Citation Information: 10.1016/j.segan.2021.100509

Assessment of Load and Generation Modelling on the Quasi-static analysis of Distribution Networks

I. S. Lamprianidou[1], T. A. Papadopoulos[1,*], G. C. Kryonidis[2], E. Fatih Yetkin[3], K. D. Pippi[1], A. I. Chrysochos[4]

[1]Department of Electrical and Computer Engineering, Democritus University of Thrace, Xanthi 67100, Greece

[2]School of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece

[3]Management Information Systems Department, Kadir Has University, Istanbul, Turkey

[4]R&D Department, Hellenic Cables, Athens 15125, Greece

[*]Corresponding author

E-mail addresses: ifiglampr@gmail.com (I. S. Lamprianidou), thpapad@ee.duth.gr (T. A. Papadopoulos), kryonidi@ece.auth.gr (G. C. Kryonidis), fatih.yetkin@khas.edu.tr (E. Fatih Yetkin), kpippi@ee.duth.gr (K. D. Pippi), achrysochos@hellenic-cables.com (A. I. Chrysochos)

Abstract: Quasi-static analysis of power systems can be performed by means of timeseries-based and probability density function-based models. In this paper, the effect of different load and generation modelling approaches on the quasi-static analysis of distribution networks is investigated. Different simplified load and distributed renewable energy sources generation timeseries-based models are considered as well as probabilistic analysis. Moreover, a more sophisticated approach based on cluster analysis is introduced to identify harmonized sets of representative load and generation patterns. To determine the optimum number of clusters, a three-step methodology is proposed. The examined cases include the quasi-static analysis of distribution networks for different operational conditions to identify the simplified modelling approaches that can efficiently predict the network voltages and losses. Finally, the computational efficiency by using the simplified models is evaluated in temperature-dependent power flow analysis of distribution networks.

Keywords: Distributed generation modelling, load modelling, load timeseries, photovoltaic systems, wind turbines.

1 Introduction

Traditionally, steady-state simulations are performed by transmission and distribution system operators to investigate cases where an electric power system may experience equipment overloads, sustained overvoltage or undervoltage conditions, excessive voltage fluctuations, or equipment control problems [1]. Steady-state studies in their most basic form refer to power flow analysis for the calculation of voltages, currents, active and reactive power flows, as well as losses. In conventional power flow studies, a snapshot of the operating condition of the power system is solved, focusing mainly on specific loading conditions that would reveal possible voltage or thermal violations, e.g., low demand – high generation.

Nowadays, distribution grids are gradually transformed to an active component of the power systems as a result of the increasing penetration of distributed renewable energy sources (DRESs). This inevitable shift towards more active distribution networks (ADNs) poses unprecedented technical challenges to transmission (TSOs) and distribution system operators (DSOs), jeopardizing the reliable operation of power systems. Focusing on distribution networks, a series of technical problems emerge, e.g., overvoltages, overloading of network equipment, etc. To thoroughly evaluate the severity and the potential impact of these problems on the network operation, as well as to effectively address them, system studies based on the power flow analysis must be performed [1]. Nevertheless, due to the operational flexibility of ADNs, the stochastic variation of loads, and the intermittent nature of DRESs [2], the applicability and effectiveness of the conventional static power flow approach is limited, since only major operational conditions at specific time instants can be analysed [1], [3].

To overcome this limitation and investigate the operational characteristics of ADNs over a wide range of operational conditions, e.g., within a set time of period, the quasi-static approach has been proposed [1], [4]. According to the IEEE Std 1547.7-2013 "*Quasi-static simulation refers to a sequence of steady-state power flow conducted at a time step of no less than 1 second but that can use a time step of up to one hour*", assuming "*no numerical integration of differential equations between time steps*". There are two main approaches to perform quasi-static analysis, namely the timeseries- and the probability density function (pdf)-based [5], [6]. The former consists of solving successive deterministic power flow calculations, using as inputs consumption and generation timeseries over a time span, e.g., a month, a year, etc. [4], [7].

In the latter, the inputs (consumption and generation) are specified by means of their pdf to imitate the system stochasticity. The most significant drawback of both approaches is the increased computational burden, since a large number of power flow calculations must be performed [8]. In this sense, quasi-static analysis may become impractical, especially in applications requiring iterative calculations, e.g., temperature-dependent power flow [9], [10], power flow for islanded microgrids [11], [12], application of evolutionary optimization methods [13]-[15], analysis of large-scale networks to study TSO-DSO interactions as well as impact studies requiring long time spans or high temporal resolution [7], [15]. Therefore, simplification methods have been proposed on the basis of reducing algorithmically the complexity of power flow calculations or by applying topological simplifications [8], [17]. An alternative solution to the above is the development of simplified load and generation timeseries-based models [7], [18], [19].

The paper aims to provide a thorough assessment of the performance of different simplified electricity load and DRES generation models for the quasi-static analysis of distribution networks. In this sense, timeseries-based models [7], [18], [19] as well as pdf analysis are applied. The strength and contributions of the analysis are summarized in the following aspects:

- A new sophisticated approach based on cluster analysis is proposed to identify sets of representative load profiles harmonized with the corresponding DRES generation patterns. Two techniques forming the input of the clustering procedure are applied, i.e., time-domain data and its projection onto the principal components
- A three-step methodology to determine the optimum number of clusters is proposed. In particular, the number of clusters is initially estimated in terms of certain metrics; subsequently, the set of clusters is redefined by applying classification analysis and is finally determined by getting feedback from the quasi-static analysis of distribution networks. To the best of our knowledge, a systematic methodology to derive representative patterns by using cluster analysis has been well investigated regarding mainly load profiles [20] - [22], neglecting harmonization with generation as in the current analysis. Moreover, such models focus mainly on buildings energy analysis and benchmarking, while their application on distribution network studies has not been investigated.
- A framework for quasi-static studies considering different operational conditions of distribution networks is proposed to identify the simplified modelling approaches that can efficiently predict the network voltages and losses. The analysis builds on a previous

study [18] that has described preliminary investigations and comparative analysis regarding the effect of some of the examined models.

- The computational efficiency of the examined simplified models is demonstrated and evaluated in an excessively computationally intensive problem, i.e., a temperature-dependent power flow analysis.

2 Load and DRES profiles

Aggregated daily electricity data of a university campus, namely Democritus University of Thrace (DUTH), located in Greece [23] are used as load profiles. DUTH campus consists of three buildings of different use, including offices, teaching rooms, lecture halls and laboratories. It is connected to the primary distribution system (20 kV) via a 20/0.4 kV transformer. The load data have been recorded at the primary side of the substation transformer from January 2014 to December 2014 with recording sampling of 15-min [24]. The original active and reactive power data are normalized assuming 430 kVA as base value.

Additionally, for the site under study, the hourly solar irradiance [25] and wind speed [26] timeseries of the Modern-Era Retrospective analysis for Research and Applications (MERRA) dataset over a one-year period (2014) are used. The corresponding DRES output is calculated by using the online simulation tool of [27]. Considering photovoltaic (PV) generation, a 1 kW_p system is used, assuming 35^o tilt, 180^o azimuth and 10% total system losses, while for wind generation a 1 kW Vestas V90 2000 wind turbine (WT) with 80-m hub height is used [27]. To comply with load data, 15-min DRES timeseries are created by applying the piecewise cubic interpolation to the hourly data.

3 Quasi-static Models

The studies within this work are performed over one-year period, by applying timeseries- and pdf-based models considering the available load and DRES data.

3.1 Timeseries-based Models

In timeseries analysis a series of data points in chronological order is used [4], [7]. On the basis of one-year analysis, the electricity load demand and DRES generation datasets consist of 35040 samples, i.e., 4 samples/hour×24hours/day×365days. This “full timeseries” dataset may be computationally demanding and expensive; thus, simplified timeseries-based models of reduced computational burden can be derived as briefly discussed below [18], [19].

3.1.1 Monthly Representative Timeseries

The concept of the monthly representative timeseries (MRT) model is based on the formulation of electricity demand and DRES generation representative profiles for each month.

In particular, regarding electricity demand (load), working (WD) and non-working (NWD) day classification is applied. The representative profiles are calculated by averaging the data of the same time segment *t* of the day (quarter of hour) within each month. In total 23 representative profiles are created regarding active and reactive power, respectively, i.e., 11 profiles for the WDs of each month (except August, since all days are assumed as NWDs in DUTH) and 12 for the NWD profiles. Accordingly, 12 representative profiles are created for solar and wind generation. In Fig. 1, the representative load profiles of active and reactive power for the WDs and NWDs of March are indicatively presented, as well as the corresponding DRES generation models.

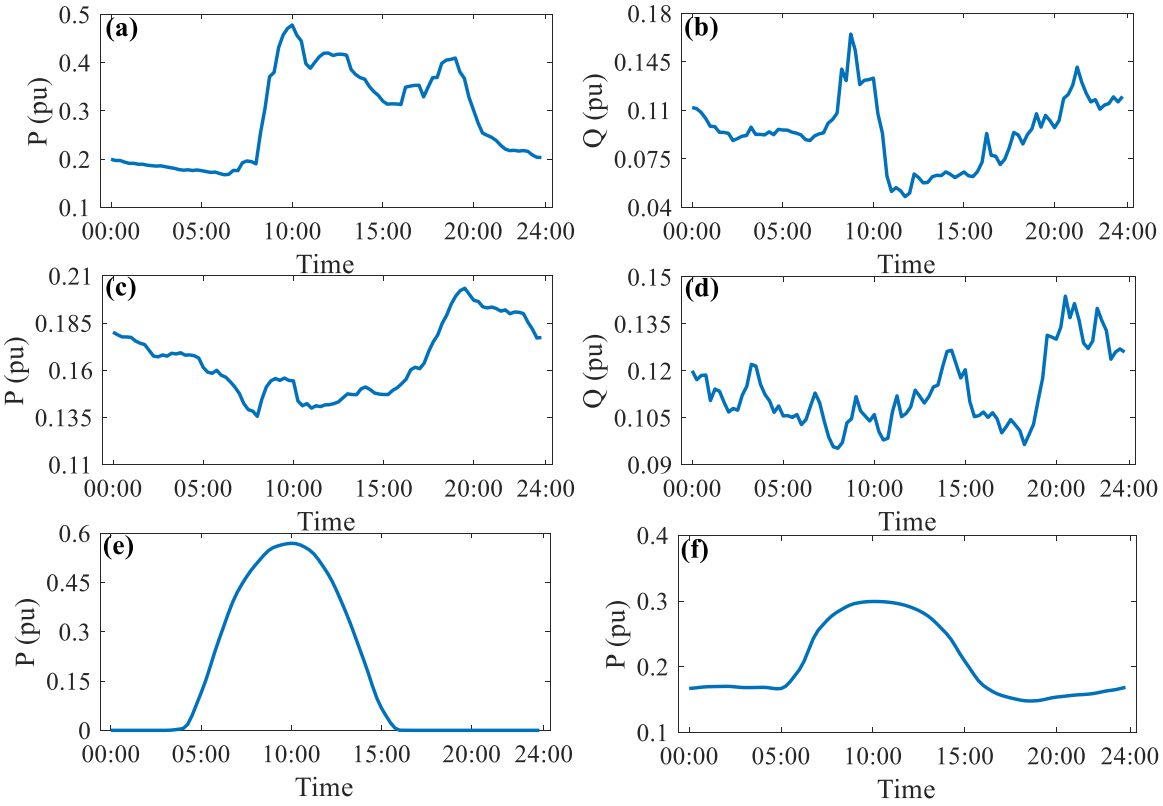

Fig. 1. Monthly representative profiles of March for the a) load active power and b) load reactive power of WDs, c) load active power and d) load reactive power of NWDs, e) PV generation, and f) WT generation.

3.1.2 Annual Representative Timeseries

The second simplified timeseries-based model is the annual representative timeseries (ART). Load (WD and NWD) and DRES representative profiles are calculated by averaging the corresponding data of the same time segment within the year. In Fig. 2, the load and DRES generation ART models are depicted.

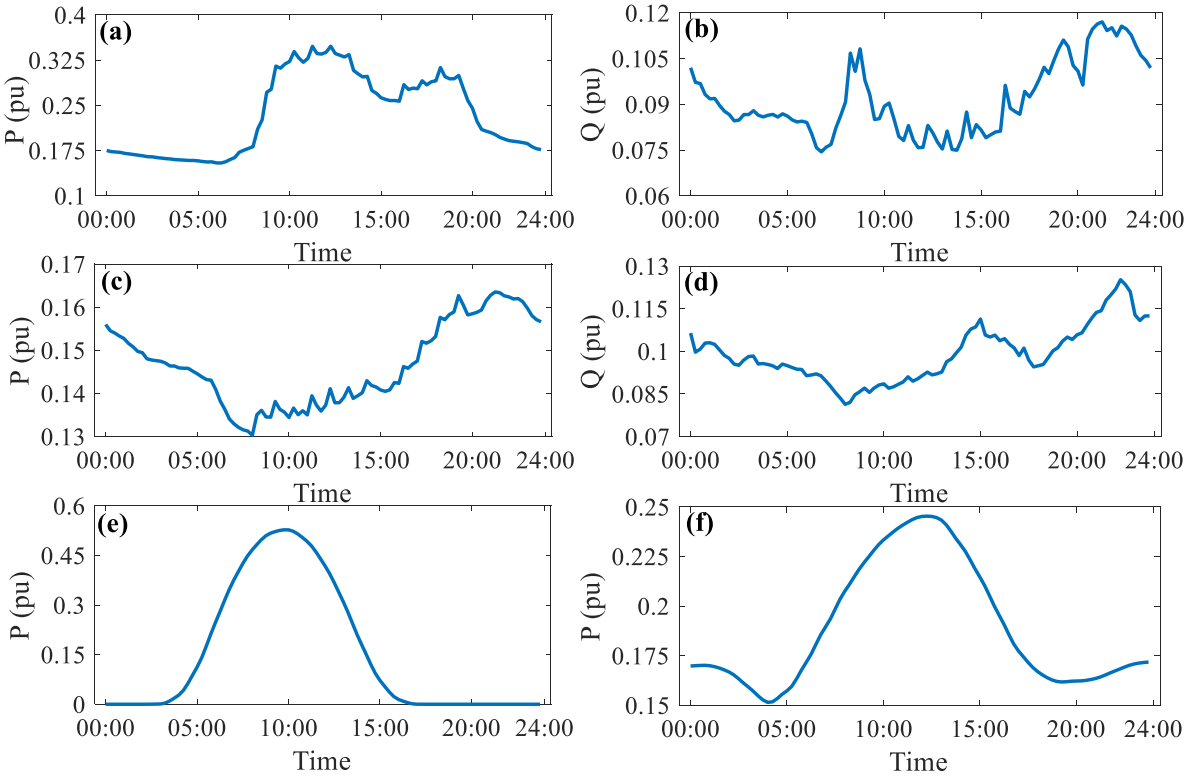

Fig. 2. Annual representative profiles for the a) load active power and b) load reactive power of WDs, c) load active power and d) load reactive power of NWDs, e) PV generation, and f) WT generation.

3.2 Pdf-based Models

In this approach, load and DRES data are generated for each time segment by using pdfs. Note that in this sense, the dataset is created in a memoryless way without taking into account their chronological variation. For active and reactive power, the normal pdf is used [28] for both WD and NWD load datasets. Random data are generated for each time segment of the day, assuming as mean (μ) the corresponding value obtained from the ART model and σ the standard deviation of the real and reactive power ART profiles. The 365 generated daily active and reactive power profiles are summarized in Fig. 3 (blue lines); red lines correspond to the ART daily profiles.

Regarding solar generation, the daily profiles are generated assuming the Beta pdf for the normalized solar insolation, described by (1) [29]:

$$f_{C_t}(C_t;\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}C_t^{\alpha-1}\left(1-C_t\right)^{\beta-1}, \tag{1}$$

where $\alpha>0$, $\beta>0$ are the shape parameters, C_t is the normalized hourly solar insolation at t and $\Gamma()$ is the Gamma function [29]. In Fig. 4, the generated PV profiles using the parameters of Table I are presented; the red line corresponds to the highest annual power generation.

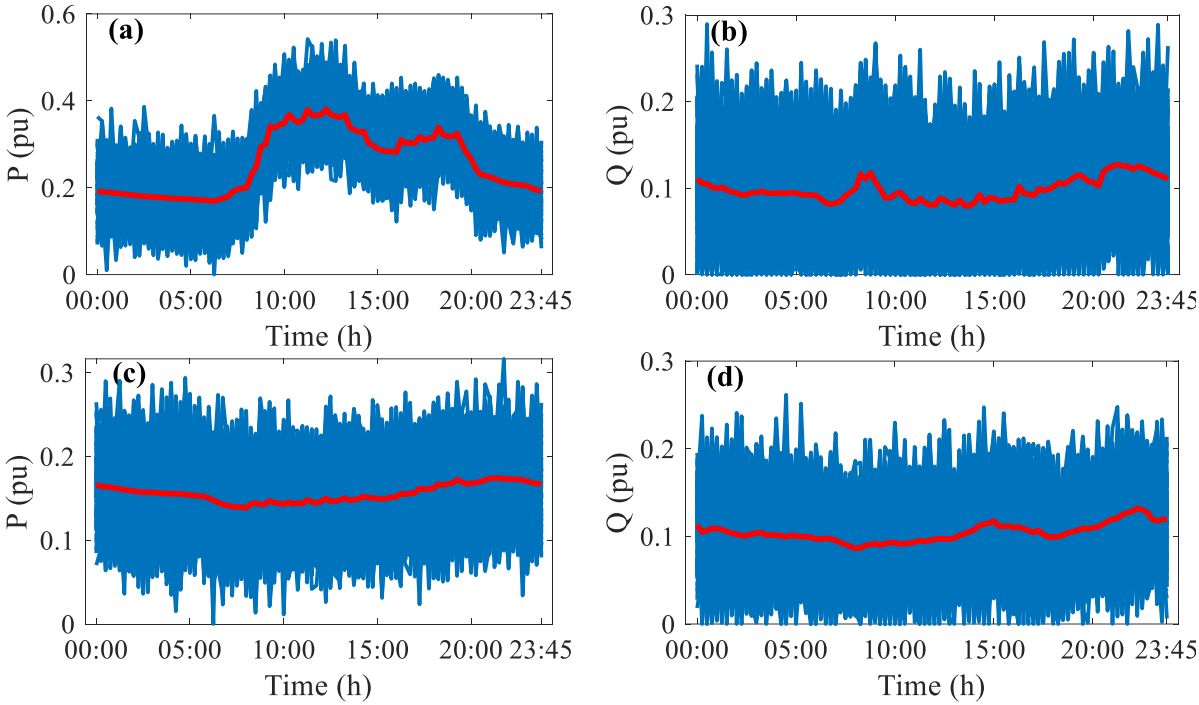

Fig. 3. Daily pdf-based load profiles of WD a) active and b) reactive power, and NWD c) active and d) reactive power.

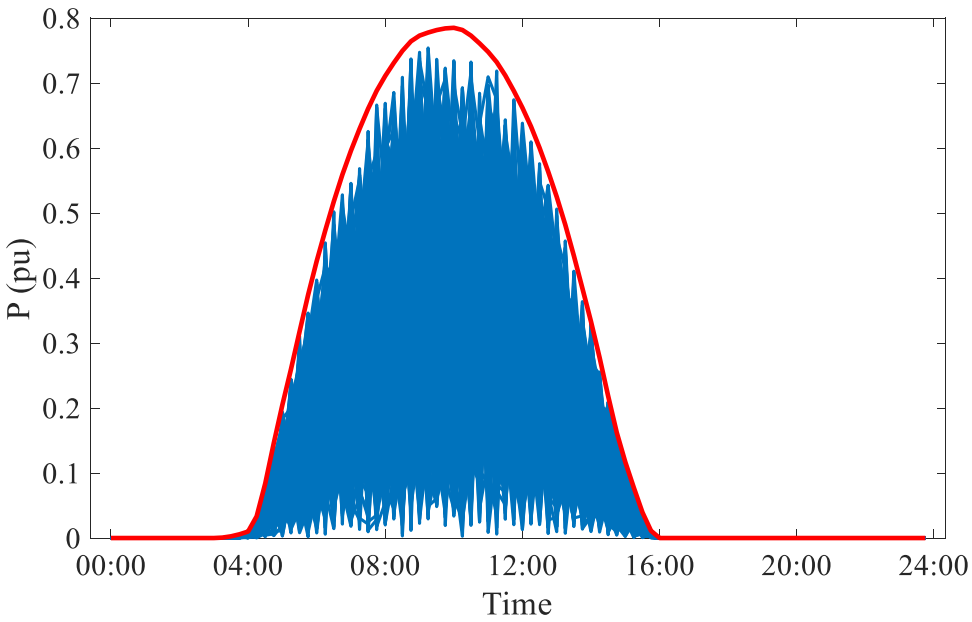

Fig. 4. Daily solar PV generation.

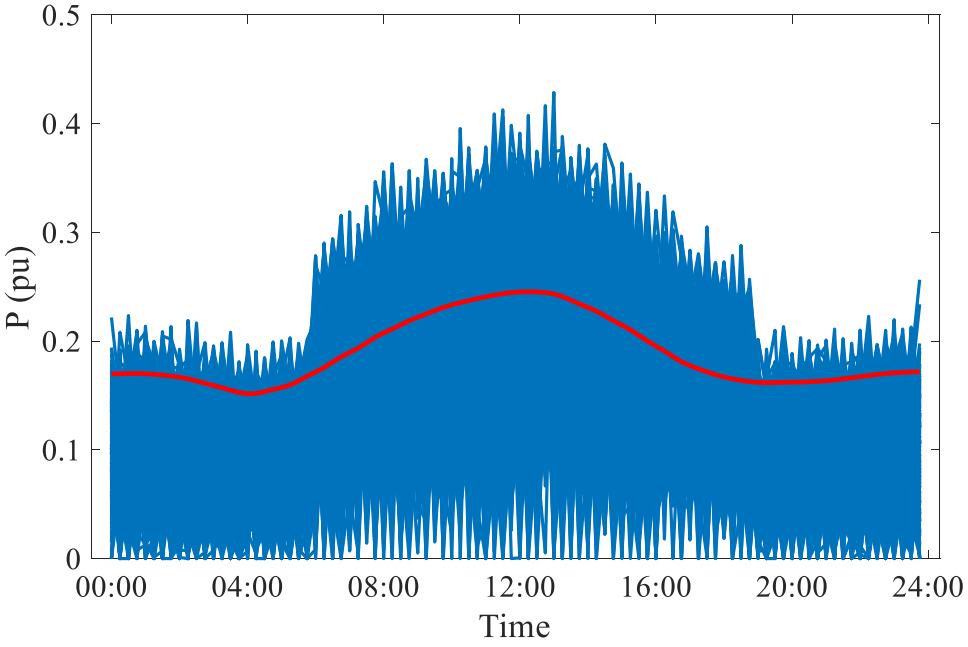

Fig. 5. Daily wind generation.

Regarding wind generation, the Weibull pdf of (2) is assumed to form the new set of the 365 daily profiles [29]:

$$f_{u_t}(u_t;\lambda,k)=\frac{k}{\lambda}\left(\frac{u_t}{\lambda}\right)^{k-1}e^{-(u_t/\lambda)^k}, \tag{2}$$

where k>0 is the shape parameter, λ>0 is the scale parameter and u_t represents the wind speed. In Fig. 5, the 365 pdf-based wind power profiles are summarized; the red line is the ART model profiles used to generate the new profiles. The adopted parameters for the wind pdf are also given in Table I [29].

TABLE I. DRES PROBABILITY DISTRIBUTION PARAMETERS [29]

Wind farms (daytime)		**Wind farms (nighttime)**		**PV power station**	
k	λ	k	λ	α	β
1.73	7.18	2.06	7.41	2.06	2.5

4 Clustering-based Models

Simplified timeseries models can be also deduced by analyzing the distinct characteristics of the load and generation patterns of the whole dataset, instead of monthly or annual analysis and derive a reduced number of patterns, though suitable to represent efficiently the original dataset. For this purpose, a new modelling approach based on cluster analysis is proposed. k-means++ is used to group the original set of 365 patterns into a specific number of representative clusters as shown in Fig. 6. Although the analysis is flexible and can be generalized to

include the application of different clustering algorithms, in this paper k-means++ is selected as one of the most widely used clustering algorithms for profiling processing [21]. The data grouping is performed by minimizing the sum of squared errors (*SSE*) between the data and the associated clusters. k-means++ starts with a systematic procedure to initialize the cluster centroids. Since the initial centroids have been determined, the points of the original dataset are assigned to the closest centroid and the centroids are updated. The reassignment procedure terminates when the centroids remain constant.

4.1 Clustering techniques

Preliminary grouping of the data into subsets can improve the efficiency of the clustering procedure and reduce the size of the input dataset [22]. Therefore, two different techniques are examined, i.e., time-domain (TD) and principal component analysis (PCA). In TD-clustering, the dataset $\boldsymbol{\theta}$ defined in (3) is used:

$$\boldsymbol{\theta} = \lfloor \boldsymbol{\theta}_P, \boldsymbol{\theta}_Q, \boldsymbol{\theta}_{PV}, \boldsymbol{\theta}_{Wind} \rfloor, \tag{3}$$

where $\boldsymbol{\theta}_P$, $\boldsymbol{\theta}_Q$, $\boldsymbol{\theta}_{PV}$ and $\boldsymbol{\theta}_{Wind}$ are matrices of order 365×96 containing the load active and reactive power, PV and wind generation data, respectively. The resulting cluster centroids are the corresponding class representative profiles.

The second technique involves the projection of $\boldsymbol{\theta}$ onto its principal components in order to perform cluster analysis on the basis of a reduced dataset containing only the most important features of $\boldsymbol{\theta}$ and exploiting also the cross-correlations between the data. PCA is a linear method that consists of redefining the coordinates of a dataset to another coordinate system [20], [30], [31]. The total number of the principal components is equal to that of the original dataset and presents the same statistical information. However, PCA allows the reduction of all variables, since the first components retain the most important statistical information from the original data. In this sense, initially, the principal components of $\boldsymbol{\theta}_P$, $\boldsymbol{\theta}_Q$, $\boldsymbol{\theta}_{PV}$ and $\boldsymbol{\theta}_{Wind}$ are calculated resulting in the corresponding principal component matrices of order 365×96. The final set $\mathbf{p}$ is formulated as follows:

$$\mathbf{p} = \lfloor \mathbf{p}_P, \mathbf{p}_Q, \mathbf{p}_{PV}, \mathbf{p}_{Wind} \rfloor, \tag{4}$$

where $\mathbf{p}_P$, $\mathbf{p}_Q$, $\mathbf{p}_{PV}$ and $\mathbf{p}_{Wind}$ are the reduced data matrices, which are calculated as $\mathbf{p}_i = \boldsymbol{\theta}_i \mathbf{V}_i$ ($i = P, Q, PV, Wind$). Here, $\mathbf{V}_i$ represents the principal components of $\boldsymbol{\theta}_i$ and is defined as the eigenvectors of the covariance matrix $\boldsymbol{\theta}_i^T \boldsymbol{\theta}_i$. The order of each block $\mathbf{p}_i$ is selected to maintain at least 96% of the original dataset ($\boldsymbol{\theta}_i$)variance [35]. k-means++ is applied to $\mathbf{p}$ to

categorize the available data to clusters, according to their distinct characteristics. The harmonized representative load and generation patterns are calculated by averaging the corresponding 15-min profiles associated to each cluster [32].

4.2 Cluster analysis assessment

To evaluate the performance of the clustering methods and select the optimum number of clusters, a comprehensive methodology consisting of three steps is proposed (Fig. 6). At the first step, an initial estimation of the optimal number of clusters is performed by using certain metrics; in the second step, classification analysis of the cluster profiles with the original data is performed. The final set of clusters is determined in the last step. Results obtained from quasi-static simulations using the derived clustering-based models are evaluated. In more detail:

Step 1: k-means++ is an unsupervised machine learning process; thus, the optimal number of clusters must be determined by the user. An initial estimate of the optimal number of clusters can be provided by applying the elbow rule to the resulting within cluster sum of squares to between cluster variation (WCBCR) curve obtained by varying the number of clusters [21]. In this step, the compactness and similarity of each cluster is also evaluated by using the Silhouette method.

Step 2: To further improve the initial set of clusters, classification analysis is performed. A label ("low", "medium" or "high") is assigned to each clustering-based representative load and generation profile according to the corresponding demand/generation levels; the participation probability of each cluster is calculated by dividing the number of the daily profiles in each class with the total number of the daily profiles. The same procedure is also applied to sub sets $\boldsymbol{\theta}_{P}$, $\boldsymbol{\theta}_{PV}$ and $\boldsymbol{\theta}_{Wind}$. To ensure adequate representation of the original dataset characteristics and adequate distribution of the clustering-based patterns, the participation probabilities of the associated categories are compared. Since consistency in the results is observed, the proposed methodology steps forward; otherwise, steps 1 and 2 are repeated assuming an increased number of clusters.

Step 3: The last step towards determining the final set of clusters includes the evaluation of the results obtained from the quasi-static analysis of distribution networks. In particular, the system losses and network voltages are calculated by incorporating the developed cluster-based profiles into the quasi-static simulation model; results are compared with the corresponding obtained by applying the full timeseries dataset. The cluster assessment terminates

when the error is lower than a specific value, e.g., ~5%; in a different case, the algorithm starts over by increasing the number of clusters by one.

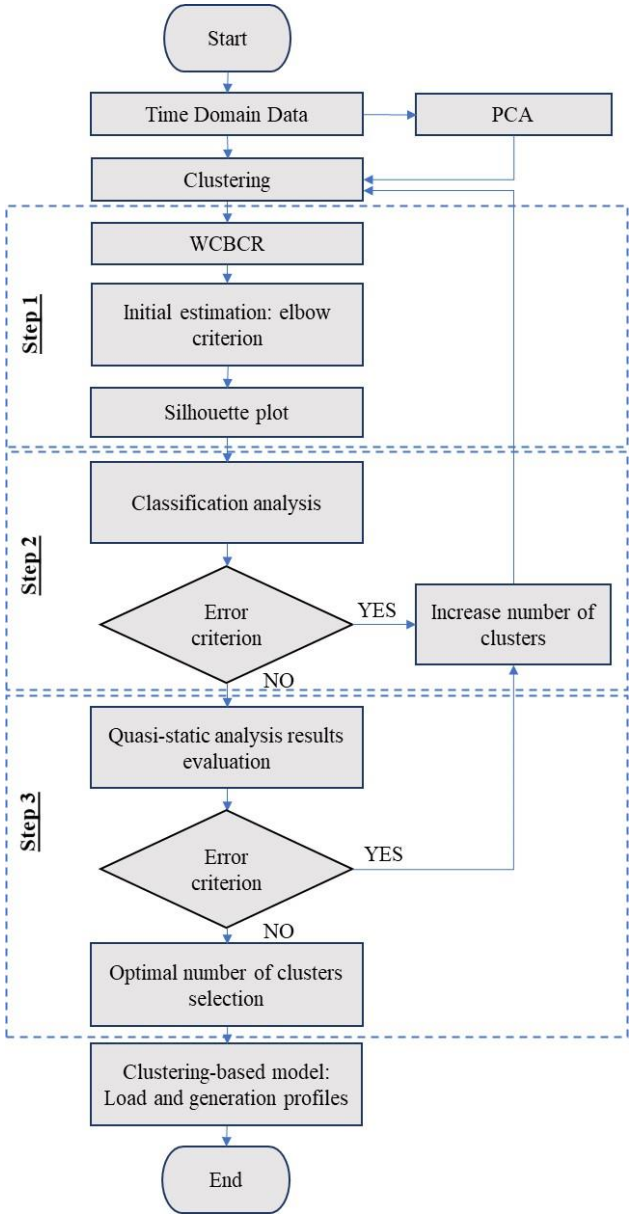

Fig. 6. Flow chart of cluster analysis methodology.

5 Results and comparative analysis

The impact of the different models on the performance of distribution networks is evaluated in terms of network voltages, losses and computational burden, assuming full timeseries (dataset of 35040 samples) results as reference [6]. Quasi-static simulations are performed in the 33-bus benchmark distribution network of [33] using MATPOWER [34]. The examined network operating cases include:

- Passive network: the original configuration of the network as presented in [33]. Generation is absent.

- PV ADN: a modified version of [33], where PVs have been added at specific nodes of the network.
- PV & WT ADN: a modified version of [33], where both PVs and WTs have been added at specific nodes of the network. In these nodes, the DRES installed capacity is equally split between PVs and WTs.

For both ADN cases two scenarios are examined:

1. *Scenario #1*: The DRES hosting nodes and rated power are given in Table II. These nodes are selected since they present the highest sensitivity to load changes [2]. The DRES rated active power is assumed equal to 50% of the load rated power at the corresponding node.
2. *Scenario #2*: The DRES hosting nodes and rated active power are summarized in Table III. This scenario is introduced to evaluate the performance of the modelling approaches under high DRES penetration levels leading to overvoltages.

Note that, in all cases, a unity power factor is assumed for the DRESs. For sake of simplicity, but with no loss of generality, the same normalized load and generation profiles are applied to all network nodes. The rated active power of all loads is equal to the original values of the benchmark case [33]. The load and generation profiles at each node are obtained by multiplying the rated power of each load or DRES unit with the corresponding normalized profiles of the examined models.

TABLE II. SCENARIO #1 DRES POWER GENERATION

Node	Rated power (MW)
#7	0.10
#14	0.06
#24	0.21
#25	0.21
#30	0.10

TABLE III. SCENARIO #2 DRES POWER GENERATION

Node	Rated power (MW)	Node	Rated power (MW)
#4	0.30	#3, #9, #10	0.50
#6, #14	0.80	#4, #8, #12	0.70
#11, #13	0.25	#2, #7, #15, #16	0.35

5.1 Passive Network

In the passive network case, only the active and reactive power load profiles of the examined models are used. Additionally to the above-mentioned models, a simplified version of the

full timeseries model is considered, i.e., the approximate reactive power (ARP) model. ARP is used to investigate the effect of reactive power modelling, since reactive power characterization in distribution systems becomes very important [32], [35]. However, most electricity load characterization techniques focus primarily to the analysis of active power data, assuming a typical load power factor cos(φ) value. In this sense, the load power factor is assumed constant and equal to 0.88 and 0.96, i.e., to the values corresponding to the 50th and the 75th percentile, respectively, of the original power factor values. The reactive power timeseries is calculated by means of $Q = P \cdot \tan(\varphi)$.

Regarding clustering-based modelling, the input data for the TD and the PCA clustering models are formulated as $\boldsymbol{\theta} = \lfloor \boldsymbol{\theta}_P, \boldsymbol{\theta}_Q \rfloor$ and $\mathbf{p} = \lfloor \mathbf{p}_P, \mathbf{p}_Q \rfloor$, respectively. The optimal number of clusters is determined by applying the proposed methodology. The initial estimate of the optimal number of clusters is 4 for both TD- and PCA-clustering models, as shown in Fig. 7. The corresponding Silhouette plots are depicted in Fig. 8, presenting mostly positive values, verifying the initial cluster number estimation. This is also evident by comparing the participation probabilities of the original dataset and the 4 extracted cluster profiles presented in Table IV. The "low", "medium" and "high" labelling is assigned according to the consumption characteristics of the examined profiles.

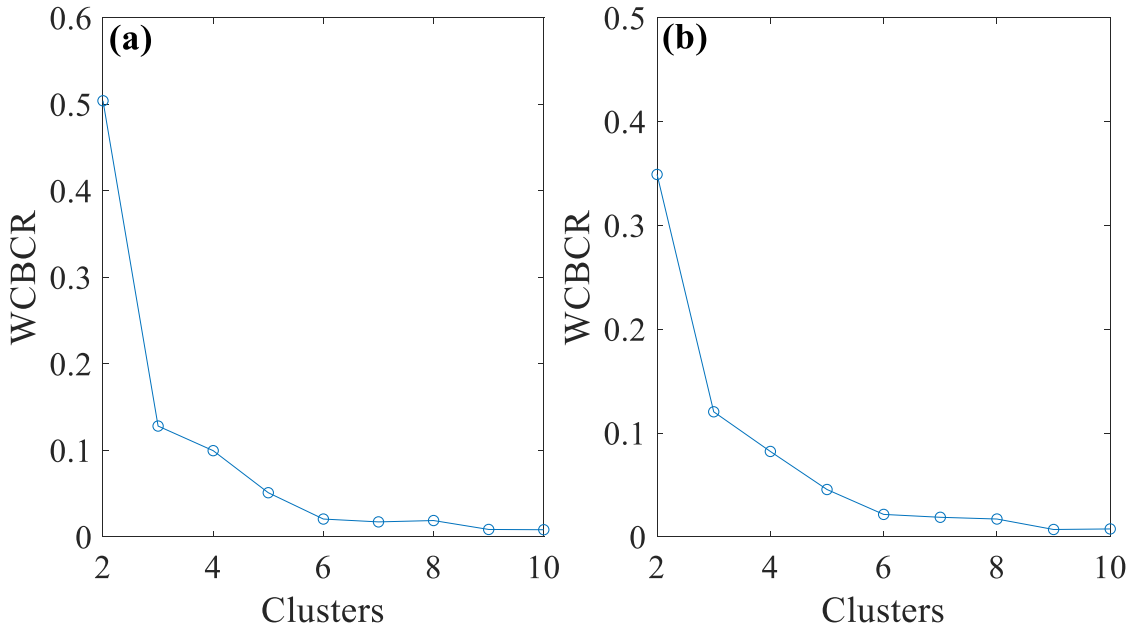

Fig. 7. WCBCR for a) TD-clustering and b) PCA-clustering.

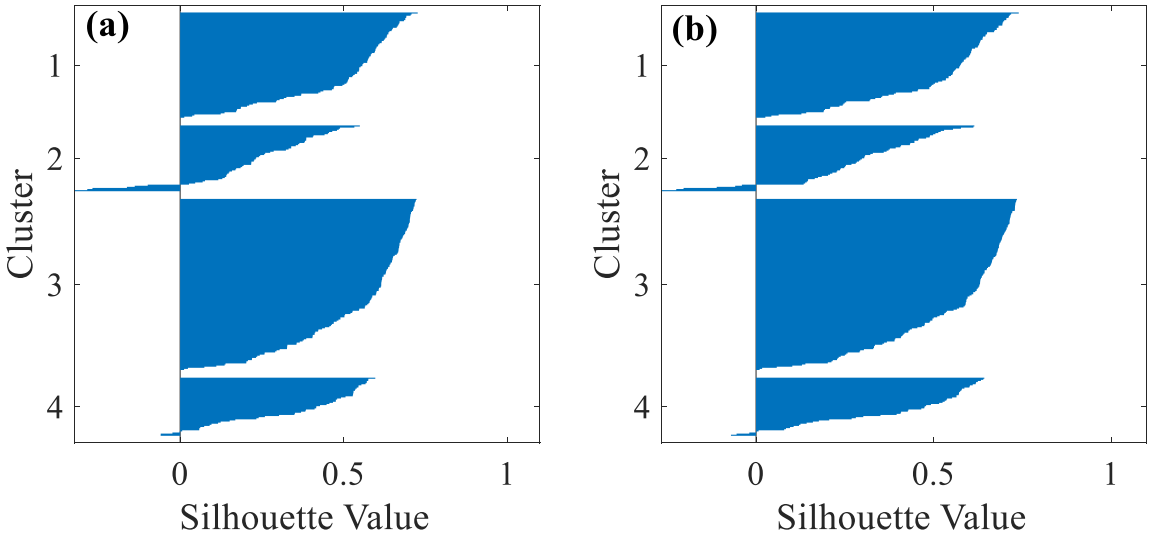

Fig. 8. Silhouette plots for a) TD-clustering and b) PCA-clustering.

TABLE IV. DETAILS OF CLUSTERING-BASED MODEL CHARACTERISTICS FOR THE PASSIVE NETWORK

Data categorization	Participation probability (%)		
	Original data (θ)	TD-Clustering	PCA-Clustering
Low	43.8%	42.7% (1 cluster)	42.7% (1 cluster)
Medium	31.0%	42.7% (2 clusters)	42.7% (2 clusters)
High	25.2%	14.5% (1 cluster)	14.5% (1 cluster)

The performance of the two clustering-based models is compared with the corresponding quasi-static simulation results obtained by the timeseries-based and the pdf-based models. First, the network voltages at nodes #18, #22, #25 and #33 are evaluated in Fig. 10; results are analysed by means of cumulative distribution function (CDF). These nodes are selected as the network terminal network nodes, presenting the lower voltage magnitude and the highest variance compared to the rest of the network nodes in the same branch. Generally, voltage results obtained by the different models exhibit similar behaviour to the full timeseries model. However, the MRT, ART and the two clustering-based models cannot accurately simulate extreme operational conditions, i.e., undervoltages. Such conditions occur rarely (CDF value < 0.05) and can be considered as outliers deviating from the typical network performance; thus, cannot be predicted by representative profile modelling approaches. A close comparison of the ARP results shows that in general both models approximate accurately the original CDF curve; however, the estimated network undervoltage level is higher compared to the full-timeseries model. On the other hand, results obtained by the pdf-based model show an overestimation of the upper voltage levels.

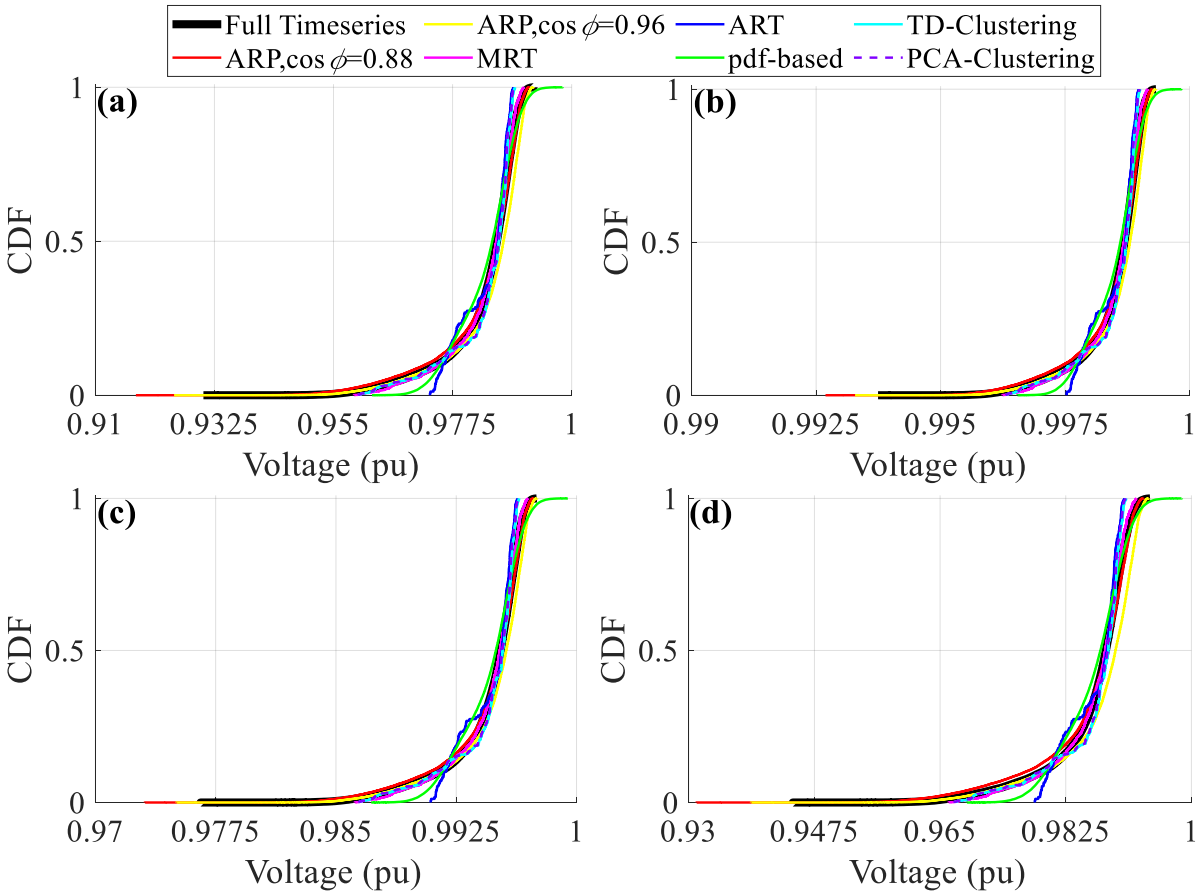

Fig. 9. Voltage CDF plots at nodes a) #18, b) #22, c) #25, and d) #33.

The effect of the different models to the calculation of the annual energy losses is evaluated in Table V; the percentage error of the active and reactive energy losses is calculated, assuming the full timeseries results as reference. For the full timeseries, the ARP and the pdf-based model, the energy losses are calculated directly as the sum of the resulting power losses occurred at each 15-min time sequence from the power flow simulations. For the MRT model the energy losses of each month are calculated using the resulting power losses of the WD and NWD profiles multiplied to the corresponding number of WDs and NWDs of the given month. Accordingly, for the ART model, the daily losses of the representative WD profile are multiplied by 209 (total number of WDs of the year), whereas the power losses obtained for the NWDs are multiplied by 156. Finally, the energy losses for the two clustering-based models are calculated by considering the power losses obtained for each cluster as well as the participation probability of the cluster to the whole year sample. It can be realized that accurate results are obtained by the ARP ($cos\varphi$=0.88) model. Small differences are also calculated with both clustering-based, the MRT, and the pdf-based models. However, differences higher than 5% are noticed for the ARP ($cos\varphi$=0.96) and the ART models.

In light of the above analysis, it is shown that the obtained results verify the accuracy of

both clustering-based models regarding the analysis of passive networks. The resulting representative patterns of the active and reactive power of the 4 clusters are summarized in Fig. 10.

TABLE V. ERROR OF ACTIVE AND REACTIVE ENERGY LOSSES CALCULATION FOR THE PASSIVE NETWORK SCENARIO

Losses	**ARP *cosφ* = 0.88**	**ARP *cosφ* = 0.96**	**MRT**	**ART**	**pdf-based model**	**TD-clustering**	**PCA-clustering**
Active Power	3.39%	5.92%	4.18%	9.53%	4.42%	4.82%	3.84%
Reactive Power	3.38%	5.88%	4.18%	9.53%	4.73%	4.82%	3.83%

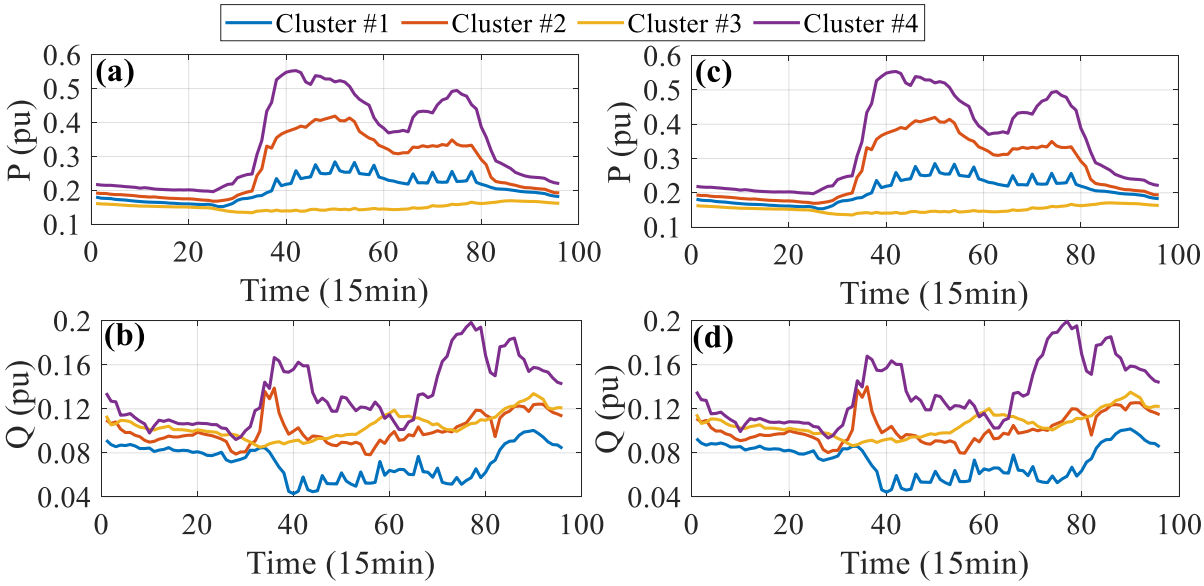

Fig. 10. Representative patterns for the load a) active and b) reactive power by TD-clustering, and for the load c) active and d) reactive power by PCA-clustering.

5.2 PV ADN

In PV ADN case, the input data of the TD- and the PCA-clustering models are $\boldsymbol{\theta} = \lfloor \boldsymbol{\theta}_P, \boldsymbol{\theta}_Q, \boldsymbol{\theta}_{PV} \rfloor$ and $\mathbf{p} = \lfloor \mathbf{p}_P, \mathbf{p}_Q, \mathbf{p}_{PV} \rfloor$, respectively. The initial number of clusters is estimated to 5 by using the Elbow and the Silhouette methods. However, due to significant discrepancy in the calculated participation probabilities between the clustering-based methods and the original dataset, the number of clusters has been increased. Note that 9 data categories are considered in the analysis as the different possible combinations of the load and PV data characterization to “low”, “medium” and “high”, respectively (see Table VI). Further investigations in terms of quasi-static analysis for the two operational scenarios resulted in determining the final set of clusters to 8. The representative patterns obtained by the two clustering-

based models are depicted in Fig. 11. The corresponding participation probabilities are compared with those of the original dataset in Table VI.

TABLE VI. DETAILS OF THE CLUSTERING-BASED MODEL CHARACTERISTICS FOR THE PV ADN CASE

Data categorization		Participation probability (%)		
Load	**PV Generation**	**Original data (θ)**	**TD-Clustering**	**PCA-Clustering**
Low	Low	6.0 %	12.1% (1 cluster)	14.0% (1 cluster)
Low	Medium	6.9 %	14.3% (1 cluster)	15.9% (1 cluster)
Low	High	31.0 %	26.0% (2 clusters)	27.4% (1 cluster)
Medium	Low	4.9 %	0.0% (no cluster)	0.0% (no cluster)
Medium	Medium	8.0 %	8.0% (1 cluster)	0.0% (no cluster)
Medium	High	18.1 %	18.9% (1 cluster)	19.2% (1 cluster)
High	Low	3.6 %	8.8% (1 cluster)	4.4% (1 cluster)
High	Medium	5.2 %	0.0% (no cluster)	8.8%(2 clusters)
High	High	16.4 %	12.1% (1 cluster)	10.4% (1 cluster)

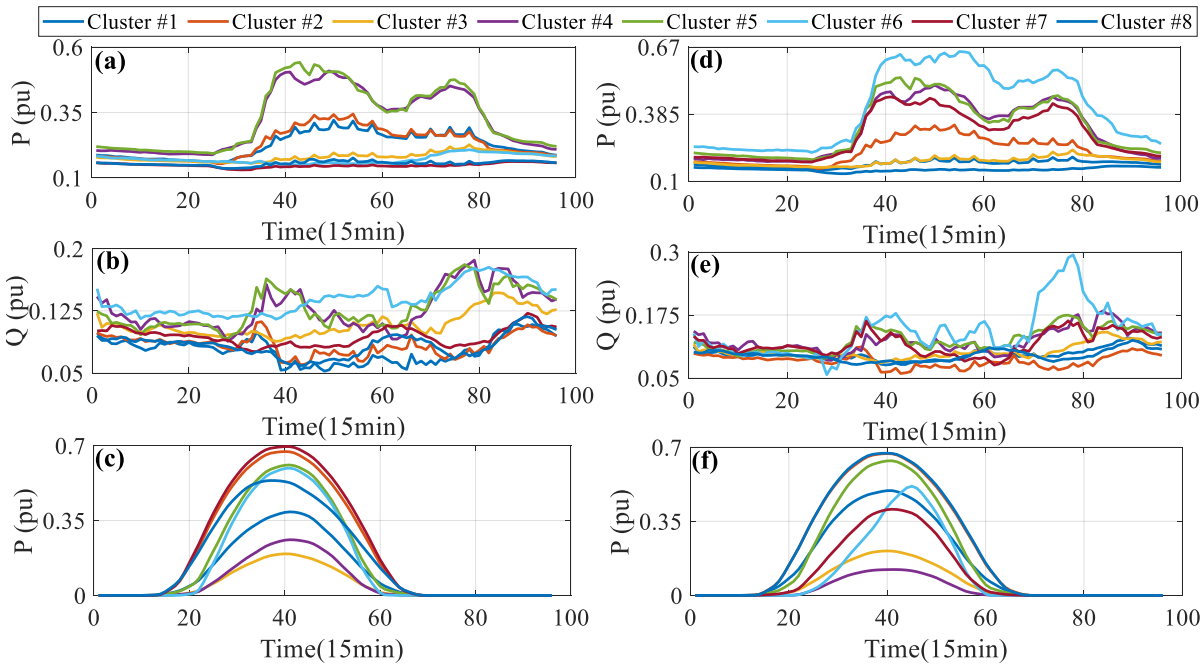

Fig. 11. Representative patterns for the a) load active power, b) load reactive power, and c) PV generation by TD-clustering, and for d) load active power, e) load reactive power and f) PV generation by PCA-clustering.

In Figs. 12 and 13, voltage CDFs obtained from the different models are compared for scenario #1 and #2. Network voltages at nodes #18, #22, #25 and #33 are considered. It can be realized that results with the different models generally approximate accurately the original CDF voltage curve; similar remarks as with the passive network analysis can be deduced. However, in scenario #2, the pdf-based model simulations present significant deviation regarding the prediction of the network overvoltages.

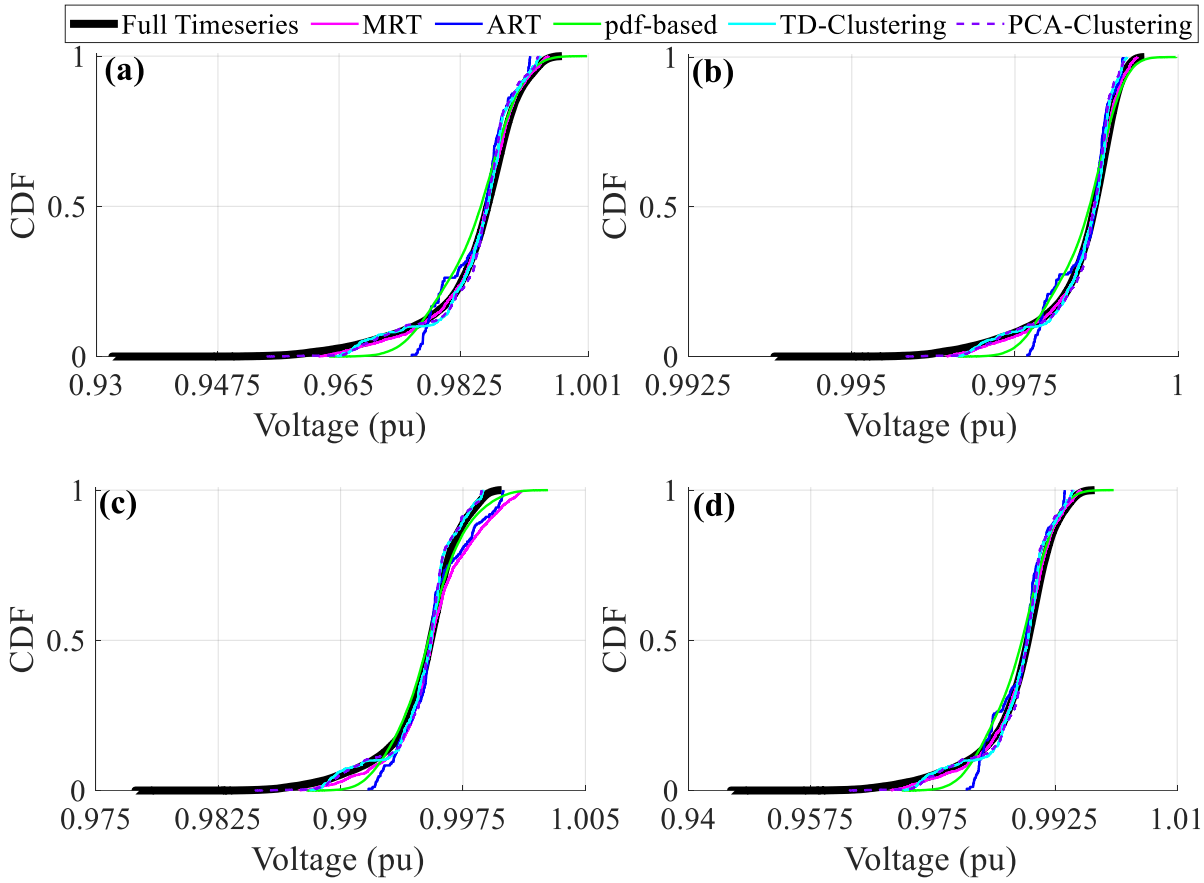

Fig. 12. Voltage CDF plots at nodes a) #18, b) #22, c) #25, and d) #33. Scenario #1.

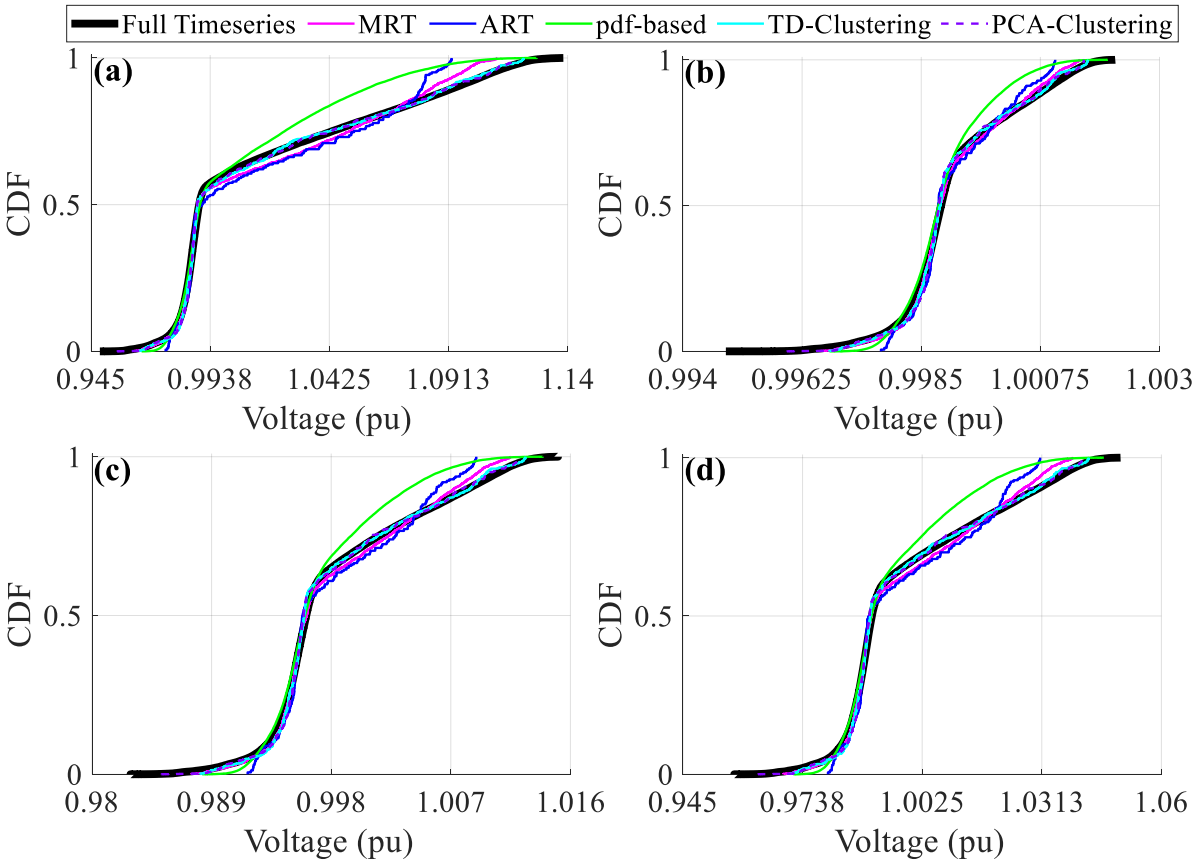

Fig. 13. Voltage CDF plots at nodes a) #18, b) #22, c) #25, and d) #33. Scenario #2.

Finally, in Tables VII and VIII the accuracy of the different models in the calculation of the active and reactive energy losses is evaluated. It is shown that clustering-based models predict very accurately the energy losses for both scenarios. Regarding the MRT and ART models, the error in energy losses calculations is higher, especially for scenario #2. In this scenario, also the pdf-based model predictions present significant error reaching ~50%, associated to the corresponding errors in the voltage calculations. This is attributed to the randomness of the load profiles and the increased stochasticity of the DRES generation of the pdf-based generation curves.

TABLE VII. ERROR OF ACTIVE AND REACTIVE ENERGY LOSSES CALCULATION FOR PV ADN. SCENARIO #1

Losses	**MRT**	**ART**	**pdf-based model**	**TD-Clustering**	**PCA-Clustering**
Active Power	7.48%	14.5%	4.2%	5.2%	5.6%
Reactive Power	7.14%	14.1%	4.0%	5.2%	5.5%

TABLE VIII. ERROR OF ACTIVE AND REACTIVE ENERGY LOSSES CALCULATION FOR PV ADN. SCENARIO #2

Losses	MRT	ART	pdf-based model	TD-Clustering	PCA-Clustering
Active Power	11.7%	19.1%	49.5%	4.3%	4.2%
Reactive Power	11.5%	18.7%	49.5%	4.2%	4.1%

5.3 PV & WT ADN

Similar to the previous cases, the number of clusters was initially determined to 6. However, due to the increased number of possible combinations of load, PV and WT generation levels, a higher number of clusters is investigated. On the basis of participation probability comparisons as well as by evaluating the quasi-static results obtained for scenario #1, it was shown that the optimal number of clusters can be set to 8. However, for scenario #2, small discrepancies in voltage and energy losses results were obtained by considering a significantly higher number of clusters, i.e., 18. The resulting representative patterns for the two clustering-based methods are depicted in Fig. 14.

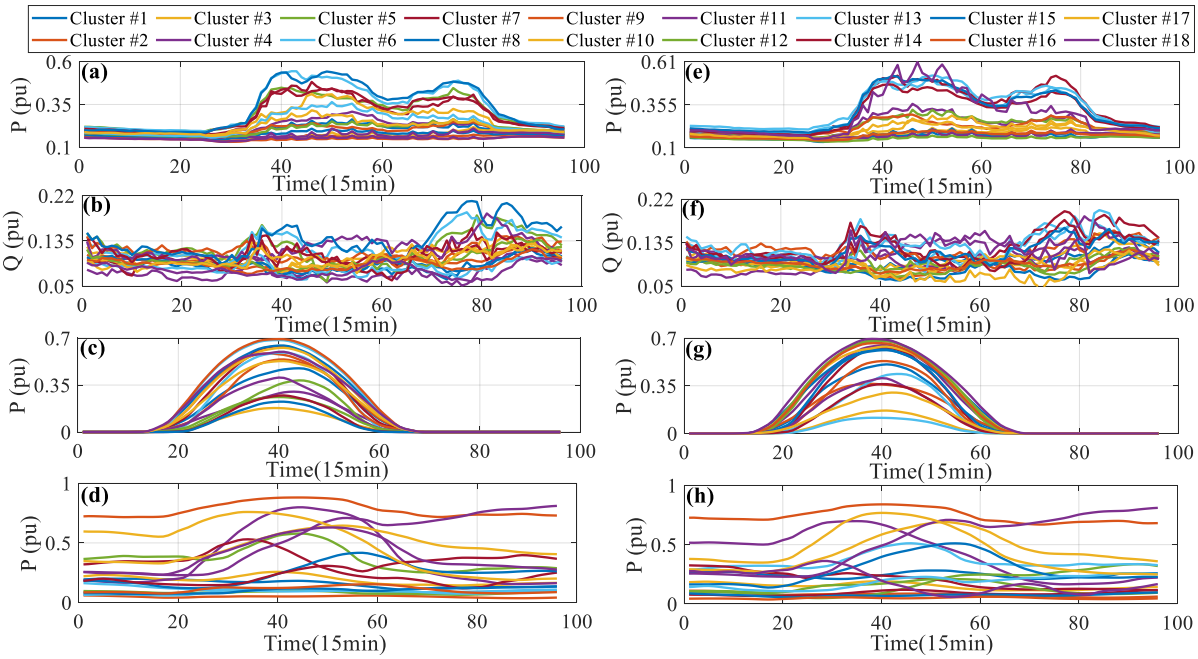

Fig. 14. Representative patterns for the a) load active power, b) load reactive power, c) PV generation and d) wind generation by TD-clustering and e) load active power, f) load reactive power, g) PV generation and h) wind generation by PCA-clustering.

Voltage results are presented in Figs. 15 and 16, for the two scenarios, respectively. Generally, voltage CDFs of the full timeseries model are accurately approximated by using the

different models. Small differences are observed for the clustering-based methods regarding scenario #1, since in scenario #2 the corresponding CDF plots overlap with those of the full timeseries model. Results in Fig. 16 reveal that network overvoltages are not accurately predicted by using the MRT, the ART and the pdf-based models.

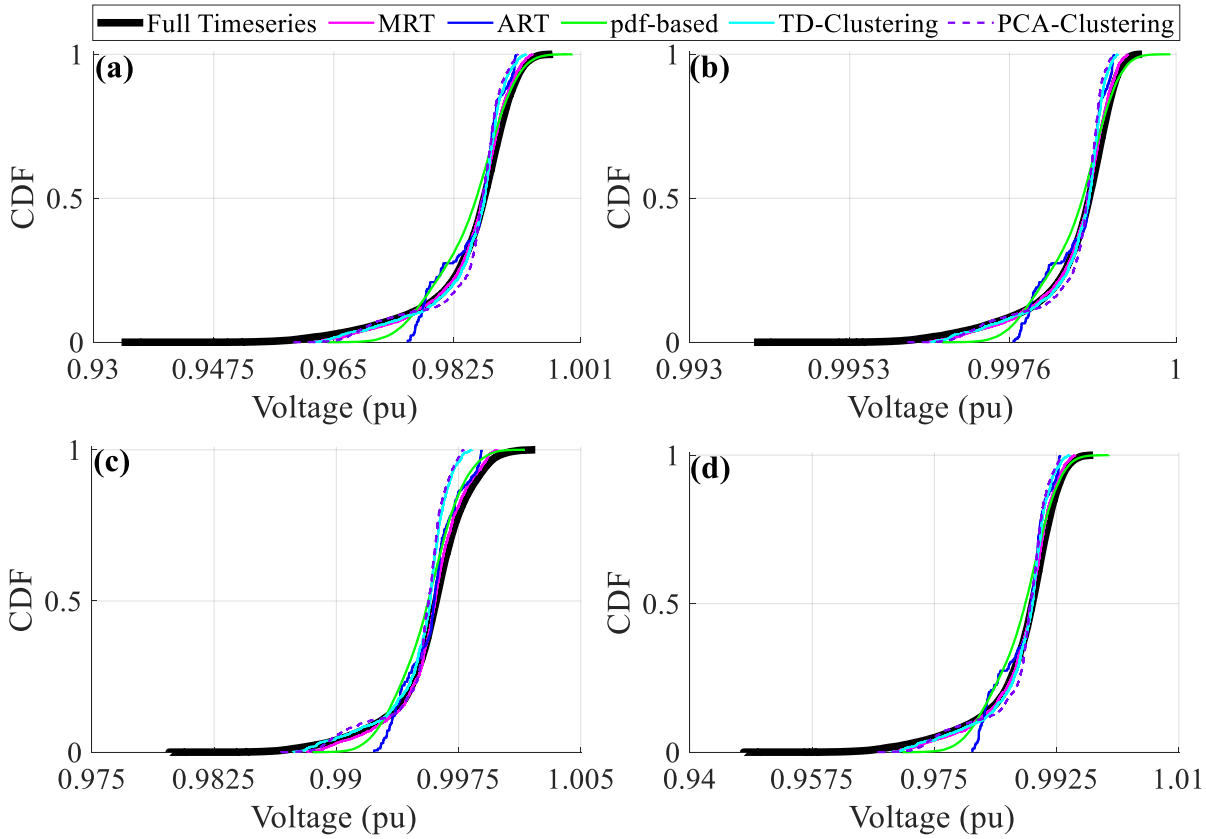

Fig. 15. Voltage CDF plots at nodes a) #18, b) #22, c) #25, and d) #33. Scenario #1.

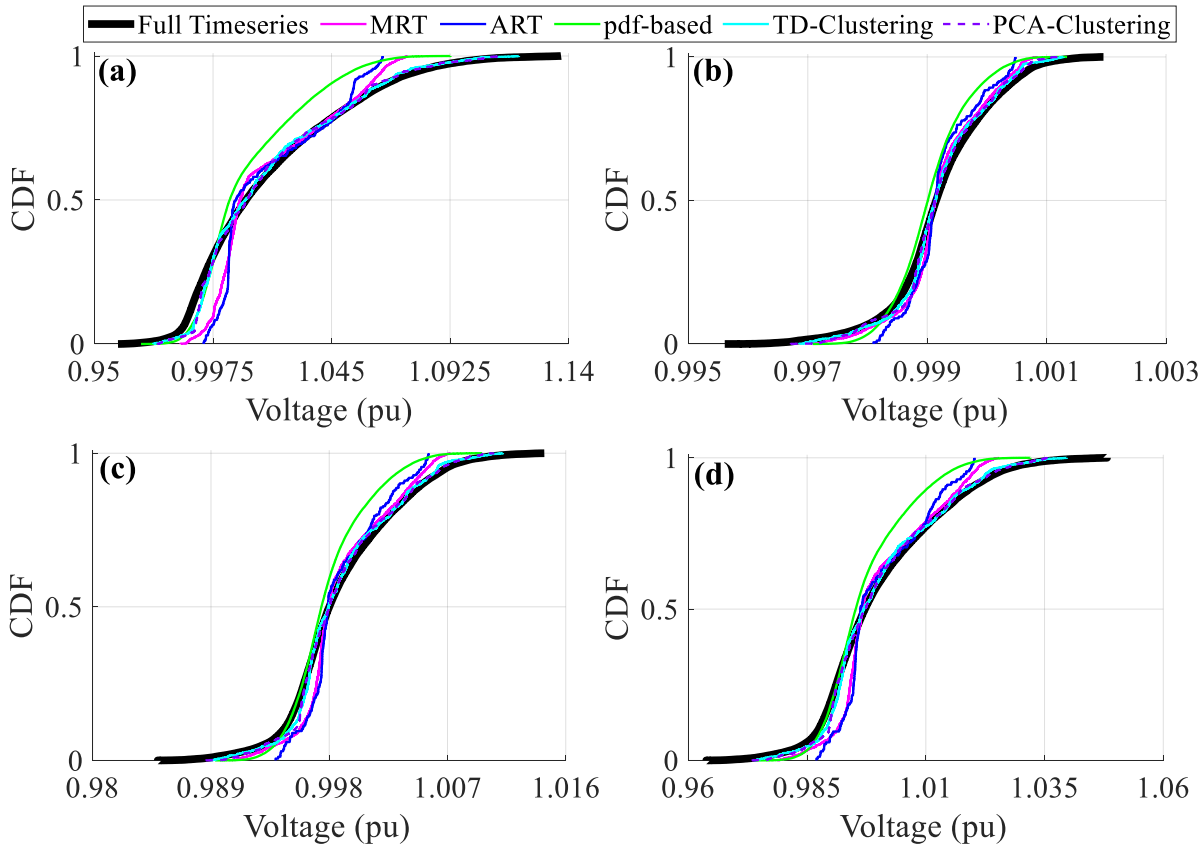

Fig. 16. Voltage CDF plots at nodes a) #18, b) #22, c) #25, and d) #33. Scenario #2.

The network losses are evaluated in Tables IX and X for the two scenarios, respectively. Small errors are calculated for the pdf-, the clustering-based and the MRT models regarding scenario #1. Higher calculation errors for all models are observed in scenario #2. However, satisfactory results are obtained only by using the two clustering-based models.

TABLE IX. ERROR OF ACTIVE AND REACTIVE ENERGY LOSSES CALCULATION FOR PV & WT ADN. SCENARIO #1

Losses	MRT	ART	pdf-based model	TD-Cluster-ing	PCA-Clustering
Active Power	5.9%	12.6%	2.1%	3.3%	4.0%
Reactive Power	5.8%	12.5%	2.2%	3.7%	4.4%

TABLE X. ERROR OF ACTIVE AND REACTIVE ENERGY LOSSES CALCULATION FOR PV & WT ADN. SCENARIO #2

Losses	MRT	ART	pdf-based model	TD-Cluster-ing	PCA-Clustering
Active Power	26.3%	31.6%	53.9%	8.2%	7.6%
Reactive Power	25.7%	30.8%	53.6%	8.0%	7.7%

5.4 Computational burden assessment

The major advantage of using load and generation simplified models is the reduced computational effort, especially in case of successive power flow calculations. Therefore, the computational time of the different models is compared in Table XI concerning the quasi-static analysis of all cases. Simulations were performed in an Intel® Core™ i7-7500U CPU @ 2.70GHz 2.90GHz processor and 8 GB RAM. Significantly lower computational burden is deduced by using the simplified MRT, ART and the clustering-based models; as expected, the pdf-based model results in comparable simulation time with the full timeseries model. In particular, the simulation time for the MRT is reduced by 93.06%. Note that for the full timeseries, MRT, ART and pdf-based models the computation time is similar for all cases, since the same number of patterns is used; thus, results in Table XI correspond to the average computational time of all cases. Concerning the clustering-based models, the simulation time is a multiple of the number of clusters, thus the computational time reduction ranges from 82% (18 clusters, i.e., PV & WT ADN case) to 96% (4 clusters, i.e., passive network case).

TABLE XI. COMPUTATIONAL TIME COMPARISON

Model	Simulation Time (s)
Full Timeseries	303
MRT	21
ART	1.7
pdf-based	289.4
TD and PCA clustering-based models	12-54

6 Temperature-dependent power flow assessment

Finally, to address the computational challenge imposed by conducting temperature-dependent power flows, the performance of the examined models is evaluated in the passive distribution network of Fig. 17. In conventional power flow analysis, the system branch resistances are typically assumed constant despite the fact that they are temperature sensitive, thus influencing also the branch loading and losses. For this purpose, dynamic calculations of the thermal conditions in distribution cables by including temperature correction of branch resistance is combined with power flow simulations by using the Matlab-OpenDSS interface [9],[10].

The examined network is a radial 20 kV distribution network depicted in Fig. 17. It consists of 11 nodes, connected via underground cables. Two backbone cables, i.e., between nodes 2

and 3 as well as 2 and 7 supply the two network branches. Table XII lists details of the MV cables. All network MV end-users consume energy. In particular, the load rated power at nodes #3 to #6 is 1600 kW and for nodes #7 to #11 1280 kW. The available normalized active and reactive power load profiles of the examined models are adjusted accordingly. The 150/21 kV transformer has rated power 20 MVA, short-circuit voltage 17.6%, whereas the full load losses are 0.38%. The transformer OLTC has 13 steps, allowing a secondary voltage regulation between -10.0% and +6.0%, in steps of 1.25%, in terms of the secondary rated voltage of 20 kV. The tap setting is determined in the power flow analysis, taking into account the regulation of the transformer to secondary node in the range of 20 kV to 20.6 kV.

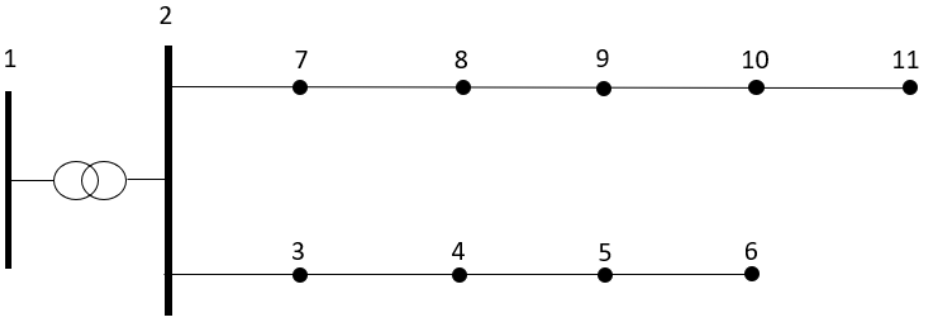

Fig. 17. MV distribution network topology.

TABLE XII. MV CABLE DETAILS

No	Connecting nodes	Cable Type	Length(m)
1	2-3	3x(1x240mm^2), XLPE	1000
2	3-4	3x(1x150mm^2), XLPE	467
3	4-5	3x(1x95mm^2), XLPE	558
4	5-6	3x(1x95mm^2), XLPE	535
5	2-7	3x(1x300mm^2), XLPE	1000
6	7-8	3x(1x240mm^2), XLPE	522
7	8-9	3x(1x150mm^2), XLPE	456
8	9-10	3x(1x95mm^2), XLPE	470
9	10-11	3x(1x95mm^2), XLPE	633

6.1 Energy losses

The total network losses, i.e., transformer and cable losses are calculated by using the different models; results are summarized in Table XIII. It can be realized that small differences are observed for most of the models regarding transformer losses, i.e., not exceeding 0.9 %; lower accuracy is only observed for the ART model, since the differences are 3.5%. Regarding cable losses, although higher differences are observed, results can be generally considered satisfactory; significant differences are only obtained with the ART model.

TABLE XIII. % ERROR OF ACTIVE ENERGY LOSSES

Network Compo-nent	MRT	ART	pdf-based model	TD-clustering	PCA-clustering
Transformer	0.66%	3.5%	0.56%	0.89%	0.76%
Cables	4.7%	23.45%	4.25%	5.75%	4.9%

6.2 Computational burden

The temperature-dependent power flow analysis involves successive iterations to determine the resulting temperature and consequently the conductor resistance in terms of the network loading conditions; thus, requiring significant computational time [9]. Therefore, a distinct advantage of the application of the simplified timeseries models is the reduced computational burden as shown in Table XIV. In this case, the simulations were performed in an Intel® Core™ i7-8700 CPU @ 3.20GHz 3.19GHz processor and 64 GB of installed memory (RAM). It is evident that the simulation time for most models has been significantly reduced (less than an hour); The pdf-based model is computationally intensive (exceeds 8.5 h), presenting similar computational burden with the full timeseries model; thus, hinders the practical implementation of temperature-dependent power flow analysis even in cases of small-area distribution networks, similar to the examined one.

TABLE XIV. COMPUTATIONAL TIME COMPARISON

Modelling Approach	Computational Time (h)	Reduction (%)
Full Timeseries	8.54	-
MRT	0.56	93.5
ART	0.04	99.5
pdf-based	8.78	-
TD and PCA clustering-based models	0.09	98.9

7 Discussion and conclusions

In this paper, the effect of different load and generation modelling approaches on the quasi-static analysis of distribution networks is investigated. For this purpose, simplified load and DRES generation timeseries-based models are considered as well as probabilistic analysis.

In addition, TD-clustering and PCA- clustering models have been introduced to identify representative patterns of load and generation profiles. To determine the optimum number of clusters, a three-step methodology is proposed. At the first step, the optimal number of clusters

is initially estimated by using certain metrics. In the second step, classification analysis of the cluster profiles with the original data is performed and the final set of clusters is determined in the last step after getting feedback from quasi-static simulations. The analysis verifies the importance of the two last steps for the determination of the optimal number of clusters, especially in case of ADN configurations. The examined cases include the quasi-static analysis of distribution networks under different operational conditions over one-year period. Finally, the computational efficiency achieved by using simplified models is evaluated for the computational intensive temperature-dependent power flow analysis of distribution networks.

The results obtained show that most of the examined models can approximate accurately the typical voltage and power losses behaviour of passive networks instead of using full timeseries. Small differences are mainly observed, since the examined models cannot accurately predict extreme operational conditions, though rarely occur. In this case, significant differences are only observed with annual representative profiling. Concerning reactive power modelling, it can be deduced that it is important in distribution network load modelling, especially regarding network power losses calculations.

Differences with the full timeseries results increase with the DRES penetration, due to the increased stochasticity and modelling complexity introduced by PV and especially by wind generation. Therefore, it seems that the monthly and annual representative timeseries as well as the pdf-based models should not be used for the analysis of overvoltages in ADNs, since in such cases marked divergences are observed. However, they can be considered as possible reliable solutions to provide indicative voltage and energy losses results in highly computationally complex problems in case of distribution networks with low DRES penetration level.

In particular, the proposed TD- and PCA-clustering models can predict accurately network voltages and power losses for all examined cases, contrary to all other simplified modelling approaches. In addition, significant computational time reduction can be achieved, depending on the number of the derived clusters. These findings indicate that the proposed clustering-based approach is robust and can be used for the analysis of distribution networks, especially for problems characterized by significant computational burden. Regarding the two different clustering-based modes, it is shown that both models can identify harmonized representative load and generation patterns, characterized by detailed separation of clusters. The main advantage of the PCA-clustering model relies on the reduced input dataset size, especially in case of multi-year dataset analysis and on taking into account the cross-correlations between data by means of principal components.

References

[1] IEEE Guide for Conducting Distribution Impact Studies for Distributed Resource Interconnection," in IEEE Std 1547.7-2013, pp.1-137, 28 Feb. 2014.

[2] Gkaidatzis PA, Bouhouras AS, Doukas DI, Sgouras KI, Labridis DP. Load variations impact on optimal DG placement problem concerning energy loss reduction. Electric Power Systems Research. 2017; 152: 36–47. https://doi.org/10.1016/j.epsr.2017.06.016.

[3] Mather BA. Quasi-static time-series test feeder for PV integration analysis on distribution systems. In: IEEE Power and Energy Society General Meeting. San Diego; 2012. https://doi.org/10.1109/PESGM.2012.6345414.

[4] Boehme T, Wallace AR, Harrison GP. Applying Time Series to Power Flow Analysis in Networks With High Wind Penetration. IEEE Transactions on Power Systems 2007; 22:951-7. https://doi.org/10.1109/TPWRS.2007.901610.

[5] Jain AK, Horowitz K, Ding F, Gensollen N, Mather B, Palmintier B. Quasi-Static Time-Series PV Hosting Capacity Methodology and Metrics. In: IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT). Washington; 2019. https://doi.org/10.1109/ISGT.2019.8791569.

[6] Martinez JA, Mahseredjian J. Load flow calculations in distribution systems with distributed resources. A review. In: IEEE Power Energy Society General Meeting. Detroit; 2011. https://doi.org/10.1109/PES.2011.6039172.

[7] Samet H, Khorshidsavar M. Analytic time series load flow. Renewable Sustainable Energy Reviews 2018; 82:3886–99. https://doi.org/10.1016/j.rser.2017.10.084.

[8] Reiman AP, McDermott TE, Akcakaya M, Reed GF. Electric power distribution system model simplification using segment substitution. IEEE Transactions on Power Systems 2018; 33:2874–2881. https://doi.org/10.1109/TPWRS.2017.2753100.

[9] Frank S, Sexauer J, Mohagheghi S. Temperature-Dependent Power Flow. IEEE Transactions on Power Systems 2013; 28: 4007-18. https://doi.org/10.1109/TPWRS.2013.2266409.

[10] IEC Electric Cables—Calculation of the Current Rating, International Electrotechnical Commission Standard Series 60287, Dec. 2006.

[11] Mumtaz F., Syed M. H., Hosani M. A., Zeineldin H. H. A Novel Approach to Solve Power Flow for Islanded Microgrids Using Modified Newton Raphson with Droop Control of DG, IEEE Trans. Sustai. Energy 2016; 7; 2; 493–503. https://doi.org/10.1109/TSTE.2015.2502482.

[12] Kryonidis G. C., Kontis E. O., Chrysochos A. I., Oureilidis K. O., Demoulias C. S., Papagiannis G. K., Power Flow of Islanded AC Microgrids: Revisited, IEEE Trans. Smart Grid; 2018; 9; 4: 3903-3905.

[13] Sharma T, Balachandra P. Model based approach for planning dynamic integration of renewable energy in a transitioning electricity system. International Journal of Electrical Power & Energy Systems 2019; 105: 642-59. https://doi.org/10.1016/j.ijepes.2018.09.007

[14] Lee KY, Vale ZA. Applications of Modern Heuristic Optimization Methods in Power and Energy Systems. Hoboken, New Jersey: Wiley-IEEE Press; 2020.

[15] Zhang S, Cheng H, Wang D, Zhang L, Li F, Yao L. Distributed generation planning in active distribution network considering demand side management and network reconfiguration 2018; 228: 1921-1936. https://doi.org/10.1016/j.apenergy.2018.07.054

[16] Ehsan A, Yang Q. State-of-the-art techniques for modelling of uncertainties in active distribution network planning: A review. Applied Energy 2019; 239: 1509–1523. https://doi.org/10.1016/j.apenergy.2019.01.211

[17] Pecenak ZK, Disfani VR, Reno MJ, Kleiss J Inversion reduction method for real and complex distribution feeder models, IEEE Trans. on Power Sys. 2018; 34; 2: 1161–1170.

[18] Lamprianidou IS, Papadopoulos TA, Kryonidis GC, Papagiannis GK, Bouhouras AS. Impact of Data-Driven Modelling Approaches on the Analysis of Active Distribution Networks. In: 54th International Universities Power Engineering Conference (UPEC), Bucharest; 2019. https://doi.org/10.1109/UPEC.2019.8893489.

[19] Atwa YM, El-Saadany EF, Salama MMA, Seethapathy R. Optimal Renewable Resources Mix for Distribution System Energy Loss Minimization. IEEE Transactions on Power Systems 2010; 25: 360-70. https://doi.org/10.1109/TPWRS.2009.2030276.

[20] Motlagh O, Berry A, O' Neil L. Clustering of residential electricity customers using load time series. Applied Energy 2019; 237: 11–24. https://doi.org/10.1016/j.apenergy.2018.12.063

[21] Tsekouras GJ, Hatziargyriou ND, Dialynas EN. Two-Stage Pattern Recognition of Load Curves for Classification of Electricity Customers. IEEE Transactions on Power Systems 2007; 22: 1120-8. https://doi.org/10.1109/TPWRS.2007.901287.

[22] Panapakidis IP, Papadopoulos TA, Christoforidis GC, Papagiannis GK. Pattern recognition algorithms for electricity load curve analysis of buildings. Energy & Buildings 2014; 73: 137–45. https://doi.org/10.1016/j.enbuild.2014.01.002.

[23] Elenkova MZ, Papadopoulos TA, Psarra AI, Chatzimichail AA. A simulation platform for smart microgrids in university campuses. In: 52nd International Universities Power Engineering Conference (UPEC). Heraklion; 2017. https://doi.org/10.1109/UPEC.2017.8231998.

[24] Greek Electricity Distribution System Operator. Medium Voltage Customer Data, https://meteringnet.deddie.gr; [accessed 18 August 2020].

[25] Pfenninger S, Staffell I. Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. Energy 2016; 114: 1251–65. https://doi.org/10.1016/j.energy.2016.08.060

[26] Staffell I, Pfenningerefan S. Using bias-corrected reanalysis to simulate current and future wind power output. Energy 2016; 114: 1224–39. https://doi.org/10.1016/j.energy.2016.08.068

[27] Renewables.ninja, https://www.renewables.ninja/; [accessed 18 August 2020].

[28] Tazi K, Abbou FM, Abdi F. Load analysis and consumptions profiling: An overview. Proceedings of the 1st International Conference on Electronic Engineering and Renewable Energy. ICEERE 2018. Lecture Notes in Electrical Engineering, vol 519. Springer, Singapore. https://doi.org/10.1007/978-981-13-1405-6_80

[29] Qin Z, Li W, Xiong X. Incorporating multiple correlations among wind speeds, photovoltaic powers and bus loads in composite system reliability evaluation. Applied Energy 2013; 110: 285–94. https://doi.org/10.1016/j.apenergy.2013.04.045

[30] Jolliffe IT. Principal Component Analysis. 2nd ed. Springer-Verlag; 2002.

[31] Korenius T, Laurikkala J, Juhola M. On principal component analysis, cosine and Euclidean measures in information retrieval. Information Sciences 2007; 177: 4893-905. https://doi.org/10.1016/j.ins.2007.05.027

[32] Yetkin EF, Ceylan O, Papadopoulos TA, Kazaki AG, Barzegkar-Ntovom GA. Active and Reactive Power Load Profiling Using Dimensionality Reduction Techniques and Clustering. In: 54th International Universities Power Engineering Conference (UPEC). Bucharest; 2019. https://doi.org/10.1109/UPEC.2019.8893443.

[33] Baran ME, Wu FF. Network reconfiguration in distribution systems for loss reduction and load balancing. IEEE Transactions on Power Delivery 1989; 4: 1401–7. https://doi.org/10.1109/61.25627.

[34] Zimmerman RD, Murillo-Sanchez CE. Matpower 6.0 User's Manual. Power Systems Engineering Research Center. New York; 2016.

[35] Kaloudas CG, Ochoa LF, Marshall B, Majithia S, Fletcher I. Assessing the Future Trends of Reactive Power Demand of Distribution Networks. IEEE Transactions on Power Systems 2017; 32: 4278-88. https://doi.org/10.1109/TPWRS.2017.2665562.