The DELTA project has received funding from the EU's Horizon 2020 research and innovation programme under grant agreement No 773960

Project Acronym:	**DELTA**
Project Full Title:	**Future tamper-proof Demand rEsponse framework through seLf-configured, self-opTimized and collAborative virtual distributed energy nodes**
Grant Agreement:	**773960**
Project Duration:	**36 months (01/05/2018 – 30/04/2021)**

DELIVERABLE D4.2
Energy Asset Segmentation

Work Package	WP4 – Adaptive Aggregator DELTA Supervisory Engine
Task	T4.2 – Energy Portfolio Segmentation
Document Status:	Final
File Name:	[DELTA]_D4.2_Final
Due Date:	31.08.2020
Submission Date:	October 2020
Lead Beneficiary:	HIT

Dissemination Level	
Public	X
Confidential, only for members of the Consortium (including the Commission Services)	

Authors List

Leading Author				
First Name		Last Name	Beneficiary	Contact e-mail
Ioannis		Koskinas	CERTH	jkosk@iti.gr
George		Karagiannopoulos	HIT	g.karagiannopoulos@hit-innovations.com
Co-Author(s)				
#	First Name	Last Name	Beneficiary	Contact e-mail
1	Stelios	Krinidis	CERTH	krinidis@iti.gr
2	Alexandre	Lucas	JRC	Alexandre.LUCAS@ext.ec.europa.eu
3	Evangelos	Kotsakis	JRC	Evangelos.KOTSAKIS@ec.europa.eu
4	Alexis	Fragkoulides	UCY	frangoullides.alexis@ucy.ac.cy

Reviewers List

Reviewers			
First Name	Last Name	Beneficiary	Contact e-mail
Andrea	Cimmino	UPM	cimmino@fi.upm.es
Carolina	Fernandes	Kiwi	cfernandes@kiwipowered.com

Legal Disclaimer

The DELTA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 773960. The sole responsibility for the content of this publication lies with the authors. It does not necessarily reflect the opinion of the Innovation and Networks Executive Agency (INEA) or the European Commission (EC). INEA or the EC are not responsible for any use that may be made of the information contained therein.

Copyright

Executive Summary

Management and exploitation of energy resources is a major priority for energy retail markets and utilities. The participation of residential customers in energy-projects expands the opportunities and benefits for both sides, not only for potential flexible demands, but also for distributed energy generation. However, at the same time, this integration of small and medium customers in energy markets increases management complexity. The requirements for the exploitation of these energy assets demand sufficient services that provide increased understanding of consumers/prosumers behaviour, grouping of customers in terms of their needs and delicate management in order to reach the next level in energy savings. This awareness of the customers' behaviour should be acquired progressively and Energy Portfolio Segmentation module is responsible to apply the segmentation techniques according to the aggregator's/retailer's policy and strategy, formulate new strategies and demand response policies and expand the business markets possibilities for the energy actors. All these strategies should be oriented to customers' needs, achieving the best services that can be offered to each user in order to extract the maximum potential flexibility from each one. Therefore, energy aggregators in the deregulated energy markets need to model and precisely define the customers' segments profiles in a comprehensive and clear way. A methodology to identify these energy profiles of customers and group them in larger scale segments is proposed in this deliverable. A variety of clustering algorithms and data pre-processing methods have been explored towards levering

The DELTA Asset Segmentation component is responsible for creating and arranging the DELTA Virtual Nodes, assign DELTA customers in them based on their overall static and dynamic characteristics towards enabling a more efficient portfolio handling. This tool is employed by the Aggregator's Decision Support System in order to timely allocate its small and medium customers to larger segments that can participate dynamically even in existing Demand Response markets.

Table of Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

Term	Description
DR	Demand Response
DVN	DELTA Virtual Node
FEID	Fog Enabled Intelligent Device
IEEE	Institute of Electrical and Electronics Engineers
AEPS	Aggregator's Energy Portfolio Segmentation
CDSS	Contractual Data Static Segmentation
TDDS	Temporal Data Dynamic Segmentation

1. Introduction

1.1 Scope and objectives of the deliverable

This deliverable is associated with the Energy Portfolio Segmentation assignment as described in Task 4.2 of the DELTA project. The report focuses on describing the methodology, the results and the process applied in order to manage the dynamic distribution of the Aggregator's assets (FEIDs) among the DVN Multi Agent Systems in the Energy Portfolio Segmentation module, and to identify the appropriate DVN candidate for a new customer in the Energy Portfolio Classification module. The system stability and efficient management of aggregator assets are the main objectives of this tool and the results are presented in the later section. Furthermore, this report studies the use of market segmentation and highlights the practical obstacles and research limitations/implications encountered providing an insightful explanation of how segmentation may be conducted.

1.2 Structure of the deliverable

The work presented in this deliverable is structured as follows:

- **Section 2** presents the literature review on the topic of Segmentation services for Energy Aggregators, including an overview of the DELTA Aggregator's Energy Portfolio Segmentation (AEPS) engine.

- **Section 3** presents the first sub-module of the AEPS, namely the Aggregator's Portfolio Classification, introducing the overall methodology and its functionalities, and how it addresses one of the core DELTA use cases,

- **Section 4** follows with the second sub-module of the AEPS, namely the Aggregator's Portfolio Segmentation that further enriches the initial creation of the DELTA Virtual Nodes based on actual dynamic characteristics. A detailed overview of the data and the methodology for the implementation is presented.

- **Section 5** presents the results of the Energy Portfolio Segmentation engine, with specific scenarios, and finally,

- **Section 6** concludes this report.

1.3 Relation to other tasks and deliverables

This deliverable is closely connected with Task 2.2 and the D2.2 that concerns an analysis of Demand Response strategies that are applied in the current Energy Markets while also providing a review of state-of-the-art research on Demand Response mechanisms that pertain to energy retail and Smart Grids. Additionally, this deliverable gives an overview of the formulation of DVN Multi Agent System that is examined in T3.2 and Deliverable 3.2 where further clustering mechanisms are applied over the assets of each DVN in terms of identifying patterns and opportunities for efficient management of energy resources. Finally, the resulting tool will be integrated to the Aggregator DSS which is part of activities of T4.4 and will be deployed at the pilot sites where it will be further evaluated under activities of T7.3 and T7.4.

2. Segmentation services for Energy Aggregators

2.1 Literature Review

The development towards Smart distribution Grids and the decentralization of the power systems requires the technology, modern buildings and other appliances to be energy efficient as well as energy flexible. The existence of flexibility in power systems is extremely crucial in order to facilitate integration of renewable energy sources and cover their intermittency with the Demand Response strategies. The achievement of the aforementioned integration requires data monitoring, which has been achieved through advance metering infrastructures such as smart meters. However, the wide variety of event information and the large volume of data pose high risks in operation and power distribution between electricity customers, which affects the reliability and the profitability of the power systems [1]. For this reason, clustering electricity customers based on their load consumptions is necessary, and an upcoming promising solution for risk elimination.

In this case, clustering/segmentation is a data mining technique where electricity customers are selected and categorized in various groups (clusters/segments) based on their load profiles. In addition, this method expedites the specification of intrinsic patterns in the big data sets that have emerged. Essentially, all the smart appliances have generated large volumes of data with limited information, and by clustering data and customers in small groups it will reduce the dimensionality in the customers' data sets, and provide quick access to useful information, directed to certain clusters. Mainly clustering advantages are for those who have access to power consumption data such as DSOs, aggregators and other decision support systems who are responsible for instant operations and fast decision making. The first surveys have been done by utilities, system operators and researchers, using the monthly usage and some fixed information (e.g. voltage levels, demand) categorized households and load profiles based on the following classes: demographics and socio-economic factors, dwelling characteristics, habits (e.g. consumption timing), energy conservation, energy efficiency goals, knowledge about electricity consumption and the attitude of use. Presently, data and detailed measurements for more than tens of thousands end-users are available and accessible. The stages of load pattern clustering are presented in the following diagram, which are analysed in [2].

In simple terms, the load profile data are collected in order to be processed and if needed repaired any missing data. Then, the input data are scaled down to obtain features that are necessary for the customers' customization. Moving to the clustering stage, methods and algorithms are applied on load patterns for accurate parameters selection and have the optimum outcome. The clustering performance assessment essentially ensures that each cluster is unique and well separated from other clusters. A post processing analysis for the generated clusters is done at the 'formation of customer classes' stage, which is based on real-life cases, such as cluster-specific tariffs or DR programs. The final stage is where the retailer or DR aggregator will apply those clustering results.
The major clustering methods as described below:

- **Hierarchical clustering** groups data, simultaneously over a variety of scales, by creating a cluster tree. The tree is a multilevel hierarchy, where clusters at one level are joined to clusters at the next level.

  To perform hierarchical clustering, it is necessary to find the similarity or dissimilarity between every pair of load profiles in the data and then group them into binary clusters based on the previously computed similarity matrix. The process is iteratively repeated by merging the clusters of each level into bigger ones at the upper level until all samples are grouped into expected clusters. The advantage of this method is that the original data is kept unchanged in the root of the cluster tree.

- **K-means clustering** method groups load profile data by determining a certain number of clusters and a centre point for each cluster. After determining the centre point of each cluster, each data set should be assigned to the nearest centre point then a recalculation of the new centre point will be done iteratively until the position of the centre point is stable.

- **Fuzzy C-means Clustering** is similar to standard K-means, the difference is that each data set has a degree of membership to each initial cluster (i.e. each data set belongs to all clusters to some degree). The degrees of membership for each data set to all clusters should sum to one. Firstly, the number of clusters and guessing the cluster centre point (most likely incorrect), which is intended to mark the mean location of each cluster is selected, then every dataset is assigned a membership grade for each cluster. The next step is updating each cluster centre point and membership grade iteratively until the position of the centre point is constant. In this step the cluster centre point moves iteratively to the correct position within the data sets. The Fuzzy C-means clustering technique does not create boundaries between data sets for the first iteration, because the clustering process involves all data. The boundaries will automatically evolve when the clustering process is completed.

- **Spectral Clustering** algorithm is a graph based algorithm that exploits the utilization of an adjacency matrix that describes the similarity distance between two energy assets in order to identify clusters through the extraction of the Laplacian analysis (eigenvalues) of the former table [32].

- **Gaussian Mixture Model (**GMM**)** is a probabilistic approach that applies soft clustering separation, focused on identifying a mixture of Gaussian Distributions. Each distribution represents a cluster center identical to KMeans that is expressed from the mean value, the covariance and its size. Hence, this algorithm identifies the matching probability of each data point to the corresponding distribution through Expectation-Maximization algorithm [35].

The following table shows some well-known clustering techniques that are utilized for customer segmentation in power systems [30].

Table 1. Clustering techniques used for power systems.

Clustering Techniques	References	Definition
K-means (KM)	[3] [4] [6] [7] [8] [9] [10]	[5]
Variations of K-means including K-medoids, K-medium and Fuzzy C-means (FCM)	[4] [10] [9] [11] [12] [13] [14]	[11] [14]
Adaptive K-means	[15] [16] [17]	[17]
Hierarchical (H)	[3] [4] [6] [10] [12] [15] [18]	[5]
Self-Organizing Maps (SOM)	[4] [19] [8] [20] [12] [21]	[12] [21]

Modified Follow the Leader (FDL)	[3] [22] [23]	[22]
GMM - Expectation maximization (EM)	[24] [25] [13]	[24]
Online clustering/ parallel clustering divide and conquer	[16] [17]	--
Spectral Clustering	[32][33][34]	[32]

The selection of the algorithms depends on the available data, for example, the *number of clusters are necessary to be pre-determined* for the K-means, Fuzzy K-means, whereas hierarchical, adaptive k-means and modified Follow the Leader do not require this parameter. In addition, hierarchical is the only method that does not require an iterative *process*, however, it does generate *boundaries between data sets*, as well as K-means and Follow the Leader. On the other hand, Fuzzy K-means and Fuzzy relations do not create boundaries. Moreover, the suitability of distance measures varies for each algorithm (e.g. the most used measure for K-mean is L_2 norm, for K-medoids L_1 is minimized). Therefore, the computation time and the complexity for each technique is different. Another comparison between these methods is the *trial and error approach* which is applicable for the Follow the Leader and Fuzzy relation, but for hierarchical, K-means and Fuzzy K-means is not [29].

Table 2 below shows the most common data sets where clustering techniques are used [30].

Table 2. Clustering application on time-series, features, or reduced data set.

Method	Examples	References
Raw consumption (time series) data	--	[8] [9]
Feature Definition: Data that are defined by the user, based on load shapes and specific application	Load factor; daily $P_{av,day}/P_{max,day}$ mean; standard deviation; skewness	[25] [26]
Feature extraction: Extracted data from load shapes, use of techniques (e.g. frequency domain analysis)	Discrete Fourier Transform (DTF); harmonics-based coefficients; discrete wavelet transform	[23] [27]
Data size reduction: Obtained from the original data	Principal component analysis (PCA); Sammon map; symbolic aggregate approximation (SAX)	[17] [28]

Another comparative analysis for clustering techniques was done in [31], where the taxonomy of clustering approaches was detailed, explained with their variants and their similarities, and finally evaluated based on certain criteria. The table 3 below outlines some of the general outcomes of this comparative analysis [31].

Table 3. Comparative study of some clustering algorithms.

Category of Clustering	Algorithm name	Time complexity	Scalability	Suitable for large scale data	Suitable for high dimensional data	Sensitive of noise/outlier
Partition	K-means	Low O(knt)	Middle	Yes	No	High
	PAM	High O(k(n-k)^2))	Low	No	No	Little
	CLARA	Middle O(ks^2+k(n-k))	High	Yes	No	Little
	GMM+EM	O(nk^2)	Middle	Yes	Yes	Little
Hierarchy	CLARANS	High O(n^2)	Middle	Yes	No	Little
	BIRCH	Low O(n)	High	Yes	No	Little
	CURE	Low O(s^2*logs)	High	Yes	Yes	Little
	ROCK	High O(n^2*logn)	Middle	No	Yes	Little
	Chameleon	High O(n^2)	High	No	No	Little
Fuzzy Based	FCM	Low O(n)	Middle	No	No	High
Density Based	DBSCAN	Middle O(n^logn)	Middle	Yes	No	Little
Graph Theory	CLICK	Low O(k^f(v,e))	High	Yes	No	High
Grid Based	CLIQUE	Low O(n+k^2)	High	No	Yes	Moderate

A brief review of the most used clustering algorithms for customer segmentation, a review of current research and possible applications of clustering techniques for power systems were presented and discussed in this section. Within the scope of DELTA, advanced clustering/segmentation techniques are developed towards delivering a state-of-the-art module that will arrange customers into the DELTA Virtual Nodes for fast applications such as dynamic Demand Response. Exploitation of dynamic and static characteristics will allow the dynamic creation and maintenance of the DELTA Virtual Nodes, arranging the customers' portfolio in an optimal manner towards leveraging the overall flexibility potential.

2.2 DELTA Segmentation Functional Overview

The Aggregator's Energy Portfolio segmentation (AEPS) module is responsible for creating and updating large virtual customers DVNs consisting of small/medium customers.

Portfolio segmentation/classification deployment includes:

- Initial allocation of FEIDs into Nodes, similar to the clustering mechanism, utilizing the contractual data (e.g. geographical, type of customer, consumption capacity, generation capacity, market type). Most common way to achieve the assignment of new customers: train supervised classifiers algorithms according to specified strategies using the labels as target variables.

- Distribution of FEIDs into Nodes, using the contractual data and the features that are extracted from time series data (consumption, generation, flexibility etc.) and the calculated reliability and availability profile of the FEIDs.

The following image depicts the workflow of the Energy Portfolio Segmentation/Classification module. The component of the segmentation tool receives as input in real time and historical measurements and a clustering process is conducted in order to identify FEIDs with similar profiles. In the case of the arrival of a new FEID, a classification model designed from unsupervised data is activated and the FEID is assigned into one of the existing DVNs. Meanwhile, the Segmentation tool periodically checks for changes into the DVNs profiles. Factors that contribute to the alteration of the profile of a cluster can be the decision of an Aggregator to change the business model, the addition, the removal or the behavioural change of a critical FEID.

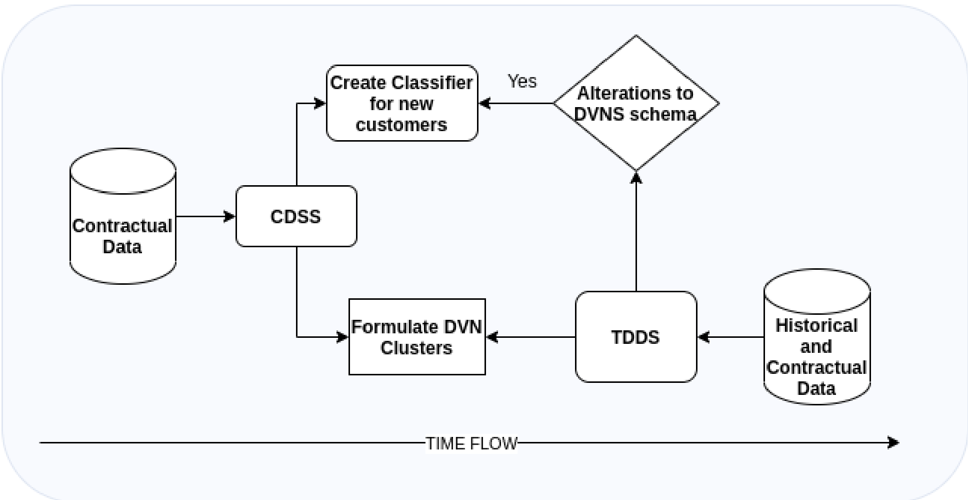

Figure 1. Aggregator's Energy Portfolio Segmentation/Classification workflow.

3. Contractual Data Static Segmentation (CDSS)

3.1 CDSS Methodology

The CDSS submodule is designed to classify customers (FEIDs) according to their intrinsic characteristics. Therefore, it is necessary to examine the contractual features of the already grouped assets in order to create a dataset that is oriented to the current structure. In that way, the learning process of the classifier recognizes the distinguishing features from the FEIDs' contractual information and links it with the TDDS results that include historical information.

One crucial constraint that affects the initial schema of the aggregator's portfolio is the Geographical location of each FEID. This information is represented from the Sector Parameter (Sector1, Sector2, etc.). Latitude and longitude of the location of each FEID is discretized in Sector domains according to the network topology. Therefore, the aggregator's portfolio that belongs to a specific Sector is examined individually and it is considered as an independent entity in terms of segmentation. In particular, location affects the constitution of DVNs, preventing FEIDs from different Sectors to participate in the same DVN as it is displayed below.

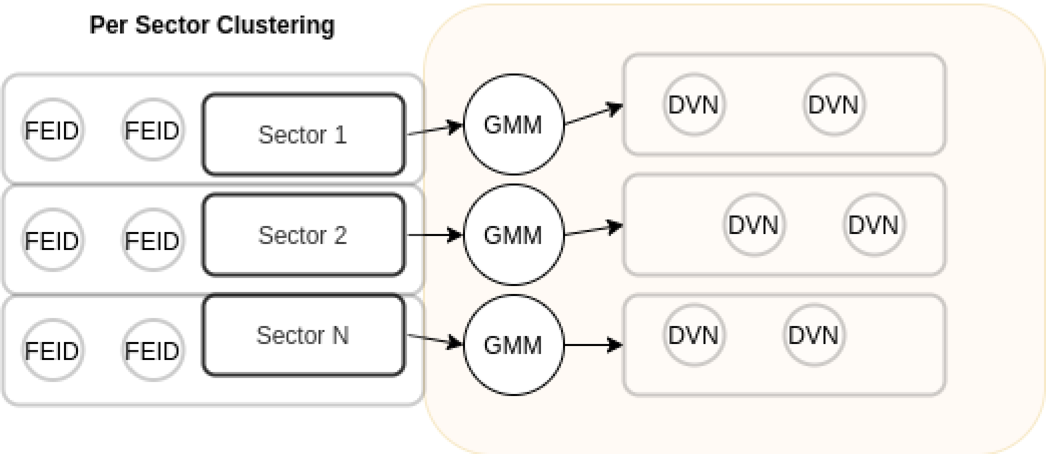

Figure 2. Per Sector Segmentation with Gaussian Mixture Models

Support Vector Machines (SVM) algorithm has been selected as the proposed classification method regarding its efficiency and effectiveness compared to other tested methods. The SVM algorithm focuses on identifying the hyperlane that distinguishes the groups of FEIDs that belong to the same DVN, maximizing the margin from the nearest data point of each group. This property incorporates scalability and sustainability to the existing ecosystem.

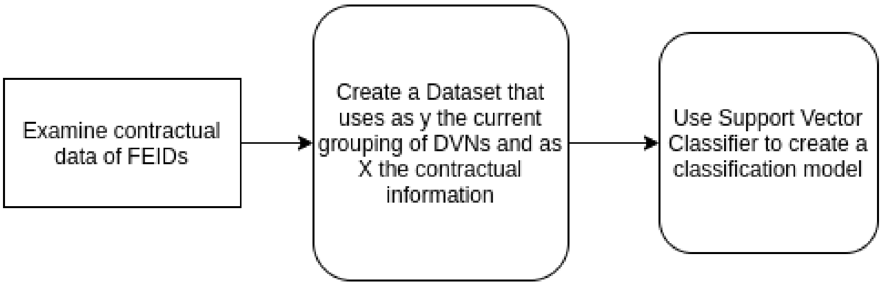

Figure 3. CDSS workflow.

3.2 Classification Process Initialisation

In the very early stage, that the aggregator initiates with an energy portfolio that dvns entities have not still been formulated, there are no historical data and no indicative labels for our dataset, the CDSS submodule distributes FEIDs among DVNs through unsupervised clustering methods over the contractual data.

One of the examined clustering algorithms that showed notable results is DBSCAN. The main advantages of this algorithm is related to its ability to examine density based spatial distances, identifying the appropriate number of clusters autonomously. Two parameters that are of substantial importance for the configuration of this property are eps and min_samples parameters. The first one is responsible for the definition of the maximum distance between two points that belong to the same neighborhood, while min_samples describes the minimum number of points that constitute a cluster. The values of these parameters that achieved the most efficient results are eps=0.6 and min_samples=10.

Going beyond common clustering approaches that are considered "hard clustering", such as the DBSCAN, and toward providing additional added-value to the DELTA AEPS, the overall approach was further enriched following a soft clustering algorithm that is capable of assigning multiple data points (FEIDs) to multiple clusters (DVNs). Furthermore, this type of clustering method estimates the probability that each data point belongs to a specific cluster or more groups. This property is further exploited as a feature that will link temporal re-formulation of the DVNs regarding the balancing of DVNs' reliability. In particular, in the case scenario that clustering results formulate one DVN that is very unreliable (the mean reliability of all FEIDs is low), through the aforementioned property of soft clustering algorithms, the intersection between two clusters is identified and swaps FEIDs with contradictory reliability that belong to an intersection area among two DVNs. This is executed in terms of supporting unreliable DVNs and creating in overall more reliable clusters. The following figure illustrates this interaction.

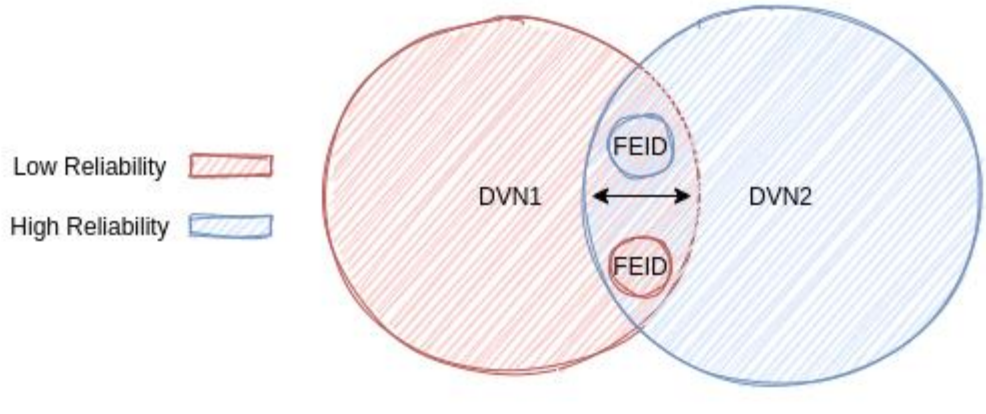

Figure 4. FEIDs swap for DVNs' reliability balance

GMM and Expectation Maximization algorithm as a soft clustering probabilistic method is the proposed approach. This method differentiates from other clustering algorithms regarding the perspective it approaches the problem. This type of Segmentation is not related with the distances of the data points with nearest centroids or the spatial density of the data. In particular, it fits a set of k Gaussians distribution to the energy profiles and extracts the corresponding parameters, such as mean value, variance and weight of the cluster. This method is achieved through the Expectation-Maximization algorithm that is an iterative process that identifies the optimal Gaussian parameters.

Selecting the optimal number of generated clusters was examined through the Bayesian Information Criterion (BIC). A grid search of the number of clusters over several GMM models extracted the corresponding BIC values. This criterion has been proved effective in model based clustering [36]. Lower values of BIC denote higher efficiency, however the key point that contributes to the final selection of clusters is the elbow point of the curve.

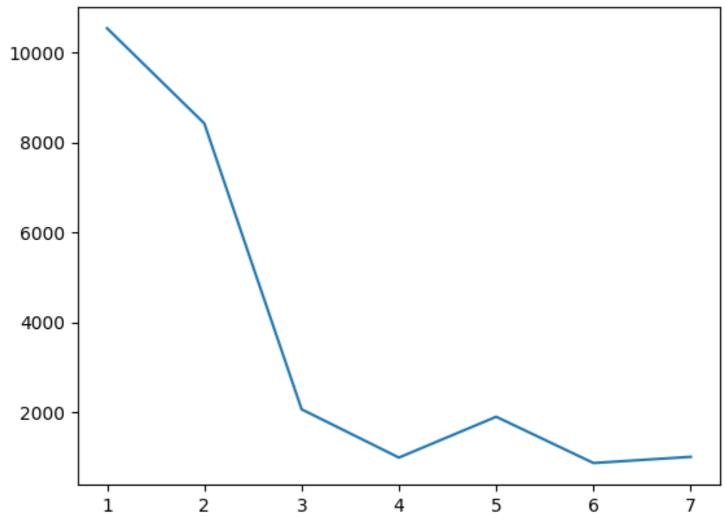

Figure 5. BIC values for several number of clusters

In order to avoid instabilities regarding the initialization points of GMM, the proposed implementation attempts to incorporate latest solutions as the initialization point of new fits. This parameterization can speed up the convergence of the algorithm and create linkages between the temporal alterations of the aggregator's portfolio schema. As a result, the clustering results are not temporally detached.

This process exploits information about the intrinsic characteristics of each asset and generates larger entities with common behaviour. This scenario describes the initial DELTA execution over the first customers' participation.

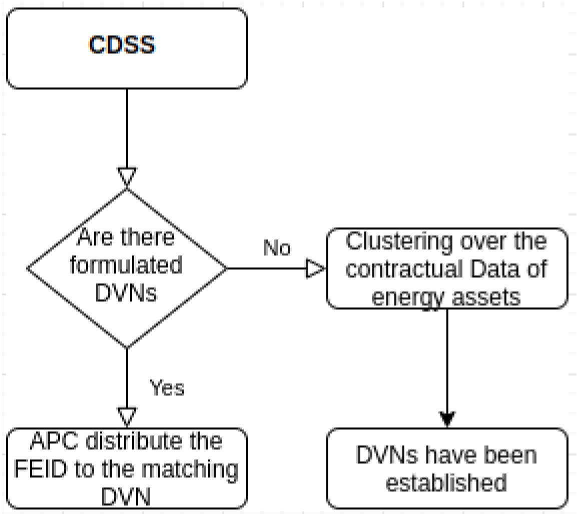

Figure 6. Workflow - Initial formulation of DVNs through unsupervised clustering over contractual data.

The CDSS module was applied over the DELTA Aggregator's portfolio as the primary step of the DVNs formulation process. DELTA portfolio consists of 150 FEIDs, that is a mixture of real FEIDs installed in a smart home and virtual FEIDs that simulate the behaviour and of real households with the corresponding contractual characteristics. Both the real and virtual FEIDs communicate with aggregators and send data in a minutely granularity. The CDSS module over the aforementioned data formulated two DVNs, while the estimated silhouette score of the generated clusters is near 0.61. This metric indicates that the clusters are well defined, while the groups of FEIDs belong to the same sector as well.

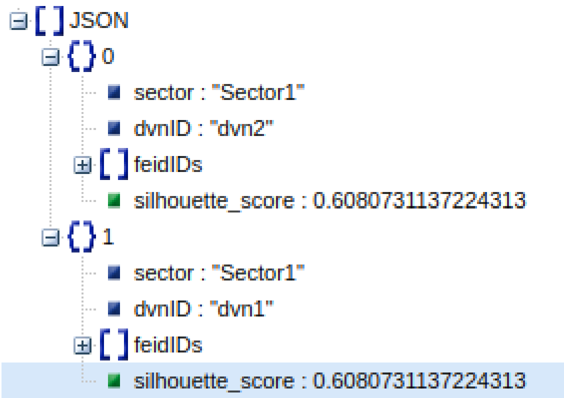

Figure 7. DELTA DVNs result through the initialization of the platform.

As soon as there are historical measurements recorded to the database about consumption, flexibility and DR contribution, the segmentation mechanism (See Section 4.2) is activated and new DVNs formulation is applied. This flow permits aggregators to preserve stability to their portfolio structure while re-sharing the assets among the generated DVNs. Configurations over the involved measurements can be applied from the aggregator, in terms of identifying the structure that meets its demands. The following figures display indicative results from the Segmentation process during the rescheduling phase, when historical data have been recorded from the aggregator. They illustrate the daily schedule of each FEID and their participation in DVN groups in accordance with some specified measurements.

3.3 New DELTA Customer Assignment

In case of a new customer the CDSS is responsible to classify new customers (FEIDs) in one of the existing DVN entities. Lack of historical measurements and information about the general profile and behaviour of customers leads CDSS module to focus on its contractual characteristics: geographical location, consumption capacity, generation capacity, storage capacity, market participation, customer type (implicit, explicit) are some of the features that are taken into account in the primary classification process. Additionally, hard constraints about the belonging sector of each FEID affect the decision of the TDDS module - DVNs consist of FEIDs from the same sectors. In that way, aggregators can preserve stability and balance to DVN entities for a short period of time without disrupting their functionality. Rescheduling process from TDDS is activated as soon as there are historical records in order to make corrective actions.

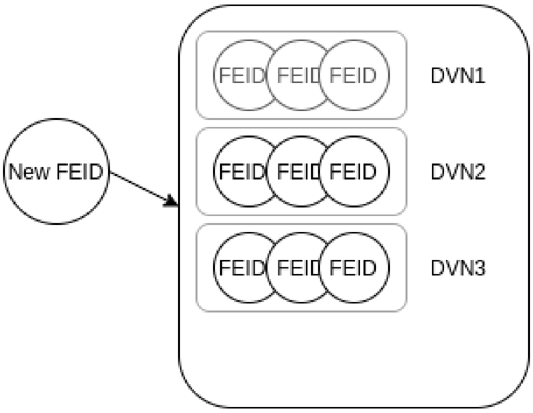

Figure 8: New Customer case that need to be assigned to a DVN.

The second scenario that concerns the extension of Aggregator's energy portfolio, when the initial generated DVNs from the TDDS module have been modified and the portfolio has been rescheduled with regard to TDSS module as it will be described in Section 4. TDDS reformulation in the existing portfolio, modifies the number of generated DVNs, rendering the already designed classification model inadequate to incorporate new customers, as it ignores the existence of the total number of DVNs. As a result, a new classification model needs to be designed, incorporating the knowledge of the generated DVNs. This is achieved through the creation of a dataset that utilizes the labels of the TDDS results and the information from the corresponding contractual data of the FEIDs. Thus, the new designed classification model combines the inference from TDDS with already the existing information from the intrinsic characteristics of each FEID.

4. Temporal Data Dynamic Segmentation (TDDS)

This section presents the second core module of the AEPS, the Temporal Data Dynamic Segmentation (TDDS). The aim of the TDDS is to establish the allocation of the DELTA Virtual Nodes underneath the DELTA Aggregator entity. The resources distribution is based on clustering algorithms performed at the Aggregator level to the entire portfolio regarding certain characteristics. The main objective of this task focuses on creating virtual medium/large clients that can be handled more efficiently and be part of a specific demand response strategy. Temporal adaptation and adjustment of DVNs resources is supported by the CDSS tool as it rearranges the DVNs autonomously based on indicators infused by the Aggregator Architectural Requirements.

Although in principal the TDDS follows the CDSS to design and deliver the DVNs, in the case of a new customer (see Section 3.3) the CDSS is employed on top of the TDDS to assign the new customer within the most appropriate DVN.

The DELTA Aggregator Portfolio Segmentation's functional flow is presented in the following flow chart.

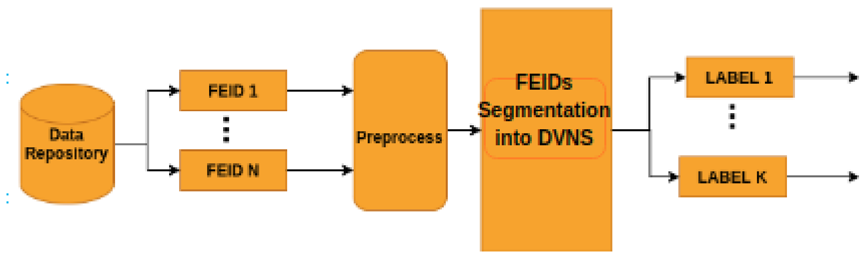

Figure 9: Aggregator's Portfolio Segmentation steps.

4.1 Data Selection

The TDDS tool that re-formulates the DVNs is based on several features, either contractual or real time and historical measurements, some of which are depicted in Figure 10. Contractual data is static information that reflects the capacities and the location of each customer as they are declared in the initial contract. Indicative contractual measurements are: consumption capacity, generation capacity and storage capacity, customer type, market participation. On the other hand, real time and historical measurements that are recorded and examined from TDDS are:

- load consumption,
- energy generation,
- upwards and downwards flexibility,
- FEIDs' reliability that represents the credibility that this asset has shown in previous DR participation, and
- FEIDs' participation that represents the total involvement of each FEID in DR programs.

Data selection and feature extraction process is associated not only with the selected DR strategy and the goals that the Aggregator has set, but also with the profile and the needs of the customers. Therefore, every customers' segment requires to be meticulously examined in order to identify these features that satisfy both sides. Applied feature engineering techniques over the historical measurements will be described in the following section.

Reliability	Participation	FEIDs	Power Consumption (W)	Power Generation (W)	Consumption Capacity (W)	Generation Capacity (W)	Storage Capacity (Wh)
1	34	50	7555.98972	0	781422	168000	168000
0.72	23	15	5599.13142	6900.26666666667	100177.55588	20000	20000
1	76	3	1245.6	703	10000	9570	9570
1	72	40	6451.60854	3632.80666666667	377050.39354	42000	42000

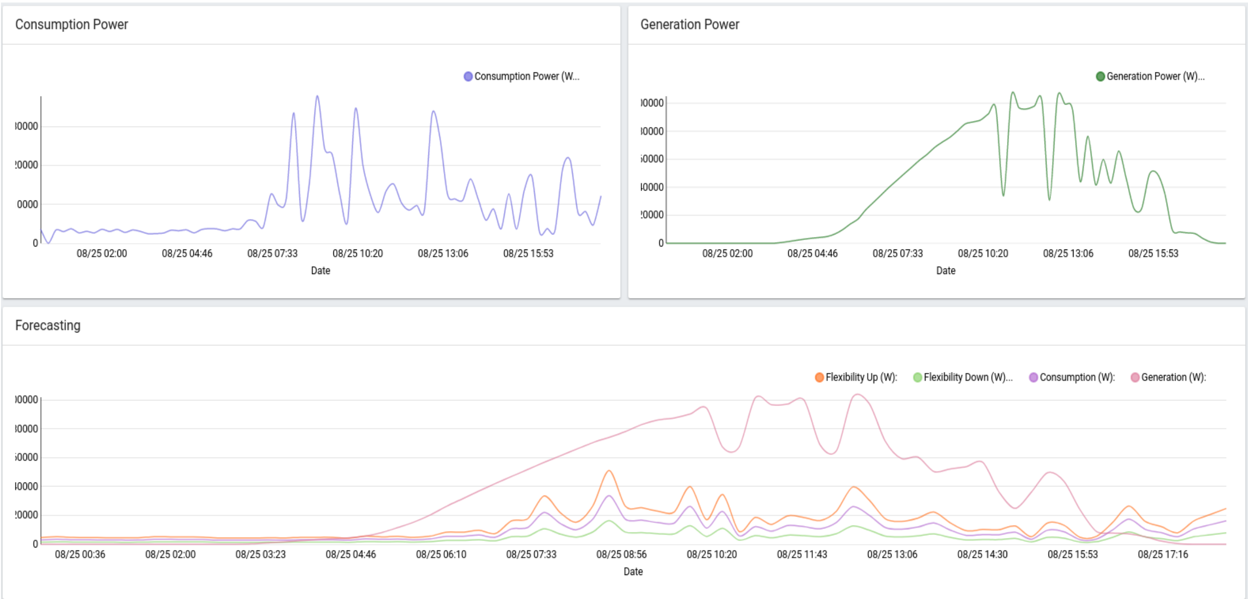

Figure 10: Contractual and real time measurements.

4.1.1 Data Pre-processing

Data manipulation and adaptation of the data structure with regard to the problem is a matter of substantial importance for the design of efficient clustering algorithms. As the structure, the content and amount of data affects the quality of clustering results, it requires a good understanding of the problem. The proposed methodology comprises three steps, as presented below:

- The primary **Data Pre-processing** step of TDDS is feature extraction of statistical metrics from real time and historical measurements of previous months: - consumption, generation, flexibility - Several statistical metrics are calculated for instance: minimum value, maximum value, mean value, standard deviation in order to reduce the dimensions of time series data. In that way, meaningful information from the latest months is preserved, while at the same time the amount of data is tractable for clustering algorithms. It is worth mentioning that outliers are detected and removed from the data focusing on a segmentation model that generalizes well and is not affected by noise.

- The **Fusion** of contractual data with the extracted features from historical measurements for each customer compose the total amount of data.

- **Feature Selection** is the following step of TDDS methodology. This step is not independent of DR strategy and policy that the aggregator/retailer has established. As the strategy and the objectives of the aggregator have been defined, a genetic algorithm implementation runs iteratively in order to identify the optimal combination of features and clustering algorithms that yield higher clustering results towards the goal. According to that methodology, the feature selection process adapts to the objectives and is not a static choice.

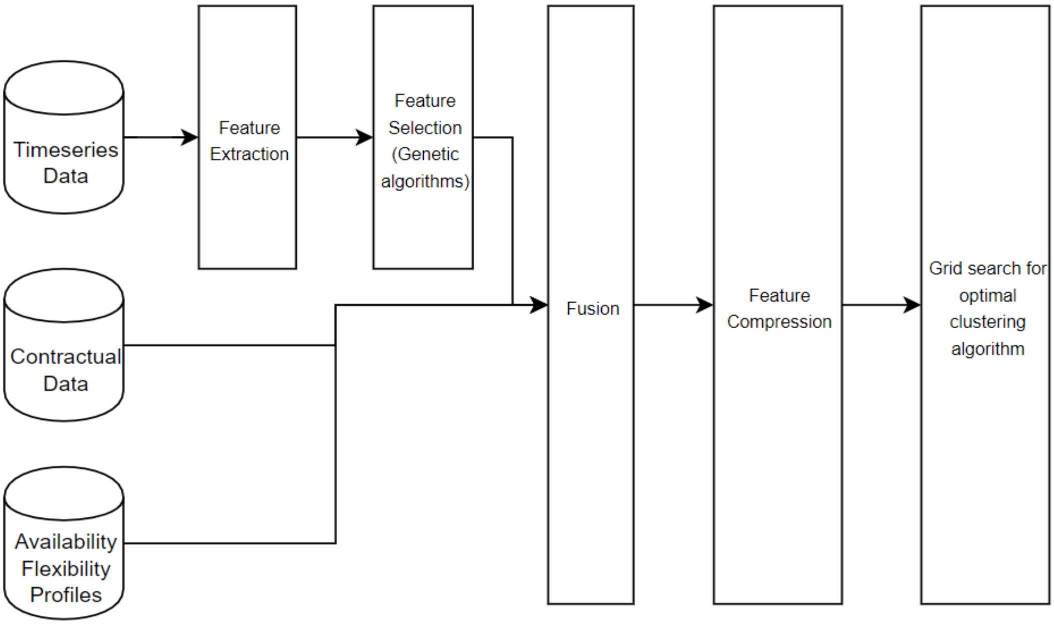

Figure 11: TDDS structural overview.

The following figure displays statistical values from DVNs with regard to consumption measurement during one day in an hourly granularity. Three tables describe the aggregated DVN behaviour of the respective generated DVNs in DELTA platform through maximum, mean and minimum metrics. It is worth mentioning that two of these DVNs are composed of virtual FEIDs. DVN3 consumption behaviour appears to reach its maximum mean value during the early hours in the morning; however, the maximum magnitude of consumption appears during the midday period. Moreover, the minimum value of the corresponding DVN does not provide any further information, as it follows a steady volume near 690 Watts during the whole day. The other two images concerning DVN2 and DVN1, they illustrate values on a larger scale. DVN1 reaches maximum mean value during the morning hours while DVN2 in the afternoon. Both of these involved DVNs contain more than twenty FEIDs in their energy portfolio.

DVN3

DELTA VIRTUAL NODE - STATISTICS

	mean	max	min
0	1074.439844	1618.200000	688.900000
1	1066.657031	1631.500000	688.500000
2	1054.195312	1557.800000	696.100000
3	1053.532031	1534.566667	687.100000
4	1032.710937	1555.800000	677.300000
5	1030.695313	1546.900000	687.200000
6	1034.695312	1950.400000	691.400000
7	1382.395067	2451.300000	690.700000
8	1926.670371	2914.700000	694.200000
9	2033.195774	2989.351471	687.700000
10	2039.628128	2997.287755	661.300000
11	1984.531895	3045.500000	675.800000
12	1937.397796	3012.600000	675.200000
13	1987.905268	2900.900000	666.200000
14	2014.322669	3229.500000	673.700000
15	1771.820875	2746.700000	678.600000
16	1335.700758	2181.100000	663.100000
17	1236.233333	1951.500000	685.300000
18	1163.100000	1869.000000	721.600000
19	1092.560606	1798.500000	683.500000
20	1086.767424	1812.400000	704.500000
21	1076.277652	1633.500000	682.200000
22	1092.586047	1737.034615	692.647273
23	1094.488411	1603.400000	714.898980

DVN2

DELTA VIRTUAL NODE - STATISTICS

	mean	max	min
0	894.91	1637.79	671.14
1	898.61	1495.33	694.17
2	1235.89	5074.96	679.30
3	10928.33	50948.57	1545.37
4	12165.63	48665.18	1522.99
5	10237.54	45706.25	1212.93
6	9088.83	32697.19	1140.07
7	7089.60	40035.45	902.86
8	7593.70	46040.21	808.99
9	4671.50	23160.87	829.64
10	6606.86	33274.20	857.63
11	9216.58	50008.59	967.56
12	7881.67	53748.91	1258.00
13	7755.74	30265.65	934.39
14	15439.13	33592.85	2222.37
15	16377.80	32167.91	7321.76
16	23550.53	39035.82	10267.67
17	13669.72	24575.53	2673.06
18	8965.72	19514.49	2590.83
19	7542.99	20508.48	1405.44
20	2545.81	12269.28	742.38
21	976.11	1646.95	32.92
22	952.99	1511.50	663.70
23	871.97	1444.90	663.56

DVN1

DELTA VIRTUAL NODE - STATISTICS

	mean	max	min
0	3005.85	3617.53	2431.02
1	2858.47	3382.16	2477.37
2	3195.52	3788.67	2608.32
3	5607.93	12006.47	2941.28
4	11412.12	20667.41	4814.53
5	15185.53	30380.01	6538.33
6	22131.40	41129.44	7258.08
7	10950.56	35549.68	3959.53
8	8555.32	19176.98	3706.79
9	11448.66	31180.42	4015.90
10	9428.79	20081.40	3285.01
11	9120.90	22103.79	3592.39
12	7301.18	24445.11	3031.74
13	9579.97	26520.97	2895.91
14	13346.44	24714.58	3445.13
15	10741.64	20353.63	4473.48
16	15593.25	34696.84	5637.23
17	17285.46	32160.04	7094.03
18	13389.48	20380.64	5518.10
19	11357.06	24289.04	5868.03
20	4488.15	6118.44	3058.86
21	3208.89	3589.94	2881.20
22	3228.20	3882.67	2591.46
23	3207.90	3872.91	2521.91

Figure 12. DVN statistical hourly measurements regarding min, maximum and mean value.

4.2 Spectral Clustering over Dynamic Warping Time Distance

Dynamic time warping (DTW) is an algorithm that estimates the distance between temporal sequences that may vary in speed. Speech recognition, Vision and graphics are some of the fields that this method has been applied to. In terms of the energy aggregator's ecosystem, there are numerous time series measurements that reflect the energy profile of customers. The basic idea of the proposed method includes the estimation DTW distance among some high importance time series (forecasted consumption, generation) and the application of Spectral Clustering (SC) algorithm in order to generate the final segments.

The basic functionality of the SC algorithm is based on the Eigen decomposition of the Laplacian matrix of our data. The Laplacian matrix is calculated through the equation:

$$L = D - A$$

where D is the Degree matrix which denotes the connections among the nodes and A is the affinity matrix of our data. The affinity matrix is a symmetric matrix that expresses the similarity of each FEID's measurement with the rest of the assets.

The first step of the incorporation of DTW method to our algorithm is to calculate the dissimilarity matrix of these measurements. The following step includes the transformation of these distances to similarities through the Gaussian Kernel function. In particular, we utilized a specific form of the kernel function that adjusts parzen window parameter for optimal results as it is mentioned in [37]. The equation of the Gaussian Kernel function is:

$$\hat{\sigma}_i^2 = \frac{\max d(i,j)^2 - \min d(i,j)^2}{2 \ln \dfrac{\max d(i,j)^2}{\min d(i,j)^2}}$$

Finally, through the decomposition of the Laplacian matrix, we extract the eigenvalues of the selected matrix, exploiting the gap between the eigenvalues that facilitates to estimate the appropriate number of clusters.

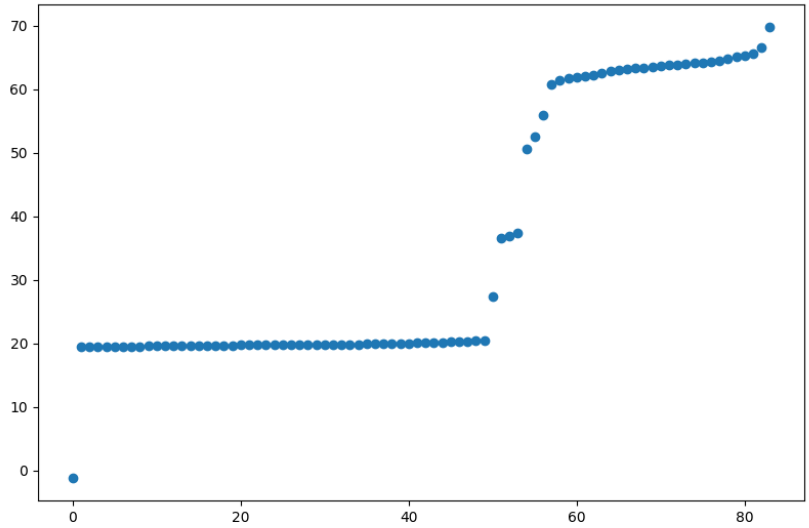

Figure 13. Eigen values of Laplacian matrix

4.3 Dynamic DVN allocation

The TDDS configures and adapts its schema according to real time conditions and modifications to the aggregator's energy resources. Potential causes of rescheduling can be:

- Behavioural alterations of FEIDs
- Energy assets (FEIDs) addition or removal
- Special temporal constraints
- Business plan changes

The former factors can trigger rescheduling processes, formulating a new individual schema consisting of several DVNs, with some energy assets (FEIDs) assigned to each one. Selection of clustering algorithms and parameterization are two factors that should be taken into consideration, as presented in the following subsections.

4.3.1 Segmentation Parameterization

The identification of the appropriate number clusters and the individual parameters of each clustering algorithm that maximize its efficiency is applied through a grid search over several clustering algorithms. The evaluation metric that is deployed to validate the segmentation process is Silhouette Score. Silhouette Score as a metric examines the coherence of points within clusters and the simultaneous separation of this cluster from neighbour segments. As far as the number of clusters concerned, clustering algorithms can be divided in two categories: the ones that detect the number of clusters autonomously and the ones that the user needs to define the appropriate number. Silhouette score is an indicative metric that facilitates the detection of this number regarding the second category of algorithms.

4.3.2 Clustering Methods

In the literature review, as it is presented in section 2.1, there are several clustering algorithms that are utilized to distinguish the groups of energy customers in terms of their energy behaviour. Within DELTA, baseline algorithms (e.g. K-means) along with more "adaptive" / improved clustering versions have been benchmarked for integration in the final AEPS, results and experiments conducted with synthetic data (as explained in the following section) in Chapter 5, where the final selection of the most appropriate method is outlined in terms of:

- their scalability as the number of users increases, and
- their feasibility to generate distinguishable groups of customers.

4.4 Dataset

The study and the results of this report are based on indicative measurements of experiments with 1000 virtual FEIDs that include historical data of 3 months. These virtual FEIDs have been generated driven by the functionality of real houses (profiles taken by actual datasets and third party tools for load profiles), in accordance with all DELTA properties. More detailed information about vFEID engine is included in D3.4. Some key characteristics of the generated vFEIDs (example provided in Table 4) that reflect the structural characteristics of our customers are:

- the 90% of our customers are consumers and only 10% are prosumers
- 80% of our customers are small/medium customers and 20% are tertiary customers
- the mean reliability of our assets is near 0.85

Table 4. Instance of Quantitative characteristics of the generated DVNs measurements

Power Consumption (W)	Power Generation (W)	Consumption Capacity (W)	Generation Capacity (W)	Storage Capacity (kWh)
3902.963	12830	790000	144000	20000
312.480	5712.2	38000	9000	5000
3122.8	4608	10000	9570	4500
7045.039	17136.6	269000	47000	9000

5. Results

5.1 Spectral Clustering over Statistical Features

Identifying customers' segments from individual measurements (consumption, generation) can aid aggregators design incentive/price based DR strategies oriented to specific objectives. An analysis of clustering algorithms for each time series measurement individually depicts that there is no clear separation to all cases as represented below. The Feature engineering step produces 40 statistical features like mean value, standard deviation, variation in an endeavour to reduce the dimensionality of time-series measurements. Isomap as a dimensionality reduction technique has been applied to data in order to be visualized in two dimensions.

Although the tests were applied over several clustering algorithms, the Spectral Clustering method showed remarkable results. Five clusters of FEIDs were identified and examined for each measurement independently. As it is observed in Figure 14, clustering in terms of consumption measurement identifies discernible groups of FEIDs, while the Silhouette score of this segmentation is near 0.76. Accordingly, regarding the downwards flexibility measurement, our algorithm achieved the identification of five groups of customers, while the Silhouette Score is close to clustering related to Consumption measurement with 0.75. On the other hand, generation and upwards flexibility measurements could not provide meaningful and discrete and separation of customers behaviour as it is reflected from visualizations and the Silhouette Score of 0.32 and 0.51 respectively. It is worth mentioning that in the case of generation measurements there are only 200 of 1000 FEIDs that are considered as prosumers, therefore there is a reduced number of manageable assets to be grouped.

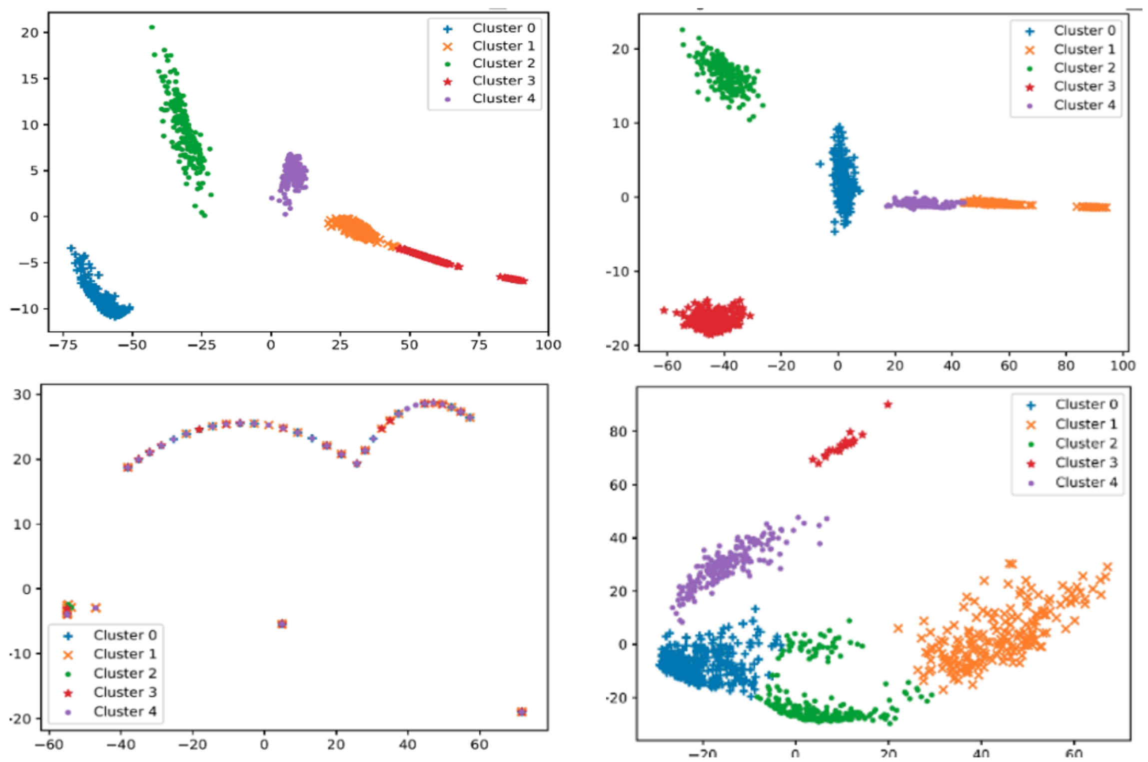

Figure 14: Consumption, Upwards Flexibility, Generation and Downwards Flexibility 2D representation through Isomap.

A general overview of the individual scores for each measurement is presented in the following table while Spectral Clustering and DBSCAN algorithms' results are compared.

Table 5. Silhouette score of spectral clustering analysis on different measurements.

Measurements	Spectral Clustering	DBSCAN
Consumption	0.76	0.61
Generation	0.32	0.30
Upwards Flexibility	0.51	0.52
Downwards Flexibility	0.75	0.71

Regarding the analysis of execution time and the feature extraction procedure, this study examines the total running time and compares different clustering algorithms for one thousand FEIDs. Spectral Clustering algorithm is proved to be the most time-consuming approach, whereas mean-shift requires the minimum duration compared to all algorithms. Generally, these algorithms seem to achieve scalability, as there is a linear relationship between the number of generated features and the execution time that is presented in Figure 6. Specifically, Figure 6 displays a diagram that correlates the execution time of each algorithm with the number of generated features. These metrics are estimated as the mean value of 15 sequential executions for each method.

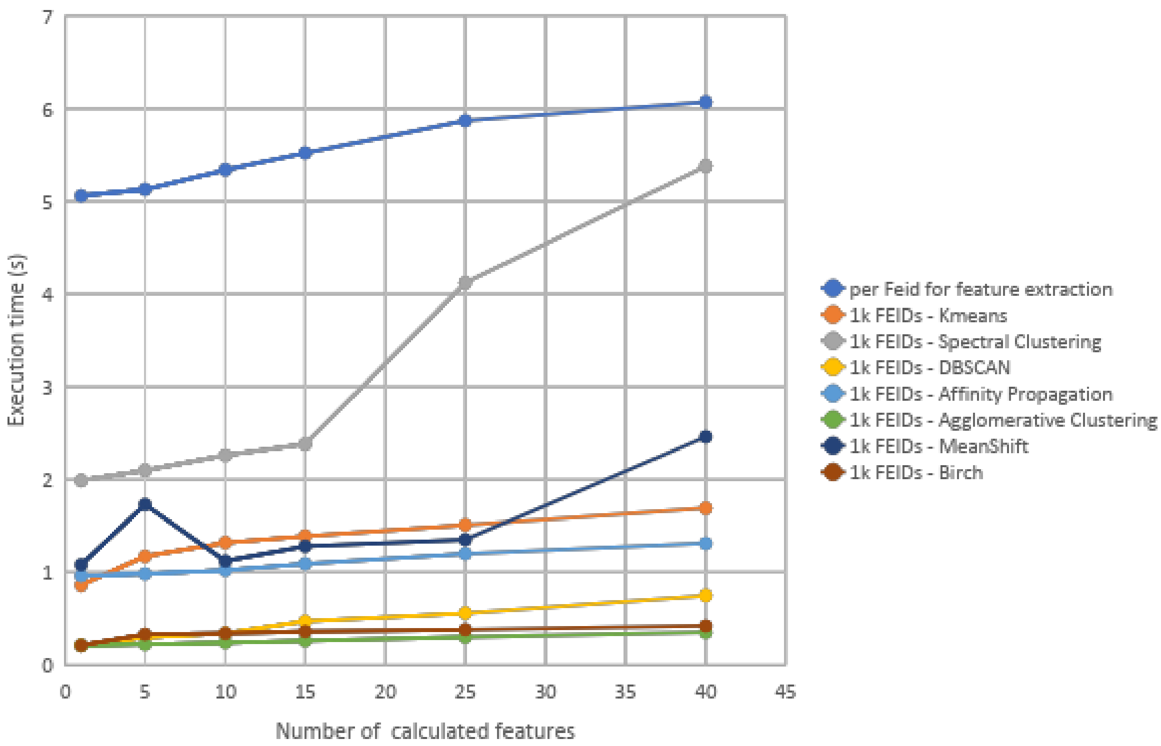

Figure 15: Execution time per Number of Calculated features for several clustering algorithms.

Furthermore, there are indicative results from statistical measurements that describe a general overview of each DVN as it has been deployed after the application of TDDS. Maximum value, minimum value, mean values and standard deviation of each established DVN during the day is displayed in figure 7, 8, 9 and 10. The temporal behaviour of the members of its group seems to have common characteristics and at the same time differentiates from other groups, validating the efficiency of our approach.

In the following figure, there are displays of statistical metrics from the generated segments of FEIDs in terms of a specified measurement. Regarding the consumption measurement, it is observed that the majority of FEIDs in the DVN1 raise their values in the middle of the day with mean value near 1500 Watts and maximum value 7500Watts, while in DVN2 and DVN3 the involved FEIDs do not surpass the 6000 Watts as maximum value and mean value near 750 Watts. The peak of the consumption measurement in all the DVNs is observed in the first half of the day.

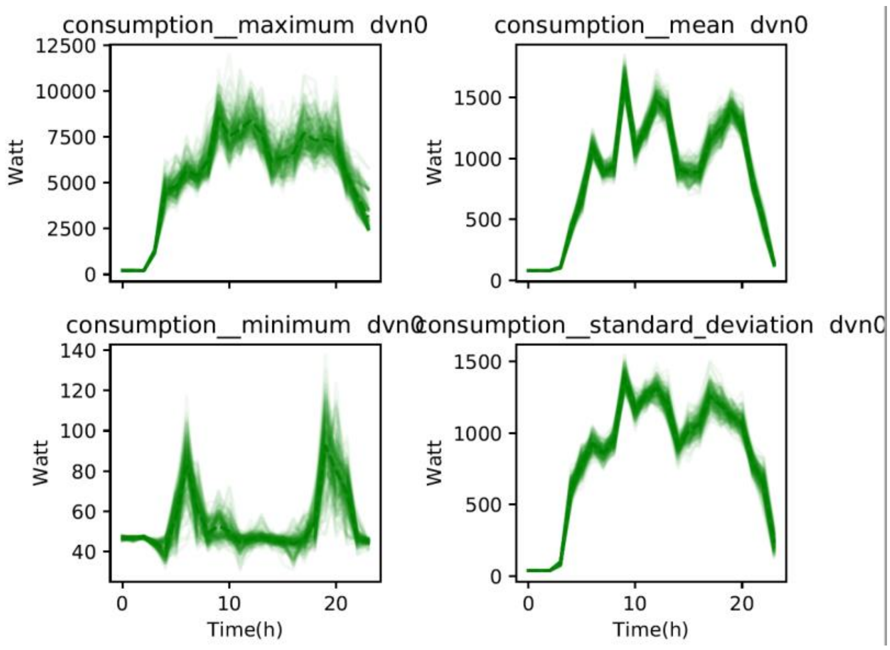

Figure 16. FEIDs' profiles to DVN0 through consumption measurement.

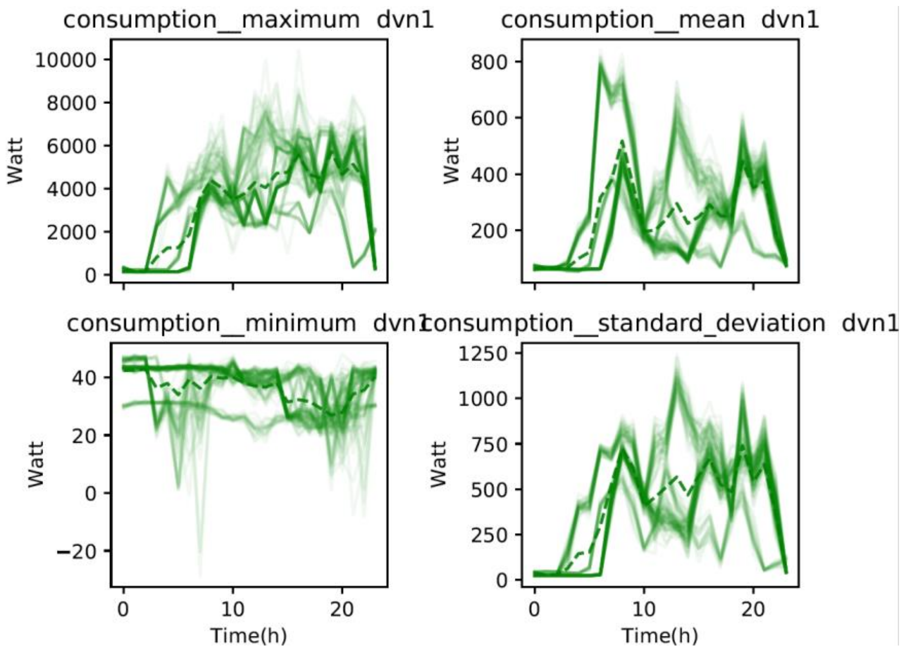

Figure 17. FEIDs' profiles to DVN1 through consumption measurement.

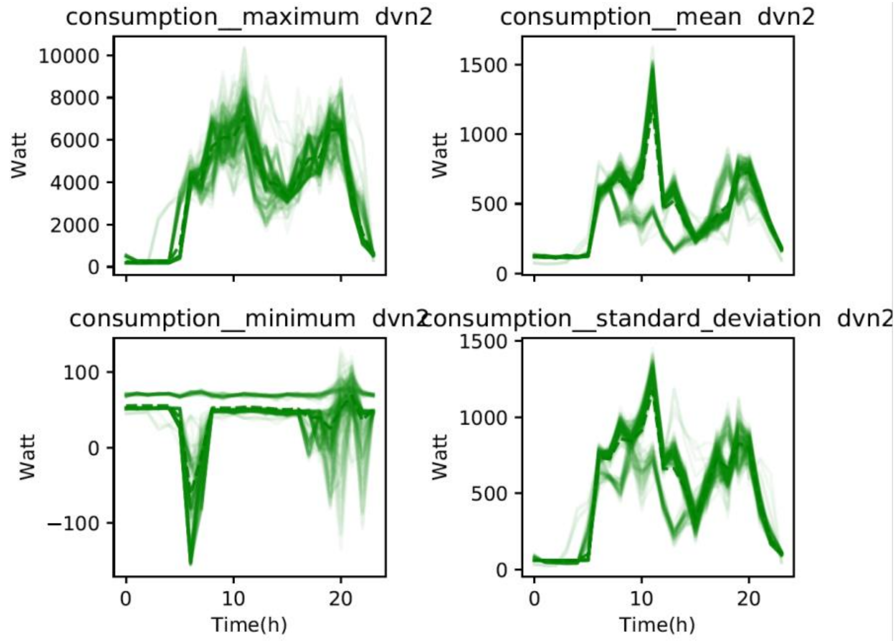

Figure 18. FEIDs' profiles to DVN2 through consumption measurement.

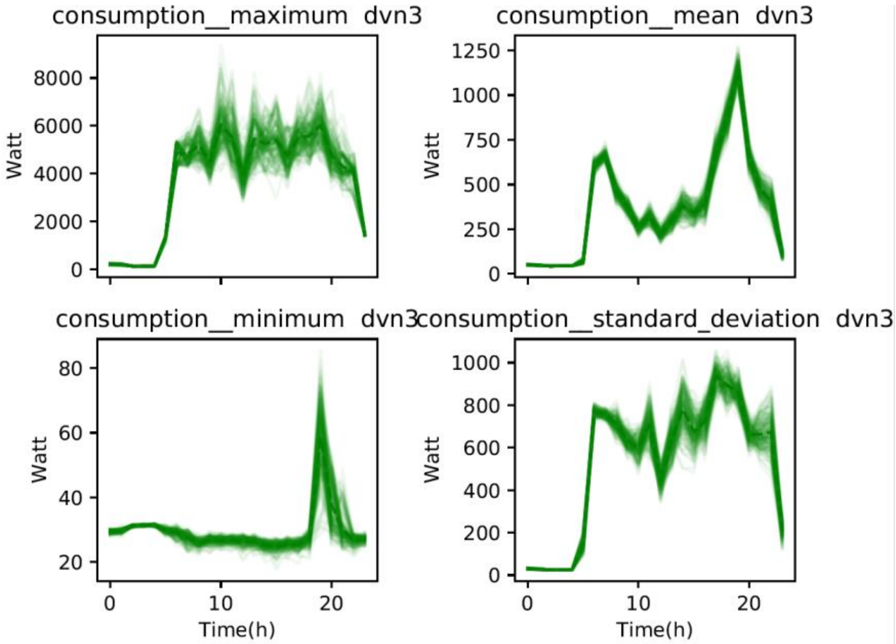

Figure 19. FEIDs' profiles to DVN3 through consumption measurement.

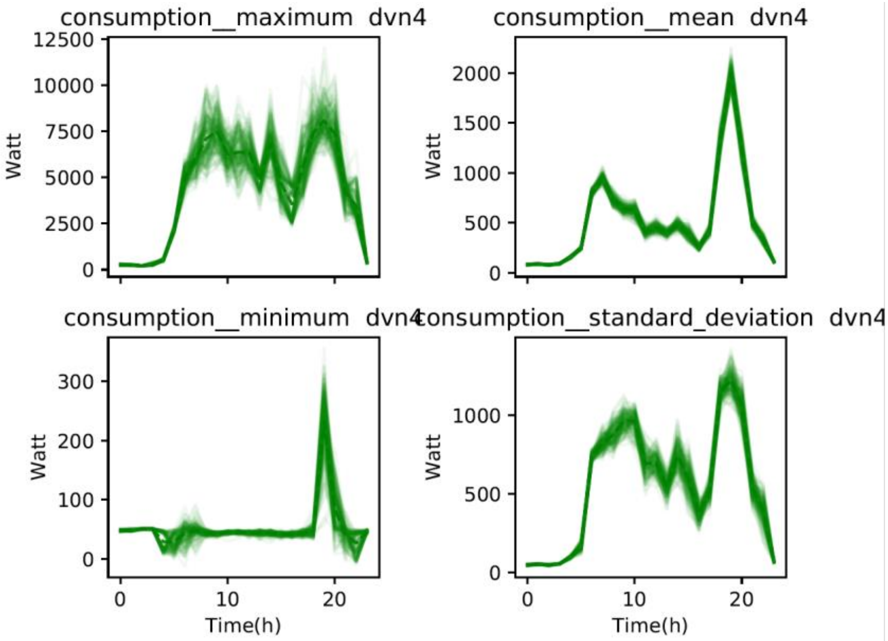

Figure 20. FEIDs' profiles to DVN4 through consumption measurement.

As depicted in the following figures, there are similar displays of statistical metrics from the generated segments of FEIDs in terms of the downwards flexibility measurement. DVN0 seems to reach the highest magnitude of downwards flexibility near 10000 Watts, while DVN3 and DVN4 reach their maximum mean values in the last 4 hours of the day. On the other side, DVN1, mean flexibility measurement surges its value abruptly in the morning and the afternoon periods. Regarding the DVN2 it reaches.

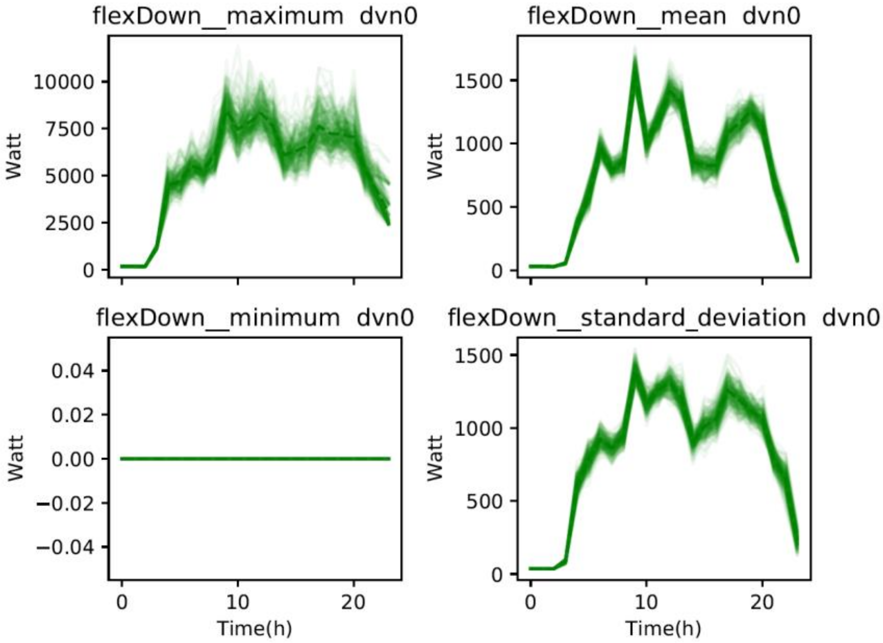

Figure 21. FEIDs' profiles to DVN0 through downwards flexibility measurement.

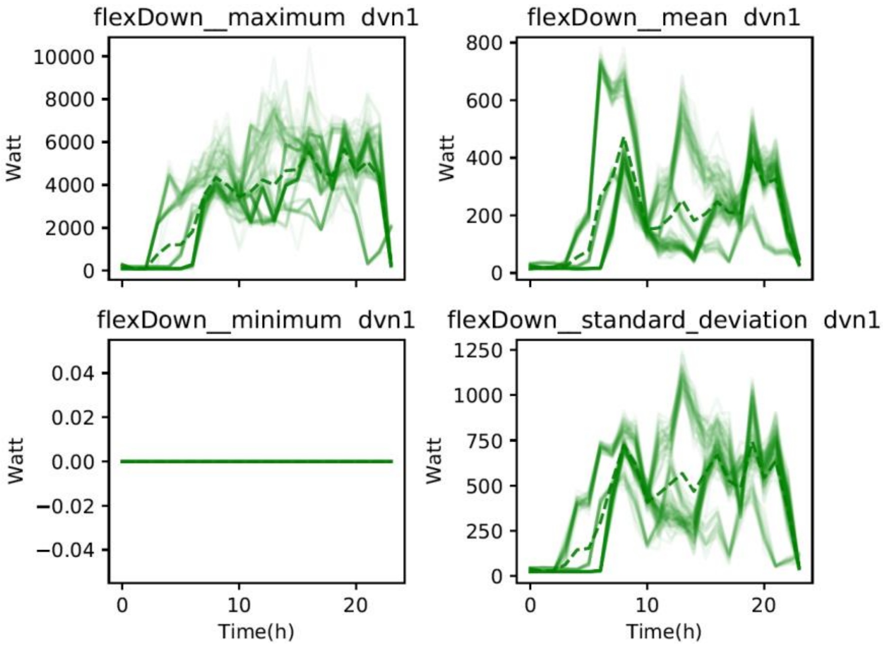

Figure 22. FEIDs' profiles to DVN2 through downwards flexibility measurement.

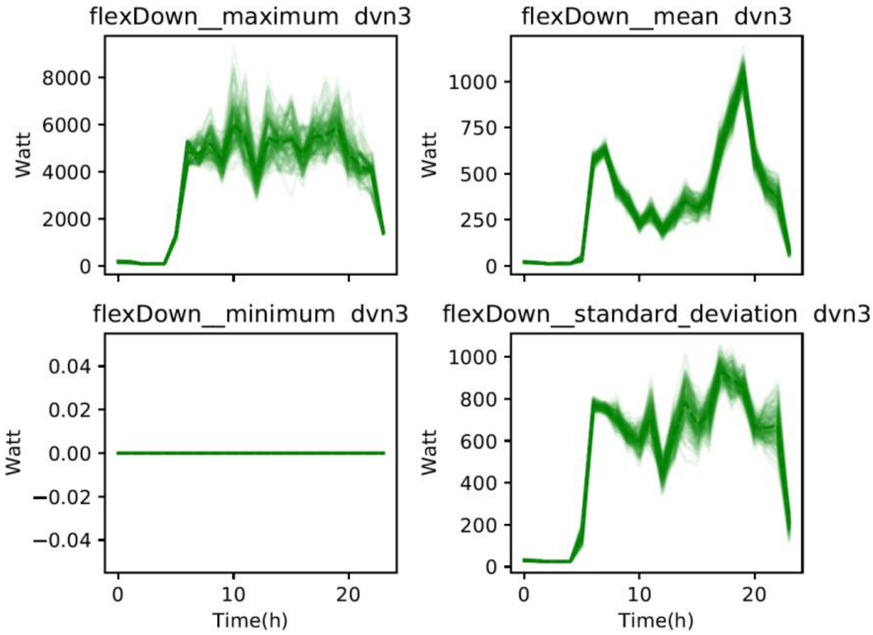

Figure 23. FEIDs' profiles to DVN4 through Downwards Flexibility measurement.

Accordingly, there are similar displays of statistical metrics from the generated segments of FEIDs in terms of the Upwards flexibility measurement. Maximum values of DVN1 tend to increase steadily its values during the day, reaching their peak of 6000 Watts in the afternoon, while DVN2 has two intensive periods of high upwards flexibility values during the morning and the afternoon. As far as DVN3 concerned, the biggest proportion of FEIDs raise their flexibility values in the morning hours and then preserve a steady volume of flexibility until the latest hours of the day.

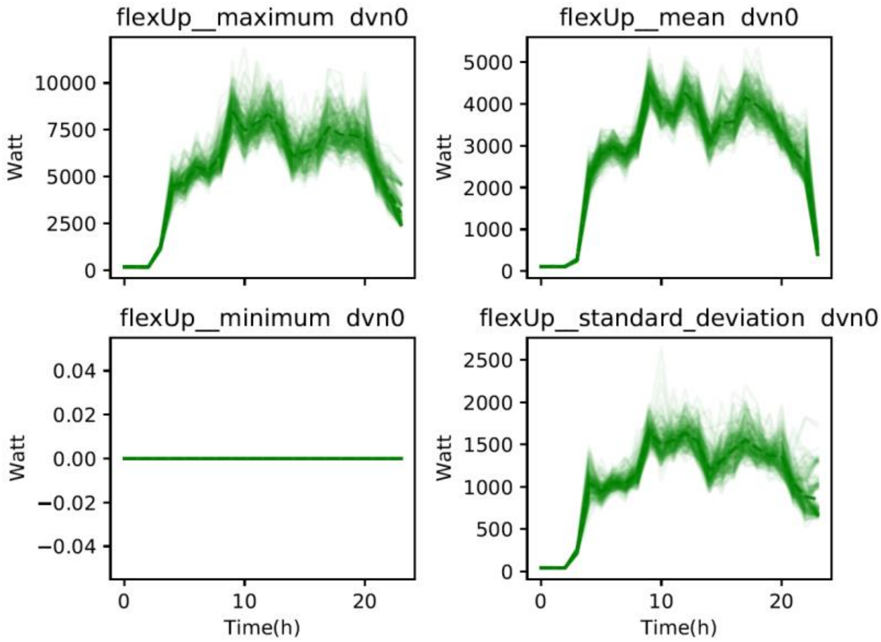

Figure 24. FEIDs' profiles to DVN0 through upwards flexibility measurement.

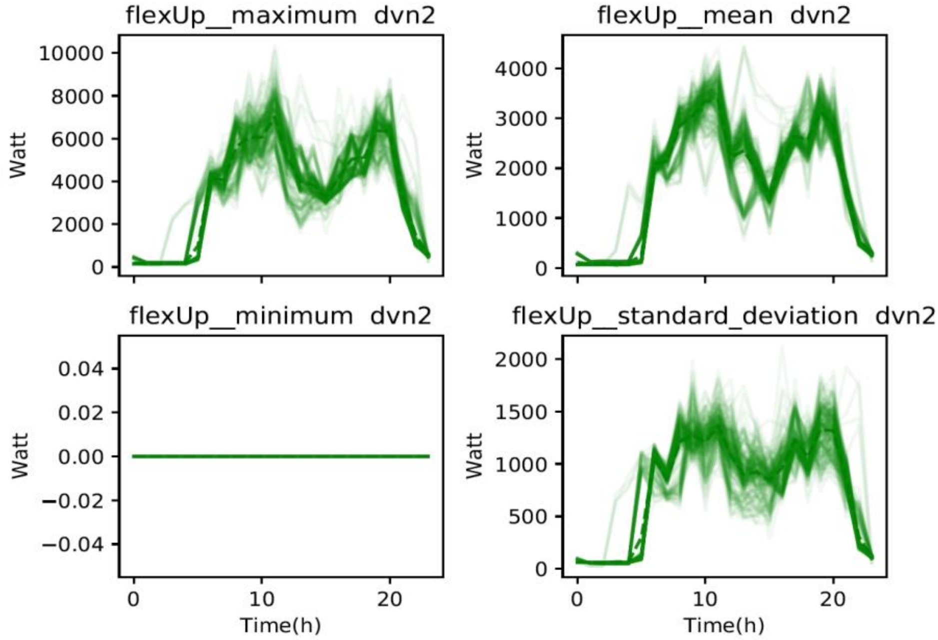

Figure 25. FEIDs' profiles to DVN2 through upwards flexibility measurement.

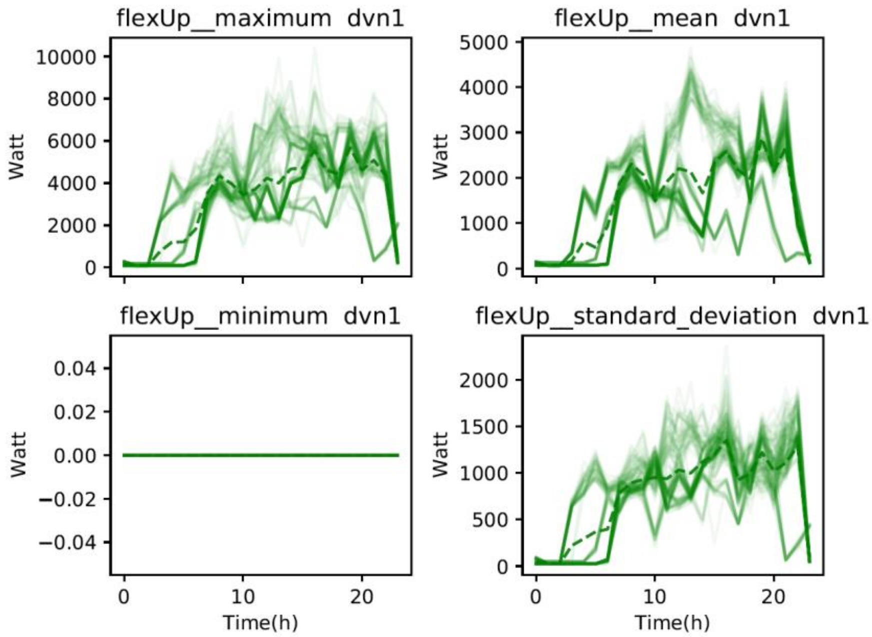

Figure 26. FEIDs' profiles to DVN3 through upwards flexibility measurement.

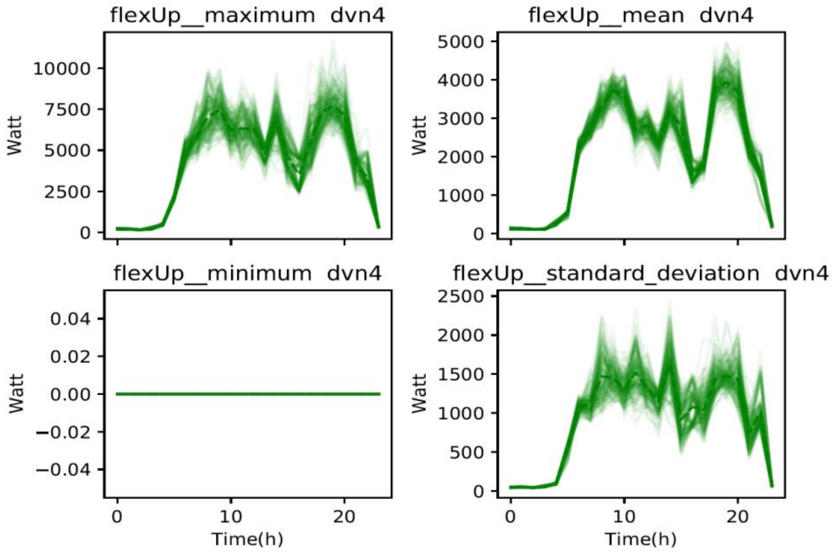

Figure 27. FEIDs' profiles to DVN4 through Upwards Flexibility measurement.

Another graphical representation of the temporal scheduling of aggregator's portfolio is illustrated in Figure 24 and Figure 25 that denotes the differentiation of individual clusters with regard to a specified measurement. Figure 24 highlights the correlation between consumption and generation in an indicative scatter plot, while Figure 25 examines the relationship between consumption and upwards flexibility. Some of the observations that have been recorded are: the fact that customers with higher loads of consumption and generation during the day are located in DVN1, while consumption in DVN0 is increased solely at 20:00. Subsequently, DVN0 contains FEIDs that are small-medium customers, whereas DVN1 has customers of larger scale.

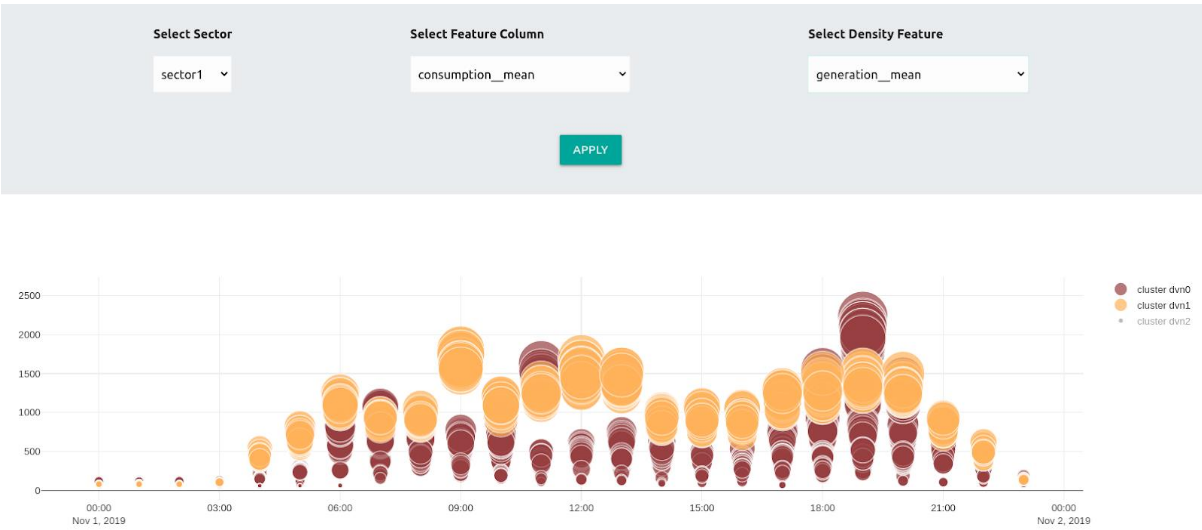

Figure 28. DELTA DVNs Segmentation schedule through Consumption and Generation measurements.

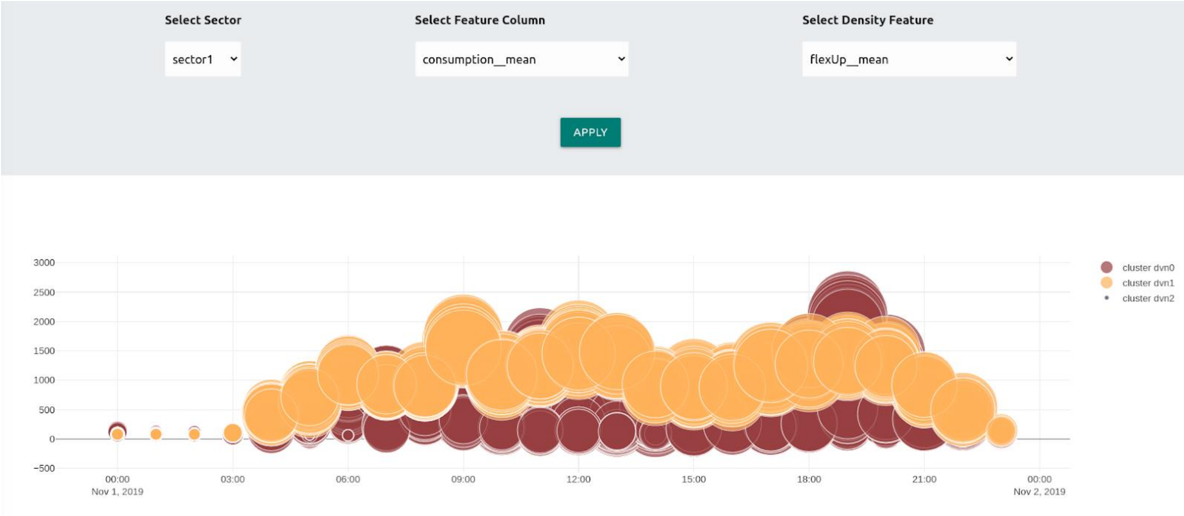

Figure 29. DELTA DVNs Segmentation schedule through Consumption and Flexibility measurements.

Some of the findings during the implementation of this engine are expected to be further elaborated and exploited through the integrated DELTA framework and during the pilot deployment.

5.2 Spectral Clustering over DWT

The results regarding Spectral Clustering over DWT are divided in two temporal continuous periods. Segmentation process applied for each time period was examined individually. In that way, it is possible to study the effects of measurement transformation regarding the formulation of the DVNs schema.

In order to discover the optimal number of clusters that segmentates the energy profiles with the most effective way, an estimation of the Silhouette Score and the Davies Bouldin Index score [38] was applied. In the following diagram and the corresponding table, there is a display of these two metrics in correlation with the number of clusters. It is discernible that the optimal point is the 5 clusters that maximize the Silhouette Score and minimize Davies Bouldin Index.

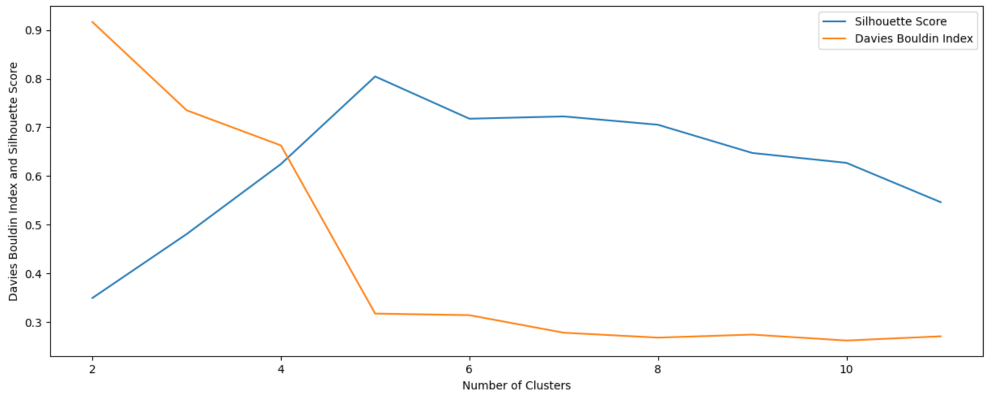

Figure 30. Silhouette Score and Davies Bouldin Index Diagrams

Table 6. Silhouette Score and Davies Bouldin Index Per number of Clusters

Number of Clusters	Silhouette Score	Davies Bouldin Index
2	0.349422	0.91
3	0.48076	0.73
4	0.62461	0.6628
5	**0.80452**	**0.3173**
6	0.71771	0.3141
7	0.72245	0.2780
8	0.7053	0.26802
9	0.6473	0.2742
10	0.62704	0.26203
11	0.546	0.270

The following figures display a representation of the Forecasted Energy Consumption of each FEID in a specific cluster DVN with 15 minutes sampling. The following DVNs that are illustrated for both time periods are a subset of the total number of generated DVNs as the optimal number of clusters was five clusters for both periods.

All of the FEIDs measurements appear to oscillate with high frequency. This fact denotes that the examined period is a peak period with active occupants in the majority of the houses. The most active DVN is DVN3, while the most FEIDs are located in DVN2. DVN4 appears to be inactive, until the latest part of the period that some FEIDs surge their loads.

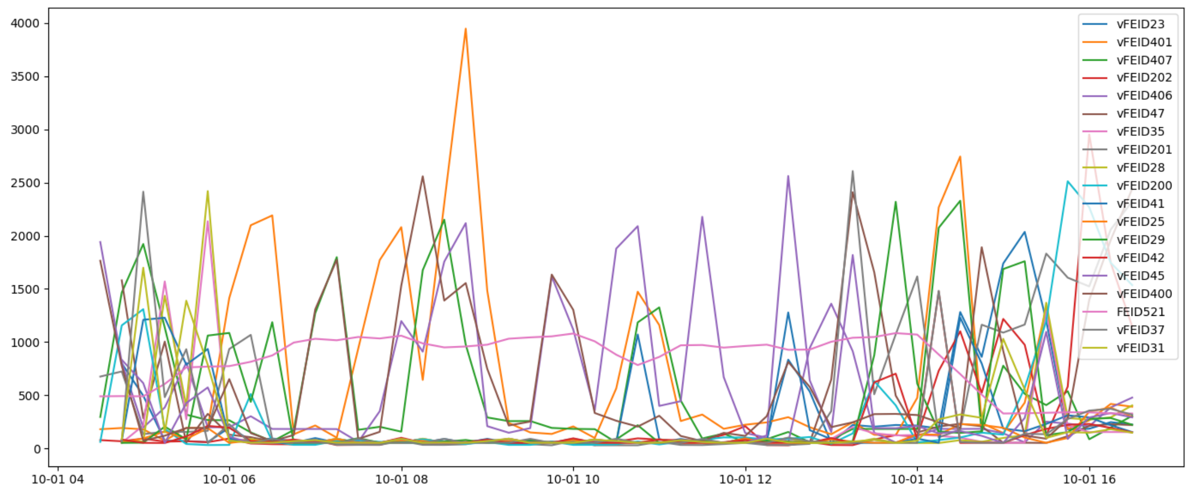

Figure 31. DVN1 - Cluster, Per FEID Forecasted Consumption in time period1

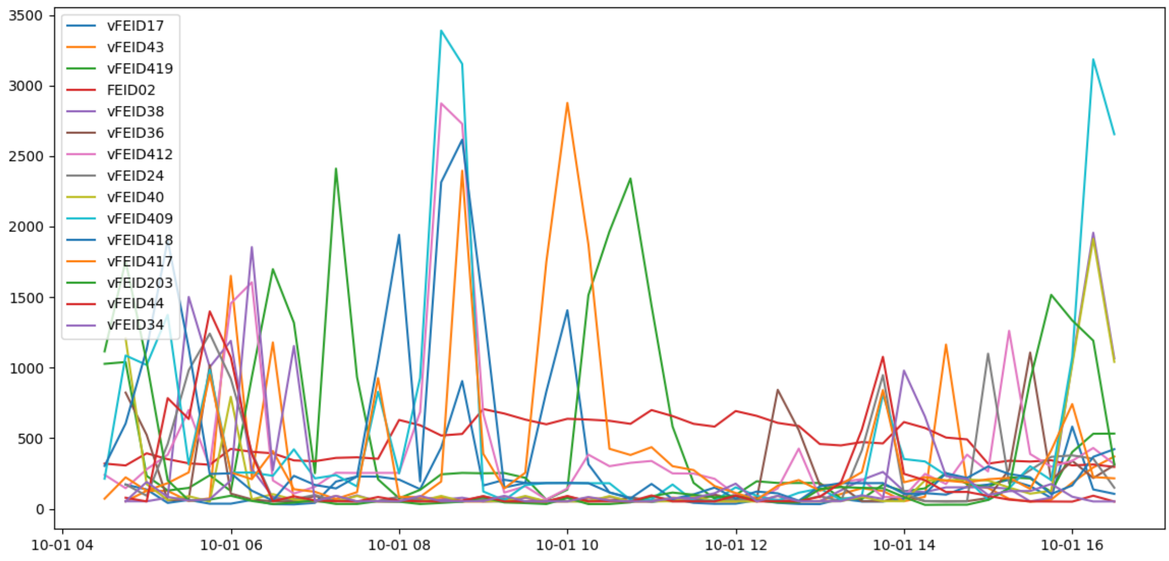

Figure 32. DVN2 - Cluster, Per FEID Forecasted Consumption in time period1.

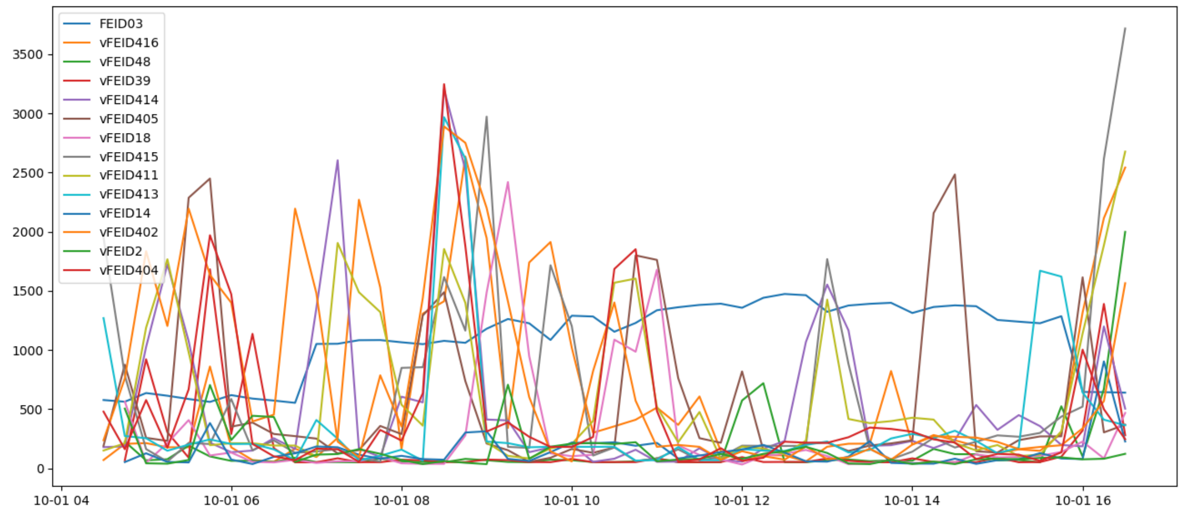

Figure 33. DVN3 - Cluster, Per FEID Forecasted Consumption in time period1

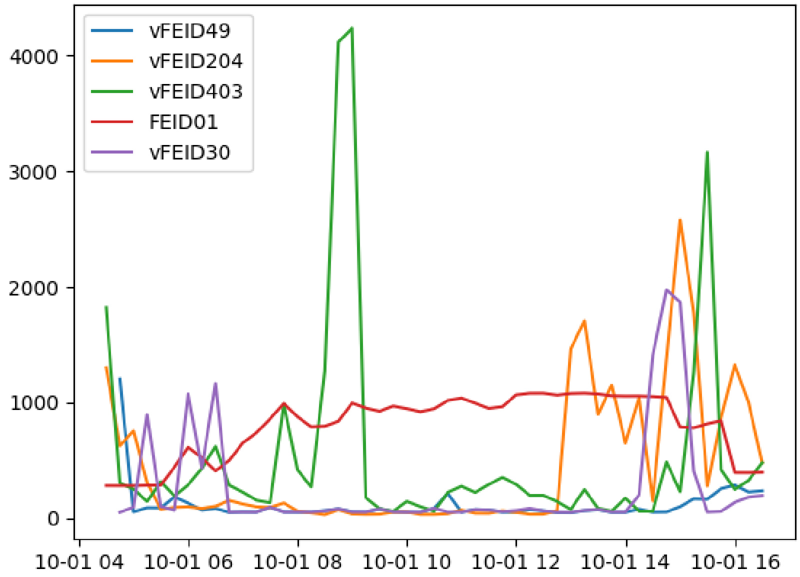

Figure 34. DVN4 - Cluster, Per FEID Forecasted Consumption in time period1

In terms of the second time period, it is discernible that all the loads decline abruptly in the first part. This fact denotes that during this period many customers reduce their loads and the occupancy level of households is low. DVN1 and DVN3 seem to have identical consumption behaviour but in different magnitudes, while DVN2 and DVN4 have a surge that lasts for a short time period. Finally, the FEIDs are shared among the DVNs in a balanced way.

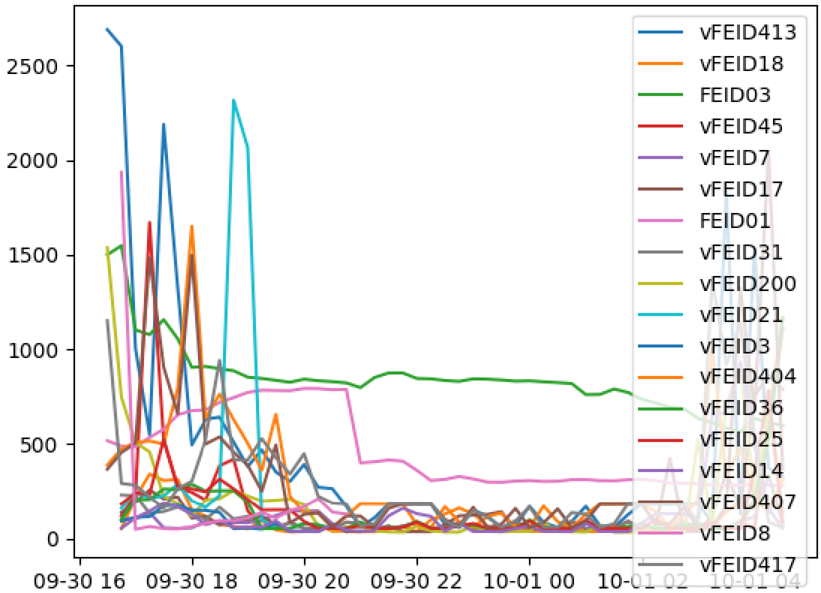

Figure 35. DVN1 – Cluster, Per FEID Forecasted Consumption in time period2.

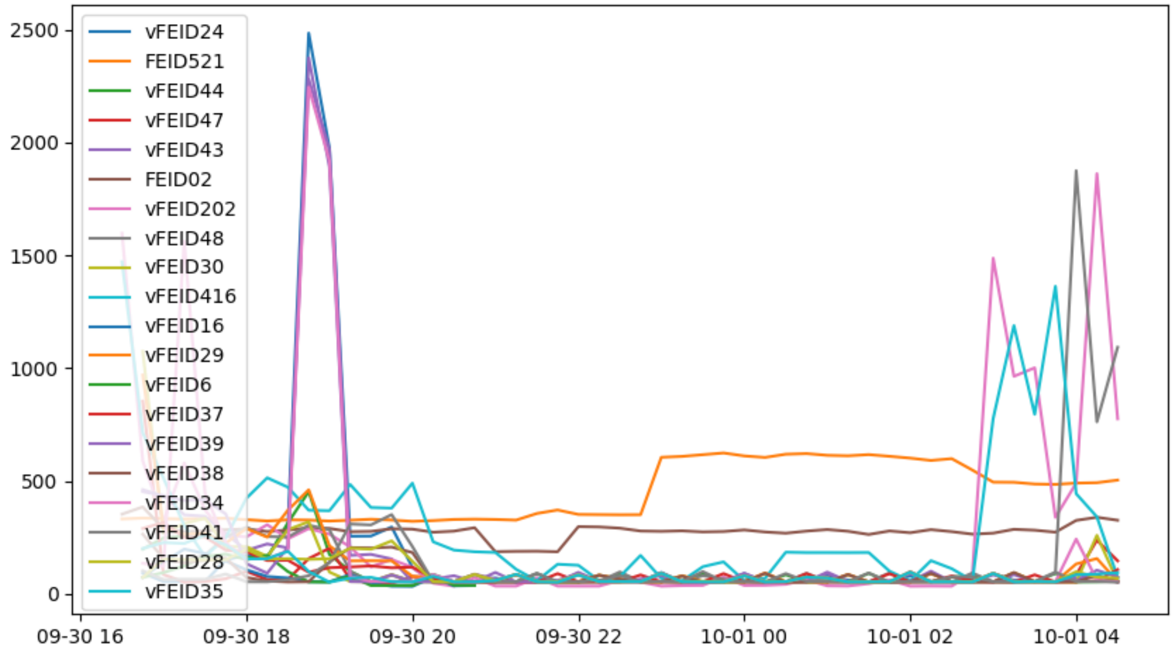

Figure 36. DVN2 – Cluster, Per FEID Forecasted Consumption in time period2.

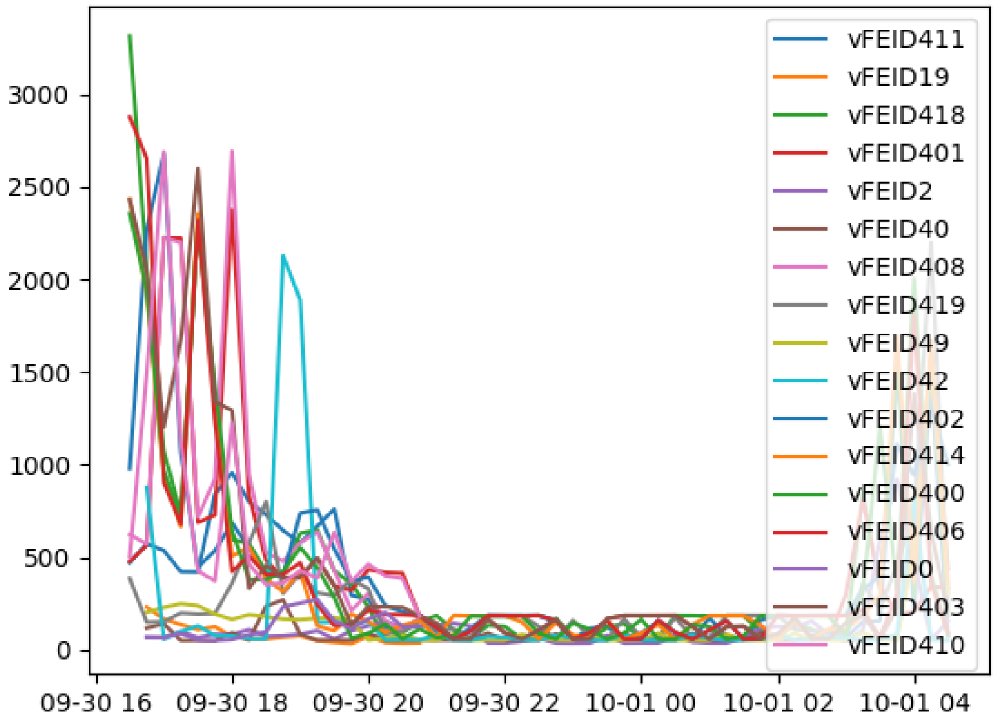

Figure 37. DVN3 – Cluster, Per FEID Forecasted Consumption in time period2.

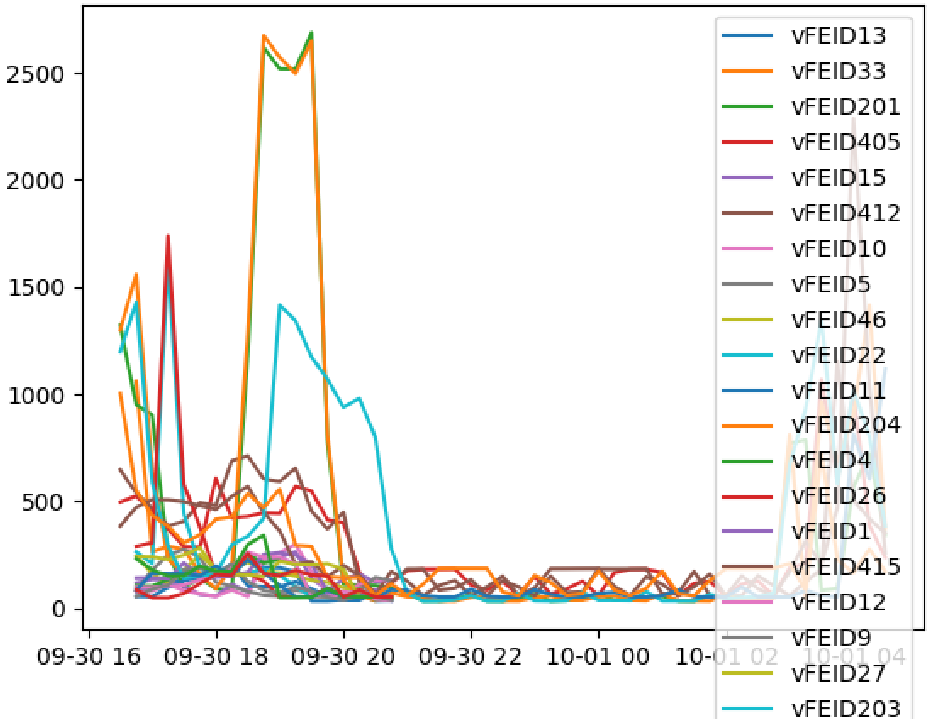

Figure 38. DVN4 – Cluster, Per FEID Forecasted Consumption in time period2.

6. Conclusions

In conclusion, AEPS is a demanding task that can solve several problems for aggregators in the modern distributed energy systems that require a small and medium customer level analysis and management of their energy assets. This report described a concrete methodology to pre-process, analyse and segment the spatio-temporal behaviour of customers in accordance with the properties of DELTA project and additionally a way to incorporate new customers in an already grouped energy system. Furthermore, this task examined several clustering algorithms with regard to scalability and efficiency towards identifying meaningful patterns. Spectral Clustering algorithm showed remarkable performance in terms of clustering efficiency especially with the incorporation of DWT distance; however, it appeared to have some weaknesses related with the execution time compared to other methods. Despite the lack of existing applied architectures similar to DELTA that reduce the possible test scenarios, TDDS and CDSS are data oriented approaches that provide flexibility to aggregators to adjust the methodology according to the adopted Demand Response business plan. Finally, applied feature engineering techniques for the extraction of statistical measurements is a topic that needs to be further explored in terms of scalability.

The tool developed through T4.2 and reported here has been integrated within the DELTA Aggregator’s Decision Support System towards enabling the addition of small and medium customers to the DR programs management. By creating larger virtual entities that can be handled dynamically it is possible to address small bits of untapped flexibility potential from numerous small and medium stakeholders, hence delivering larger sums that are prerequisites to the current DR markets. Real-life evaluation of the AEPS is expected through the DELTA pilots deployment and evaluation activities within WP7.

References

[1]. Q. Gong, Y. Jin, Y. Han, C. Sang, R. Wang and B. Zhou, "Abnormal Electricity Customer Clustering Method Based on Electricity Big Data," 2019 IEEE Sustainable Power and Energy Conference (iSPEC), Beijing, China, 2019, pp. 2221-2226, doi: 10.1109/iSPEC48194.2019.8974928.

[2]. Rajabi, A., Eskandari, M., Ghadi, M., Li, L., Zhang, J. and Siano, P., 2020. A comparative study of clustering techniques for electrical load pattern segmentation. Renewable and Sustainable Energy Reviews, 120, p.109628.

[3]. G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping," Energy, vol. 42, pp. 68-80, 2012.

[4]. G. Chicco, R. Napoli, and F. Piglione, "Application of clustering algorithms and self organising maps to classify electricity customers," in Power Tech Conference Proceedings, 2003 IEEE Bologna, 2003, p. 7 pp. Vol. 1.

[5]. C. C. Aggarwal and C. K. Reddy, Data clustering: algorithms and applications: Chapman and Hall/CRC, 2013.

[6]. M. Crow, "Clustering-based methodology for optimal residential time of use design structure," in North American Power Symposium (NAPS), 2014, 2014, pp. 1-6.

[7]. J. D. Rhodes, W. J. Cole, C. R. Upshaw, T. F. Edgar, and M. E. Webber, "Clustering analysis of residential electricity demand profiles," Applied Energy, vol. 135, pp. 461-471, 2014.

[8]. F. McLoughlin, A. Duffy, and M. Conlon, "A clustering approach to domestic electricity load profile characterisation using smart metering data," Applied energy, vol. 141, pp. 190-199, 2015.

[9]. J. L. Viegas, S. M. Vieira, R. Melício, V. Mendes, and J. M. Sousa, "Classification of new electricity customers based on surveys and smart metering data," Energy, vol. 107, pp. 804-817, 2016.

[10]. Y.-I. Kim, J.-M. Ko, and S.-H. Choi, "Methods for generating TLPs (typical load profiles) for smart grid-based energy programs," in Computational Intelligence Applications In Smart Grid (CIASG), 2011 IEEE Symposium on, 2011, pp. 1-6.

[11]. K. Zhou, S. Yang, and Z. Shao, "Household monthly electricity consumption pattern mining: A fuzzy clustering-based model and a case study," Journal of Cleaner Production, vol. 141, pp. 900908, 2017.

[12]. G. Tsekouras, P. Kotoulas, C. Tsirekis, E. Dialynas, and N. Hatziargyriou, "A pattern recognition methodology for evaluation of load profiles and typical days of large electricity customers," Electric Power Systems Research, vol. 78, pp. 1494-1510, 2008.

[13]. A. Albert and R. Rajagopal, "Smart meter driven segmentation: What your consumption says about you," IEEE Transactions on Power Systems, vol. 28, pp. 4019-4030, 2013.

[14]. I. Benítez, A. Quijano, J.-L. Díez, and I. Delgado, "Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers," International Journal of Electrical Power & Energy Systems, vol. 55, pp. 437-448, 2014.

[15]. J. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," IEEE Transactions on Smart Grid, vol. 5, pp. 420-430, 2014.

[16]. R. Pal, C. Chelmis, M. Frincu, and V. Prasanna, "Time Series Clustering for Demand Response, An Online Algorithmic Approach," Online: www.cs.usc.edu/assets/007/93954.pdf.

[17]. Y. Wang, Q. Chen, C. Kang, and Q. Xia, "Clustering of electricity consumption behavior, “Authorized licensed use limited to: University of Cyprus. Downloaded on August 18,2020 at 08:11:46 UTC from IEEE Xplore. Restrictions apply dynamics toward big data applications," IEEE Transactions on Smart Grid, vol. 7, pp. 24372447, 2016.

[18]. R. Li, C. Gu, F. Li, G. Shaddick, and M. Dale, "Development of low voltage network templates—Part I: Substation clustering and classification," IEEE Transactions on Power Systems, vol. 30, pp. 3036-3044, 2015.
[19]. V. Figueiredo, F. Rodrigues, Z. Vale, and J. B. Gouveia, "An electric energy consumer characterization framework based on data mining techniques," IEEE Transactions on power systems, vol. 20, pp. 596-602, 2005.
[20]. O. Motlagh, G. Foliente, and G. Grozev, "Knowledge-mining the Australian smart grid smart city data: A statistical-neural approach to demand-response analysis," in Planning Support Systems and Smart Cities, ed: Springer, 2015, pp. 189-207.
[21]. S. V. Verdú, M. O. Garcia, C. Senabre, A. G. Marín, and F. J. G. Franco, "Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps," IEEE Transactions on Power Systems, vol. 21, pp. 1672-1682, 2006.
[22]. G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader, "Customer characterization options for improving the tariff offer," IEEE Transactions on Power Systems, vol. 18, pp. 381-387, 2003.
[23]. E. Carpaneto, G. Chicco, R. Napoli, and M. Scutariu, "Electricity customer classification using frequency–domain load pattern data," International Journal of Electrical Power & Energy Systems, vol. 28, pp. 13-20, 2006.
[24]. W. Labeeuw, J. Stragier, and G. Deconinck, "Potential of active demand reduction with residential wet appliances: A case study for Belgium," IEEE Transactions on Smart Grid, vol. 6, pp. 315-323, 2015.
[25]. S. Haben, C. Singleton, and P. Grindrod, "Analysis and clustering of residential customers energy behavioral demand using smart meter data," IEEE Transactions on Smart Grid, vol. 7, pp. 136-144, 2016.
[26]. T. Räsänen and M. Kolehmainen, "Feature-based clustering for electricity use time series data," in International Conference on Adaptive and Natural Computing Algorithms, 2009, pp. 401-412.
[27]. S. Zhong and K.-S. Tam, "Hierarchical classification of load profiles based on their characteristic attributes in frequency domain," IEEE Transactions on Power Systems, vol. 30, pp. 2434-2441, 2015.
[28]. A. Notaristefano, G. Chicco, and F. Piglione, "Data size reduction with symbolic aggregate approximation for electrical load pattern grouping," IET Generation, Transmission & Distribution, vol. 7, pp. 108-117, 2013.
[29]. I. Prahastono, D. King and C. S. Ozveren, "A review of Electricity Load Profile Classification methods," 2007 42nd International Universities Power Engineering Conference, Brighton, 2007, pp. 1187-1191, doi: 10.1109/UPEC.2007.4469120.
[30]. A. Rajabi, L. Li, J. Zhang, J. Zhu, S. Ghavidel and M. J. Ghadi, "A review on clustering of residential electricity customers and its applications," 2017 20th International Conference on Electrical Machines and Systems (ICEMS), Sydney, NSW, 2017, pp. 1-6, doi: 10.1109/ICEMS.2017.8056062.
[31]. Amit Saxena, Mukesh Prasad, Akshansh Gupta, Neha Bharill, Om Prakash Patel, Aruna Tiwari, Meng Joo Er, Weiping Ding, Chin-Teng Lin, A review of clustering techniques and developments, Neurocomputing, Volume 267, 2017, Pages 664-681, ISSN 0925-2312, doi.org/10.1016/j.neucom.2017.06.053.
[32]. A. Carlos, M. E. Bart De Moor, and J. A. K. Suykens. "Identifying Customer Profiles in Power Load Time Series Using Spectral Clustering." Artificial Neural Networks – ICANN 2009 Lecture Notes in Computer Science, 2009, pp. 315-24. doi:10.1007/978-3-642-04277-5_32.

[33]. Haben, Stephen, et al. “Analysis and Clustering of Residential Customers Energy Behavioral Demand Using Smart Meter Data.” IEEE Transactions on Smart Grid, vol. 7, no. 1, 2016, pp. 136–144., doi:10.1109/tsg.2015.2409786

[34]. Auder, Benjamin & Cugliari, Jairo & Goude, Yannig & Poggi, Jean-Michel. (2018). Scalable Clustering of Individual Electrical Curves for Profiling and Bottom-Up Forecasting. Energies. 11. 1893. 10.3390/en11071893.

[35]. Puente-Gil, Álvaro De La, et al. “Electrical Consumption Profile Clusterization: Spanish Castilla y León Regional Health Services Building Stock as a Case Study.” Environments, vol. 5, no. 12, 2018, p. 133., doi:10.3390/environments5120133.

[36]. Q. Zhao, M. Xu and P. Fränti, "Knee Point Detection on Bayesian Information Criterion," 2008 20th IEEE International Conference on Tools with Artificial Intelligence, Dayton, OH, 2008, pp. 431-438, doi: 10.1109/ICTAI.2008.154.

[37]. Bhissy, Kanaan & Faleet, Fadi & Ashour, Wesam. "Spectral Clustering Using Optimized Gaussian Kernel Function", International Journal of Artificial Intelligence and Application for Smart Devices, 2014, vol. 2, pp 41-56, doi:10.14257/ijaiasd.2014.2.1.04.

[38]. Dent, I., Craig, T., Aickelin, U., & Rodden, T. (2012). An approach for assessing clustering of households by electricity usage. Available at SSRN 2828465, doi: doi:10.2139/ssrn.2828465.