses has, to the best of the authors' knowledge, not been presented in literature before, the choice of an adequate clustering algorithm in not trivial. Therefore, four clustering algorithms of different notion and complexity are tested.

3. CLUSTERING ALGORITHMS

3.1 K-means++

K-means++ is one of the most commonly deployed clustering algorithms, since it has shown remarkable results in a wide range of engineering applications (Pham and Afify (2007)). It basically partitions the data into K clusters $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_K\}$. To do so, k-means++ tries to minimize the "Within Cluster Sum of Squares" (WCSS) in an iterative manner (Arthur and Vassilvitskii (2007)). This minimization can be written as:

$$\arg\min_{\mathcal{C}} \sum_{k=1}^{K} \sum_{\boldsymbol{F}_i \in \mathcal{C}_k} \|\boldsymbol{F}_i - \boldsymbol{c}_k\|^2, \quad (4)$$

where $\boldsymbol{c}_k = \frac{1}{N_k} \sum_{\boldsymbol{F}_i \in \mathcal{C}_k} \boldsymbol{F}_i$ indicates the centroid of $\mathcal{C}_k$, as estimated by averaging its N_k entries. In our application, the maximum number of iterations was set to 1000, while 100 tries were executed using new initial cluster centroid positions.

3.2 K-medoids

Contrary to k-means++ which uses the mean of cluster entries as center (centroid), k-medoids employs the median cluster member as center (medoid). This k-medoids feature offers robustness to noise and outliers (Pham and Afify (2007)). As a general rule, the partitioning around medoids (PAM) algorithm is commonly used for small data sets in order to solve the k-medoids problem. Its main principle is similar to k-means++ algorithm but instead of estimating a new cluster centroid in every iteration, PAM algorithm checks each one of the cluster members as a potential medoid based on their generated WCSS (Kaufman and Rousseeuw (2009)). In our application, similar parameters were deployed as in k-means++.

3.3 Fuzzy c-means (FCM)

FCM is based on the general idea that each data entry can partially belong to more than one cluster using a degree of membership (Bezdek (2013)). In an iterative manner, FCM partitions the data into clusters and assigns the adequate degrees of membership by minimizing the objective function:

$$J_m = \sum_{i=1}^{M} \sum_{k=1}^{K} \mu_{ik}^m \|\boldsymbol{F}_i - \boldsymbol{c}_k\|^2 \quad (5)$$

where μ_{ik} indicates the degree of membership of $\boldsymbol{F}_i$ in the k-th cluster, whereas m denotes the fuzzy partition matrix exponent for controlling the degree of fuzzy overlap. In our application, m was set to 2 and since each cluster's members will be used for parameter estimation, each entry is assigned to the cluster with the maximum membership.

3.4 Hierarchical clustering

The agglomerative hierarchical clustering is the only algorithm among those proposed which employs a bottom-up approach. It starts with the hypothesis that each data entry belongs to its own cluster. Then, it estimates the similarity or dissimilarity, e.g. euclidean distance, for each

data entry with the rest of the data set, forming an $N \times N$ symmetrical matrix. Once the similarity matrix has been formed, the pair of entries with the highest similarity are merged together into a new cluster. A new $(N-1)\times(N-1)$ similarity matrix is created and again the most similar pairs are merged together until a hierarchical tree is developed (Ward Jr (1963)). The process is repeated until the specified number of clusters is reached. In our application, the euclidean distance was deployed as a similarity metric.

3.5 Optimal number of clusters

In order to determine the optimal number of clusters, the knee-point criterion of WCBCR is deployed (Tsekouras et al. (2007)), at which a further increase in the total number of clusters does not produce a significant decrease in WCBCR. In the numerator, WCBCR expresses the sum of the distances of all data entries to their center while the sum of distances between cluster centers forms the denominator:

$$\mathrm{WCBCR} = \frac{\sum_{k=1}^{K}\sum_{i=1}^{M_k} d^2(\boldsymbol{c}_k, \boldsymbol{F}_i)}{\sum_{1\le k_1\le k_2}^{K} d^2(\boldsymbol{c}_{k_1}, \boldsymbol{c}_{k_2})} \tag{6}$$

where M_k denotes the total number of entries assigned to cluster $\mathcal{C}_k$ and $d(\bullet,\bullet)$ denotes the euclidean distance.

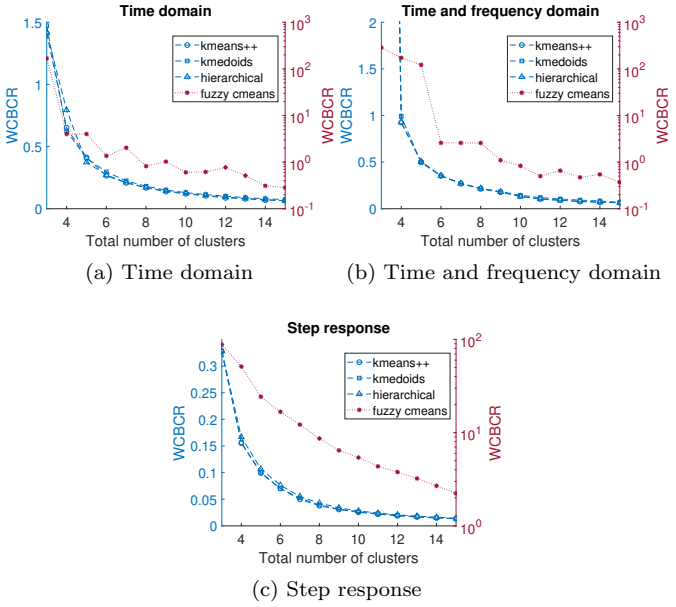
4. CASE STUDY USING REAL MEASUREMENTS

The aim of this section is to highlight the existence of different dynamics within an ADN, which cannot be neglected during the modeling procedure. In this context, high resolution data were recorded from August 2017 to July 2018 in six real substations in Southern Germany. The measurement units were triggered every time there was a voltage drop greater than 1% of the previous measured value. Within that period of time, around 1200 measurements were acquired at each substation. Due to limited space, the clustering results of one representative substation are presented. Nevertheless, the rest of substations showed behavior similar to the one illustrated below.

4.1 Clustering algorithms' comparison

As a first step of the proposed unsupervised learning technique, the different clustering algorithms are compared in terms of WCBCR. To do so, WCBCR is estimated for different numbers of clusters, ranging from 3 to 15, and the results are depicted in Fig. 2, for each one of the proposed strategies. Voltage and active power data were deployed.

For all three strategies, k-means++, k-medoids and hierarchical clustering (blue curves) enjoy a relatively small WCBCR without any significant difference between them. In contrast, FCM algorithm (red curves) yields extremely big WCBCR for all three strategies. This result can be justified by the fact that each data entry was assigned to the cluster with the maximum membership value, even though it belongs to more than one cluster with different degree of membership. In addition, similar clustering performance was observed in case of voltage-reactive power combination, with k-means++, k-medoids and hierarchical clustering outperforming FCM; while there was no



(a) Time domain (b) Time and frequency domain

(c) Step response

Fig. 2. WCBCR for different strategies and clustering algorithms.

notable deviation between their WCBCR. It should be also mentioned that WCBCR is not suitable for the comparison between the three proposed strategies, since different features are used. Therefore, qualitative results are presented in the next paragraph.

4.2 Clustering results

To compare the three different strategies, k-means++ was randomly picked as a clustering algorithm and the

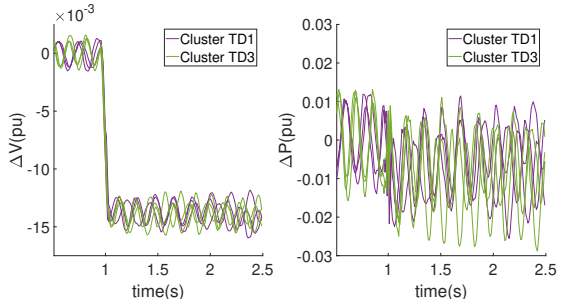


Fig. 3. Clusters with same dynamic behavior, as generated by the "Time domain" method (TD).

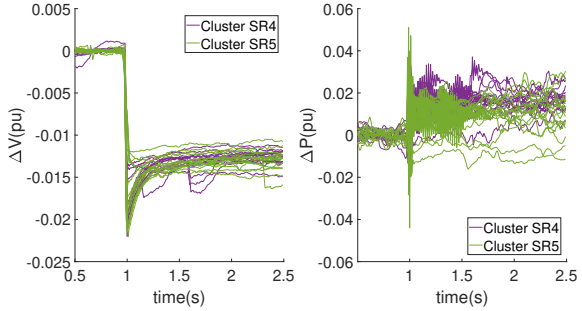


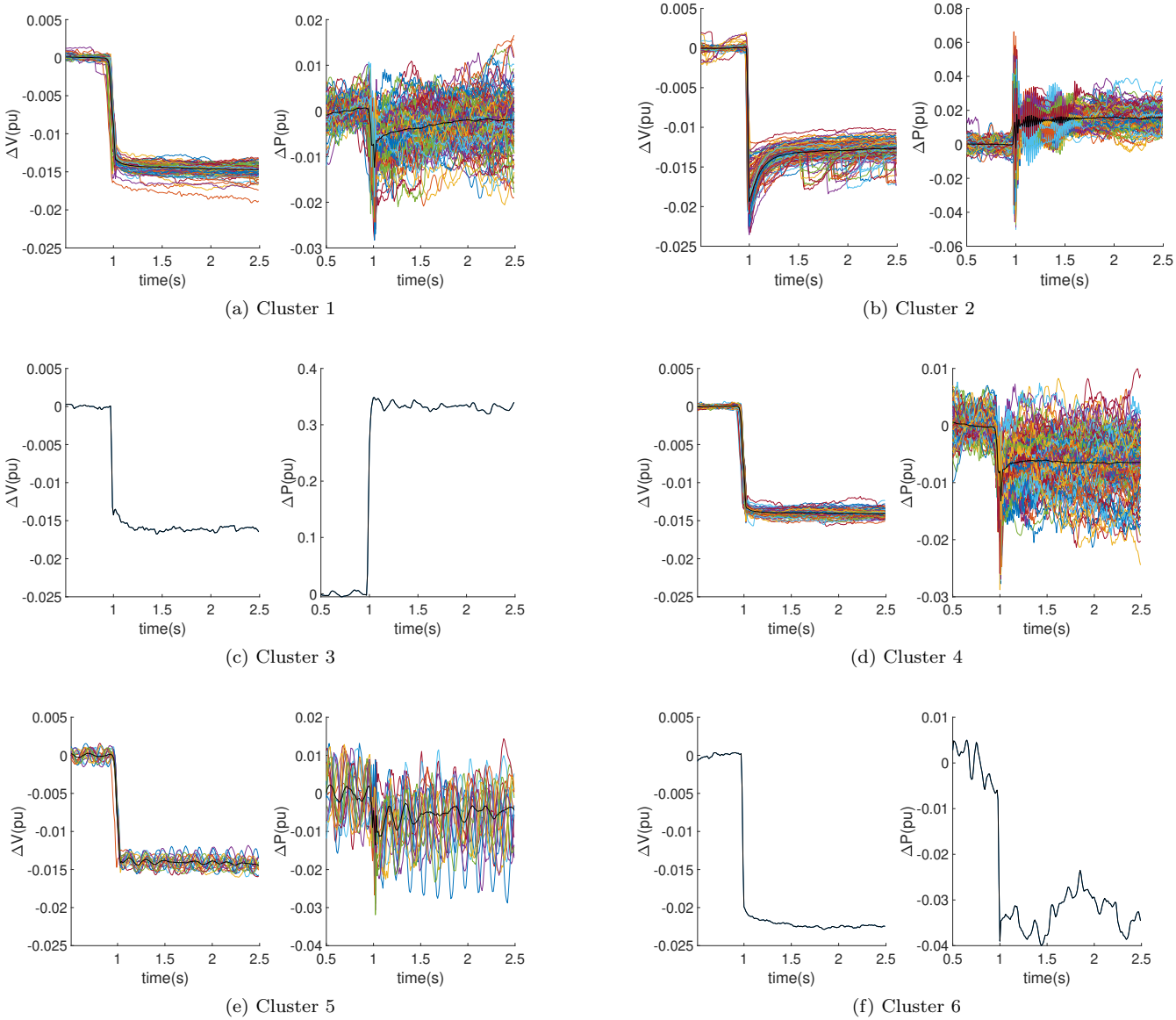
Fig. 4. Clusters with same dynamic behavior, as generated by the "Step responses" method (SR).

knee-point criterion of WCBCR was employed in order to determine the optimal number of clusters.

Both "Time domain" and "Step responses" methods can distinguish different dynamic behaviors and group them into clusters. However, in case of oscillatory input or output, more than one cluster representing the same dynamics were developed as shown for example in Fig. 3 and 4. The magenta curves indicate the entries that were assigned to cluster TD1 (Fig. 3) and cluster SR4 (Fig. 4), while the green curves denote the entries of cluster TD3 (Fig. 3) and SR5 (Fig. 4), using the "Time domain" and "Step responses" strategy, respectively. Since both approaches are exclusively based on TD features, redundant clusters are developed for oscillatory responses due to small phase shifts of the corresponding signals. Therefore, in the event of oscillations in our system, it is likely that the two aforementioned approaches will yield redundant clusters.

This issue is tackled using the "Time and frequency domain" method, as shown in Fig. 5. The whole dataset of voltage and active power measurements was grouped into six distinguishable clusters. The left graphs present the voltage change and the right ones the active power response. The colored curves indicate the individual recordings, whereas the black curves denote the average of each cluster's measurements. It is worth mentioning that although strong noise is present in the active power response due to the continuous switching of the consumers, the proposed method can accurately capture the signals' dynamics, as is clearly highlighted by the average cluster responses.

Importantly, cluster 1 and 4 illustrate the partial active power recovery after a voltage change, due to the induction machines' presence within the system (Karlsson and Hill (1994)). Although both clusters look similar, they differ mainly in the power's new steady state, as shown by the average cluster responses, with cluster 1 entries being characterized by a bigger active power recovery. Furthermore,



(a) Cluster 1

(b) Cluster 2

(c) Cluster 3

(d) Cluster 4

(e) Cluster 5

(f) Cluster 6

Fig. 5. Clustering results in terms of ΔV and ΔP using "Time and frequency domain" method (TF).

the entries of cluster 2 expose a dynamic behavior, where active power rises even though voltage drops. Besides that, high frequency oscillations occur for a few time steps after the voltage change, which possibly originate from complex control strategies of DG. Cluster 3 and 6 contain only one entry with a significantly bigger active power change than the rest of the measurements. In particular, the active power change of cluster 3 reaches an actual difference of 0.35 p.u., before and after the disturbance. Finally, all oscillations with similar amplitudes and frequencies were correctly assigned to cluster 5, independent of the phase shift between them.

It is now clear that the "Time and frequency domain" method can accurately distinguish and group different dynamic behaviors within an ADN. Furthermore, the results confirmed the existence of considerably different dynamics that should not be assumed to be similar. To this end, the need for multiple models describing the same ADN is imperative. It should be mentioned that similar results were acquired for voltage and reactive power data using the "Time and frequency domain" method. However, the reactive power signals do not contain significant noise coming from the switching of the consumers and thus, the clustering results are much more clear.

Most importantly, the proposed methodology can be deployed to identify outliers, e.g. internal faults within the examined ADN, which should be discarded from the parameter estimation procedure.

5. CONCLUSION

This paper presents a novel unsupervised learning methodology for clustering different dynamic behavior in the context of ADN modeling. Importantly, no knowledge of load and generation mix is required and the clustering is performed exclusively based on the measurements themselves. Furthermore, the proposed method can be easily integrated as a pre-processing step into most of the existing ADN modeling approaches. Therefore, different dynamic models can be accurately generated representing the same ADN under various operating conditions, since each measurement is mapped to a different cluster.

Importantly, the combination of TD and FD features has been proved to work most effectively, since it avoids generating redundant clusters with the same dynamics. Furthermore, k-means++, k-medoids and hierarchical clustering produced almost identical response in terms of WCBCR and they can be successfully deployed on clustering dynamic behavior. On the contrary, FCM showed poor clustering performance and it is not suggested in the proposed methodology.

In future work, artificial intelligence algorithms will be implemented to link the derived clusters with the various weather and seasonal conditions. In this context, the system operator will be able to analyse the influence of those parameters to the grid dynamics and prevent undesired scenarios.

REFERENCES

Arif, A., Wang, Z., Wang, J., Mather, B., Bashualdo, H., and Zhao, D. (2017). Load modeling—a review. *IEEE Transactions on Smart Grid*, 9(6), 5986–5999.

Arthur, D. and Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 1027–1035.

Bai, H., Zhang, P., and Ajjarapu, V. (2009). A novel parameter identification approach via hybrid learning for aggregate load modeling. *IEEE Transactions on power systems*, 24(3), 1145–1154.

Bezdek, J.C. (2013). *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media.

Garnier, H., Mensler, M., and Richard, A. (2003). Continuous-time model identification from sampled data: implementation issues and performance evaluation. *International Journal of Control*, 76(13), 1337–1357.

Karlsson, D. and Hill, D.J. (1994). Modelling and identification of nonlinear dynamic loads in power systems. *IEEE Transactions on Power Systems*, 9(1), 157–166.

Kaufman, L. and Rousseeuw, P.J. (2009). *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons.

Kontis, E.O., Papadopoulos, T.A., Chrysochos, A.I., and Papagiannis, G.K. (2017). Measurement-based dynamic load modeling using the vector fitting technique. *IEEE Transactions on Power Systems*, 33(1), 338–351.

Kontis, E.O., Papadopoulos, T.A., Syed, M.H., Guillo-Sansano, E., Burt, G.M., and Papagiannis, G.K. (2019). Artificial-intelligence method for the derivation of generic aggregated dynamic equivalent models. *IEEE Transactions on Power Systems*, 34(4), 2947–2956.

Metallinos, K.S., Papadopoulos, T.A., and Charalambous, C.A. (2016). Derivation and evaluation of generic measurement-based dynamic load models. *Electric Power Systems Research*, 140, 193–200.

Papadopoulos, P.N., Papadopoulos, T.A., Crolla, P., Roscoe, A.J., Papagiannis, G.K., and Burt, G.M. (2014). Black-box dynamic equivalent model for microgrids using measurement data. *IET Generation, Transmission & Distribution*, 8(5), 851–861.

Pham, D. and Afify, A. (2007). Clustering techniques and their applications in engineering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 221(11), 1445–1459.

Renmu, H., Jin, M., and Hill, D.J. (2006). Composite load modeling via measurement approach. *IEEE Transactions on power systems*, 21(2), 663–672.

Resende, F., Matevosyan, J., and Milanovic, J. (2013). Application of dynamic equivalence techniques to derive aggregated models of active distribution network cells and microgrids. In *2013 IEEE Grenoble Conference*.

Tsekouras, G.J., Hatziargyriou, N.D., and Dialynas, E.N. (2007). Two-stage pattern recognition of load curves for classification of electricity customers. *IEEE Transactions on Power Systems*, 22(3), 1120–1128.

Ward Jr, J.H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301), 236–244.

Zali, S.M. and Milanović, J.V. (2013). Generic model of active distribution network for large power system stability studies. *IEEE Transactions on Power Systems*, 28(3), 3126–3133.