Pseudo measurements based on smart meters prosumer's characterization for distribution system state estimation.

Ales Svigelj

***Abstract*—**Growing complexity of conventional distribution systems due to increasing penetration of the distributed generation such as wind and solar plants, calls for better observability or monitoring capabilities in order to allow higher and higher levels of penetration of distributed energy resources such as electric vehicles and batteries. The distribution system state estimation is seen as the key technology for providing the full observability of distribution grid. The most important input parameters into distribution system state estimation are topology parameters, phasor measurements and pseudo measurements. To this end, in this paper we are focusing on calculation of pseudo measurements, based on real prosumers characterization. More precisely, we are proposing clustering based pseudo measurements calculation based on real smart meters data. The results show that proposed method can improve the estimated values of pseudo measurements.

***Keywords*—**pseudo measurements, prosumer characterization, smart grid, state estimation.

I. Introduction

CHARACTERIZED by a very strict configuration conventional power systems are still based on centralized generation, where electricity follows a very well defined path, from generation to consumption, going through transmission and distribution grid, until it reaches final consumers as depicted in Fig. 1 (left) [1]. However, due to the ongoing widespread deployment of distributed generation (DG) and energy storage elements, commonly called distributed energy resources (DER) in industrialized countries typical consumers are also expected to become electricity producers. DG, considered as small generators (in terms of power production), are connected at different levels (e.g. Medium Voltage (MV), Low Voltage (LV) and are depending on their size and geographical location as depicted in Fig. 1 (right) [1]. Hence, the structure of the distribution networks is thus substantially changing. Furthermore, the electrical grid capacity is also questioned with the prospects of increased load due to foreseen gradual replacement of conventional Internal Combustion Engine (ICE) vehicles by Electrical Vehicles (EVs) [2]. Thus, we are witnessing a shift from a mainly unidirectional power flows towards a fully bidirectional paradigm, where traditional consumers and producers are changing into so called prosumers (consumers & producers). In fact, DG expansion raised significant challenges to the grid operation and compels significant changes in the way the electrical power system is regarded from planning to operation of the electrical power system.

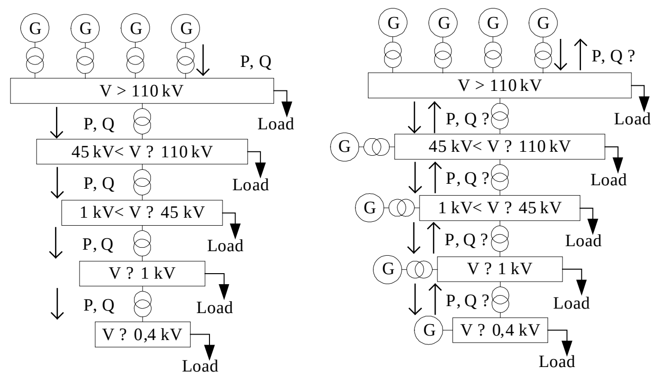

Fig. 1 Power system: Conv. (left), DEG (right) [1]

Moreover, due to renewable plants peculiar nature of operation their output may change drastically over a short span of time. E.g. the wind can vary significantly from one part of the day to another, and even over the course of a few minutes the power output of a wind farm can change dramatically. Solar power can be described in a similar manner. For instance, a small cloud can reduce the output of a photovoltaic solar farm for a few minutes and then raise it back to its previous level [3]. The less predictable generation and the continuously changing behavior of the consumers make power flows more uncertain. This requires increasingly sophisticated measurement instruments and techniques for power quality monitoring, fast detection of anomalous events and, accurate distribution system state estimation (DSSE) [4].

Some inherent level of flexibility designed to balance supply and demand at all times have all power systems. Variability and uncertainty are not new to power systems because loads change over time in sometimes unpredictable ways, and conventional energy resources fail unexpectedly. Variable

This work is partially funded by the EU under Grant agreement no. 619437 "SUNSEED." The author also acknowledges the financial support from the Slovenian Research Agency (research core funding No. P2-0016).

Ales Svigelj is with the Department of Communication Systems at Jozef Stefan Institute, Jamova cesta 39, Ljubljana, SI-1000 Ljubljana, Slovenia (phone: +386 1 4773-112; fax: +386 1 4773-111; e-mail: ales.svigelj@ijs.si).

renewable energy supply, however, can make this balance harder to achieve. In order to achieve higher proportion of renewable energy the grid control needs to be extended to the area where DER are [5], and where new high consumers are expected (e-cars) i.e. from the transmission network closer to users, i.e. to the distribution grid, thus start performing functionalities that were traditionally performed on the transmission network [6].

Distribution systems lack observability or monitoring capabilities, despite the fact that they need this due to their considerable diversity, variability, and vulnerability to disturbances. Several Phasor measurement units (PMU) can generally be found in the distribution network, but not in the quantity and density to provide all the information required as imposed with future demands. Advanced Metering Infrastructure (AMI) with smart meters (SM) are in the installation phase in the majority of the western countries and improve grid observability; still, 15 min data from them gives insight into power consumption, though real time state cannot be inferred. Further, renewables are changing their output on much more rapid time scale than that. According to [4] empirical measurements are required for several purposes:

- diagnosis of specific impacts of DER,
- establishing a baseline against which to compare impacts of DER,
- validation of new distribution feeder models
- ongoing monitoring to support operations and planning,
- exploration of as yet unknown phenomena on distribution systems.

In the grid the voltage (or current) phase angle is the key to power flow, dynamics and stability. Voltage phase angle measurement from the distribution system might address both known and as yet poorly understood problems, such as dynamic instabilities on the distribution grid, to enable new applications in the context of growing distributed intelligence and renewable resource utilization.

To achieve reliable and accurate knowledge of the grid condition, the real time DSSE is of key importance [6]. The benefit of the DSSE is that it can take into account all types of available measurements, thus reducing the investment costs into the required measurement infrastructure. Further, DSSE provides estimation of the grid state also on the grid nodes where measurement devices are not located.

A. State estimator

State estimator is a general name for a framework of several interconnected functional blocks of which the state estimation algorithm is the central one. Measurements need to be pre-processed, before they are used in the state estimation algorithm to detect possible errors. AMI measurements are used in two ways, firstly, as an input to the load flow algorithm. From the calculated load flow a number of results are used as pseudo measurements. These results are specified in a pseudo measurement scheme which is calculated in advance by the measurement placement algorithm. Secondly, AMI measurements are used in a so-called load estimator, a module that calculates a 15 minute load estimates for that particular AMI location with a corresponding standard deviation that is used as a weight in the state estimation algorithm. All these measurements are tested against a couple of predefined conditions, and in case they are not met the measurements are eliminated from the state estimation cycle. Prior to each state estimation cycle an observability analysis needs to be performed, to ensure that the state estimation algorithm will converge to a solution. If the observability analysis shows that the given measurement set is not adequate for a state estimation algorithm to converge, the measurement placement algorithm calculates a new pseudo measurement scheme, that ensure the state estimation algorithm convergence with updated measurement set [7]. The same measurement placement module is used to calculate optimal placement of SPM devices in the network. But generally that is only used prior to the system setup, since it is not very convenient to install and remove devices frequently. Fig. 2 shows the block diagram for the developed framework [8].

As for the SE algorithm different optimization algorithms can be used [7] to name a few:

- Weighted Least Squares estimator;
- Least Absolute Value estimator;
- SHGM estimator based on projection statistics;
- Least Median of Squares estimator.

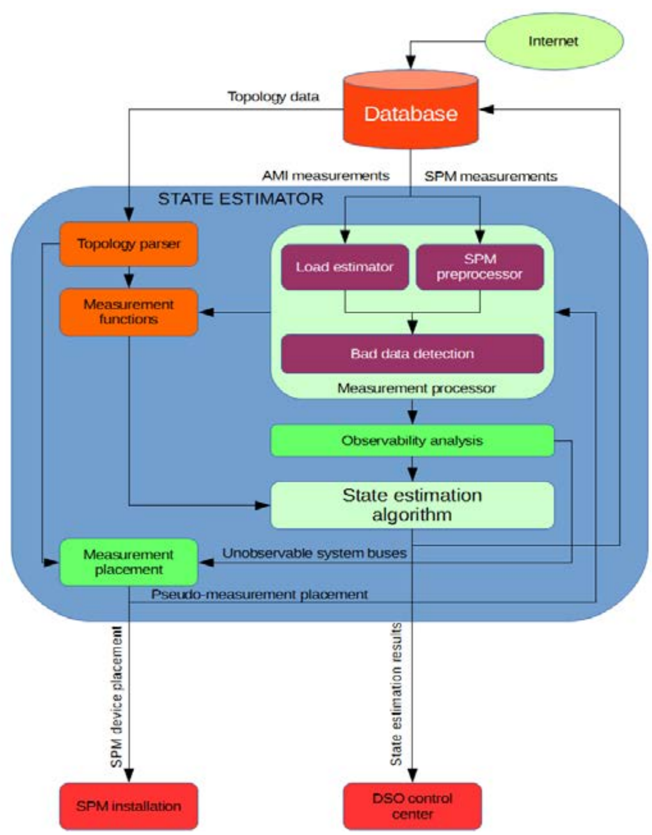

Fig. 2: State estimation block diagram [8]

With regard to power systems, the state estimation (SE) has been extensively studied and implemented in the transmission grid. In order for distribution grid to be able to facilitate renewable energy sources, large batteries and EV it is necessary to develop and implement the SE in distribution grids. Because of some inherent differences between transmission and distribution grids, straight forward implementation of the transmission SE software into distribution grids is not possible. Main distribution (in comparison to transmission) system characteristics affecting the DSSE are [9, 10]:

- High resistance to reactance ratio in lines and cables. Since lines and cables have predominantly a resistive character, this eliminates the possibility of adopting simplifications commonly used in estimators developed for transmission systems. Also this fact calls for very tightly synchronized and accurate measurement devices.
- Phase imbalance. Loads in distribution grids are mostly non-symmetrical in nature; this causes the three-phase quantities to be non-symmetrical. This characteristic is made worse with growth of distributed generation. The state estimation software needs to take into account phase imbalance.
- Large number of nodes. In comparison to transmission networks, distribution networks have much larger number of nodes. This in turn increases the computational complexity of the state estimation problem. Another aspect of this fact is, that, to achieve the necessary accuracy of the state estimate, the number of measurement nodes required is larger.
- Uncertainty of network parameters. State estimation framework requires values of line impedances. These are generally computed from known line length and known line material. These parameters can have large uncertainty because some infrastructure is very old and was not accurately measured at the time of installation. This leads to degradation of state estimator results.

In development of the DSSE this characteristics need to be addressed [9, 8]. Within the FP 7 SUNSEED project one of the main goals were to establish the observability of distribution system using the DSSE in the real testbed [8]. Furthermore, SUNSEED proposes an evolutionary approach to utilization of already present communication networks from both energy and telecom operators. These can be suitably connected to form a converged communication infrastructure for future smart energy grids offering open services.

In general for the SE the following inputs are required [9]:

- topology data obtained from the DSO,
- synchro phasor measurement,
- AMI and pseudo measurements.

However, in this paper we are focusing on AMI and pseudo measurements calculations that can be used in DSSE calculation. Thus, the real AMI measurement was analyzed. We describe our approach for obtaining near real time load estimation based on the latest AMI measurements. More specifically, we are dealing with the characterization of prosumers, based on AMI measurements from the observed testbed(s).

The paper is structured as follows. Section 2 introduces prosumers characterization. The pseudo measurements calculation based on prosumer characterization from SM is proposed and evaluated in Section 3. Section 4 concludes the paper.

II. Characterization of DS prosumers

The data collection of electric energy consumption, processing and analysis capabilities can be significantly enhanced with the deployment of SM in low voltage distribution networks, within the context of AMI systems. In order to ensure higher reliability and efficiency in distribution power systems, the information on prosumers electric consumption/production patterns is particularly important. In fact, the communication between utilities companies and electricity customers aims to provide information on consumption of consumers in order to make a better use of electrical energy.

Furthermore, the knowledge about costumer’s behavior can be a useful decision tool, not only for utilities, but also for consumers [11]. The knowledge resulting from the study of load profile can be used by the utilities to identify the aspects that cause the increase of the diagrams’ peaks and development of specific prosumer’s contracts.

In addition to above demand response scenarios, the usage of AMI is of paramount importance in the light of the limited number of real-time measurements in distribution systems. Thus, the increased smart metering data availability from LV consumers is necessary to provide near real time load estimates, which can be used also for calculation of pseudo-measurements, used as an input in the DSSE. Though the planned frequency update of the state is at the highest 1 second, the AMI measurements are reported only once every 15 minutes. In general this means that at the time of new state calculation only obsolete AMI measurements are available.

A. *Description of AMI measurements*

The vital element of an advanced AMI dedicated to measuring ‘end users’ (prosumers) energy consumption and production is SM. Nowadays, the main AMI functionality is billing improvement and remote asset monitoring. Billing data is usually generated on 15 min interval where intervals are collected and sent via data concentrator to billing management center (HES) once per day, where they are analyzed. Since SMs allow us to monitor not just energy (active, reactive) but also different parameters (varies from AMI vendor to vendor), DSO started collecting various data used for distribution power grid stability, energy loss calculation, fault detection, power quality monitoring, etc. Nevertheless not all data could be sent to the HES due to communications limitations (PLC, GSM/GPRS) and the big data handling. Due to that those reasons the DSO has to make a compromise what parameters/data should be transmitted to HES and used for further analyses, besides billing purposes.

Current AMI measurements that DSOs monitor/gather from

AMIs can be summarized as:

- Grid visualization values (delivery point load and network status - *V*, *P*, *Q*, *f*, etc.); with AMI meters dispersion in a certain region we are improving power grid visualization in a more transparent way due to AMI measurements that AMI meters could provide, for example prosumers load profile (aka delivery point load);
- Fault management values (alarms, last gasp, anti-tamper, multi-energy events, etc.);
- Asset management values (endpoint consumption measurement, billing, profiling, end user interaction, etc.).

AMI meter captures instant voltage, current and further calculates effective value that is used for calculating energy values (kWh) in compliance with standard SIST EN 50470-1 and SIST EN 50470-3. AMI meter could use more logs but main ones are Load Profile and Log Book. In Load Profile we can find active and reactive power, *f*, *U*, *I*, etc. In Log Book we can find different messages (power outage, power up/down, various alarms, etc.) Due to the vast number of choices what AMI meter can capture every DSO has its own demands for 'must have' values.

Within the SUNSEED project [8] the data from AMI meters is transferred securely through established DSO communication infrastructure and is gathered in the DSO databases, from where it is forwarded to SUNSEED main database storage [6].

B. Prosumer's behavior characterization

Based on the measurement we identify the load profiles for individual users. It is worth noting that we are investigating the real measurements from the selected testbeds. Load profiles for users are a baseline for pseudo measurement calculations as described in section 3.

The initial overview of the historical AMI data shows that loads depends on the day of the week (working day, Saturday, Sunday). This is in particular true for the big industry consumers as shown in Fig. 3 (time is starting at Monday midnight).

It is clearly seen that the both consumers have typical daily routine during the weekday, while on Saturdays and Sundays the consumption is much lower in particular if the Saturday and Sunday are not working days. Furthermore, also during weekdays the consumption can be very low due to Bank holidays.

In the observed testbeds there are few solar plants whose profile heavily depends on the weather conditions and the month of the year (January … December), due to sun insensitivity. The raw data showing the daily production in kW (15 min measurements) for a year (envelop is easy visible) are depicted in Fig. 4. The envelop showing the difference between winter and summer is easy visible. The daily production for one day, (i.e. 28 one-day consecutive curves) is shown in Fig. 5. The production heavily depends on the weather conditions (e.g. sunny, cloudy …). The effect of cloudiness is further evaluated in [3] while the effect of wind turbines on the stability of the power system is evaluated in [12].

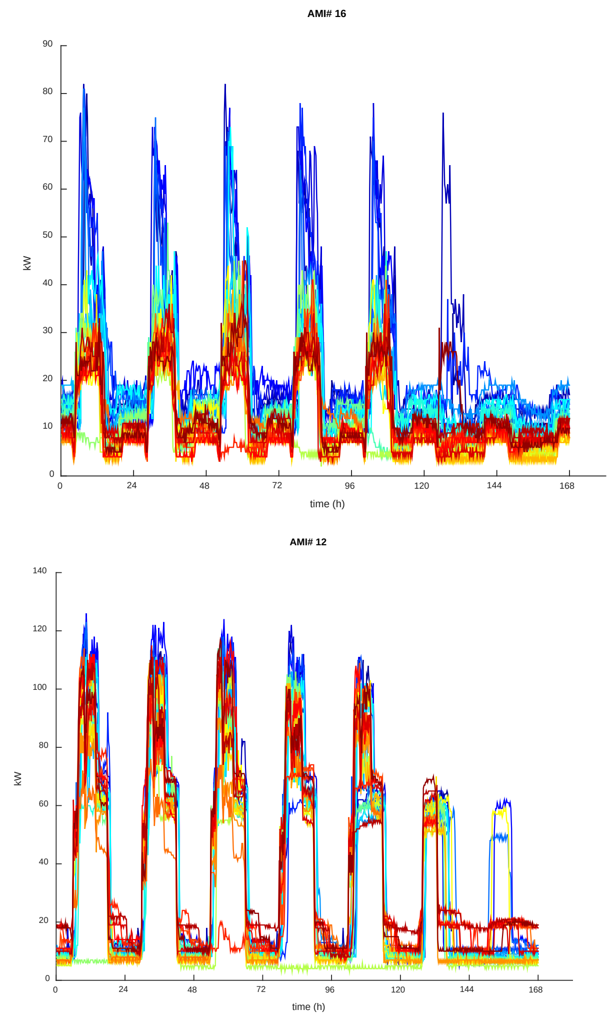

Fig. 3: Weakly loads (28 weeks) (P) raw data for two selected consumers a higher dimensional feature space

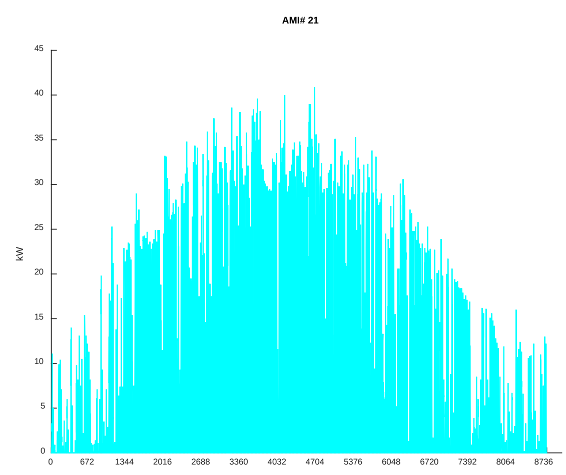

Fig. 4: Daily production of solar plant (*P*) for 1 year

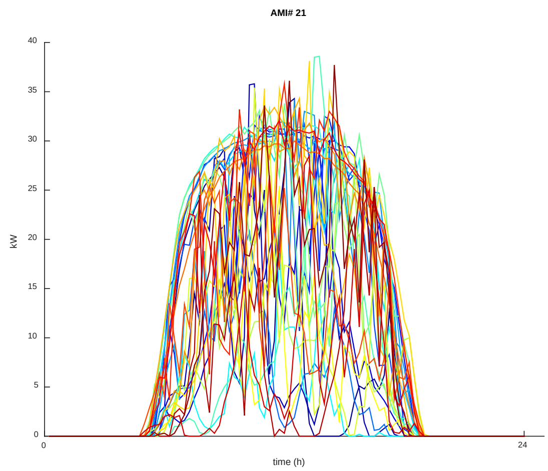

Fig. 5: Daily production of solar plant (*P*) for 1 day (28 consecutive days)

Based on historical real measurement in order to identify typical load profiles we apply clustering algorithms to each individual AMI measurement data. Clustering can be defined as the process of partitioning a large database into groups (or clusters) based on a concept of similarity or proximity among data. There is a wide variety of clustering algorithms, although there is no single algorithm that can, by itself, discover all sorts of clusters shapes and structures. Good clusters present high similarity within a certain group and low or a very different similarity among objects of others classes [11]. Despite the several clustering methods for cluster analysis, for a certain database, each clustering method may identify groups which member objects are different. Thus, two important questions need to be addressed, associated to the clustering procedure, namely, what is the best clustering method that produces the best data partition and how many clusters should be presented in the data [11]. It should also be noted that the number of groups may be strongly influenced by the analysis and suggestions of experts.

The objective of data clustering techniques consists of dividing a data set X composed of n data patterns $\{x_1,...,x_n\}$ into K clusters $\{C_1,...,C_K\}$, such that similar data patterns, x_i ,x_j , are placed in the same cluster, i.e. $\{x_i, x_j\} \in C_k$, and dissimilar data patterns are grouped into different clusters, i.e. $x_i \in C_k$, $x_j \in C_k$, $k \neq 1$. The set of clusters P = $\{C_1,...,C_K\}$ is referred to as data partition

In [9] authors assembled the representative load diagrams of electric customers in clusters, according to a similarity criterion. Several clustering algorithms have been used, namely: Complete-link (CL); Average-link (AL); Ward´s-link (WL); Normalized Cut algorithm (NC) and K-means algorithm (KM). The clustering algorithm results showed that the best partition according to the clusters validity indices was the K-means algorithm. It was identified that the suitable number of clusters (K) to this case study was two taken into account their validity indices. However, with a visual analysis of the obtained load profiles for each cluster it was verified that this number was not adequate [11]. In [11] the authors had chosen a partition of 8 clusters and selected the most adequate from the obtained results. For the profiling we used the well-known K-means algorithm. However, we are aware that some other clustering methods should be evaluated as well as they might better utilise WLS (weighed least squares) algorithm used in the State Estimation [9, 13].

The K-means algorithm [14] is probably the best known data clustering algorithm. K-means tries to minimize the sum of squares within cluster:

$$\sum_{k=1}^{K} \sum_{x_i \in C_k} \left\| x_i - \overline{x_k} \right\|^2 \tag{1}$$

where $\left\| x_i - \overline{x_k} \right\|^2$ is the Euclidian distance between pattern xi and its closest cluster centroid x_k.

This algorithm takes as parameter the desired number of clusters K and randomly chooses K data patterns as the initial centroids $\{x_1,...,x_K\}$ of each cluster. Then, K-means algorithm iterates between two steps: find for each pattern $x_i \in X$ the closest centroid x_k and assign it to the corresponding cluster C_k and update each centroid x_k as the mean vector of the corresponding cluster C_k, i.e.:

$$\overline{X}_k = \frac{1}{|C_k|} \sum_{x_j \in C_k} X_i \tag{2}$$

where $|C_k|$ is the number of data patterns that belong to C_k.

This process is repeated until no pattern assignments are changed from one iteration to the next one meaning the algorithm converged to a (local) minimum.

The identified clusters on per user bases will be used also for near real time load estimation as described in section 3. For the determination of number of clusters per user the visual evaluation shows that in general 3 main clusters covers the main profiles. As an example in Fig. 6 we show clustering of industry consumer to 3 clusters (28 consecutive days). In addition to typical representative of the cluster we also show raw data, clustered data and standard deviation (STD) of clustered data from their representative.

However, because we would like to automatically determine the appropriate number of clusters we compare sums of the distance from every pattern in a cluster to the average of the cluster (denoted as *SUM(CK)* of each curve for different number of clusters K started with one cluster K= 1,2,3…. Obviously, with the increase of the number of clusters K the sum is decreased.

ISSN: 1998-4464

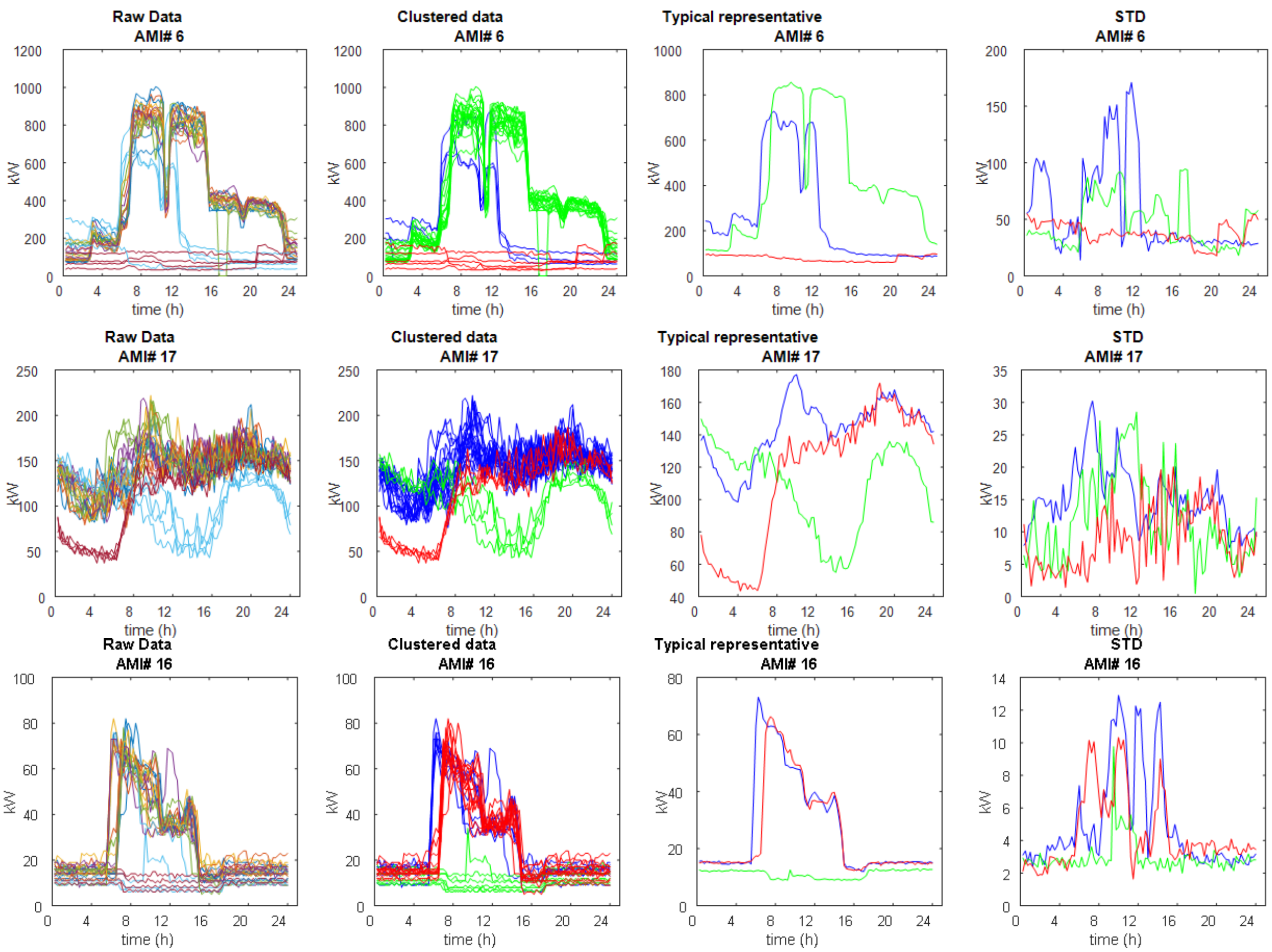

Fig. 6: Example of clustering of an industry consumer (high load).

For easier comparison we also defined the normalized sum of cluster C_k as:

$$SUM_{norm}\left(C_K\right)=\frac{SUM\left(C_K\right)}{SUM\left(C_1\right)} \tag{3}$$

Thus, if there is only one cluster the SUM_{norm} is 1. Graphical representation of SUM_{norm} for up to 5 clusters for some typical AMI meters is shown in Fig. 7. It can be seen that in the case of more than 3 clusters, is not improved significantly in the majority of cases.

The same analysis was done for the reactive power (Q) as depicted in Fig. 8. We present the same number of clusters as in upper case, although it can be seen that in this case of (AMI# 6) the 2 clusters are sufficient, as their typical representatives are very similar.

In the case of solar plants they can be put in general in at least two clusters, either working fully or not, as shown in Fig. 9 (three clusters are shown as an example).

Based on above characterization and clustering, we developed clustering based algorithm for load estimation, described in the next section, which serves as an input to DSSE.

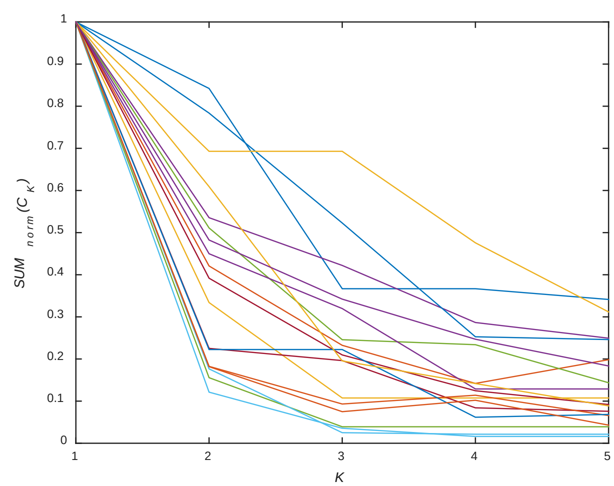

Fig. 7: SUM_{norm} in dependence of number of clusters

ISSN: 1998-4464

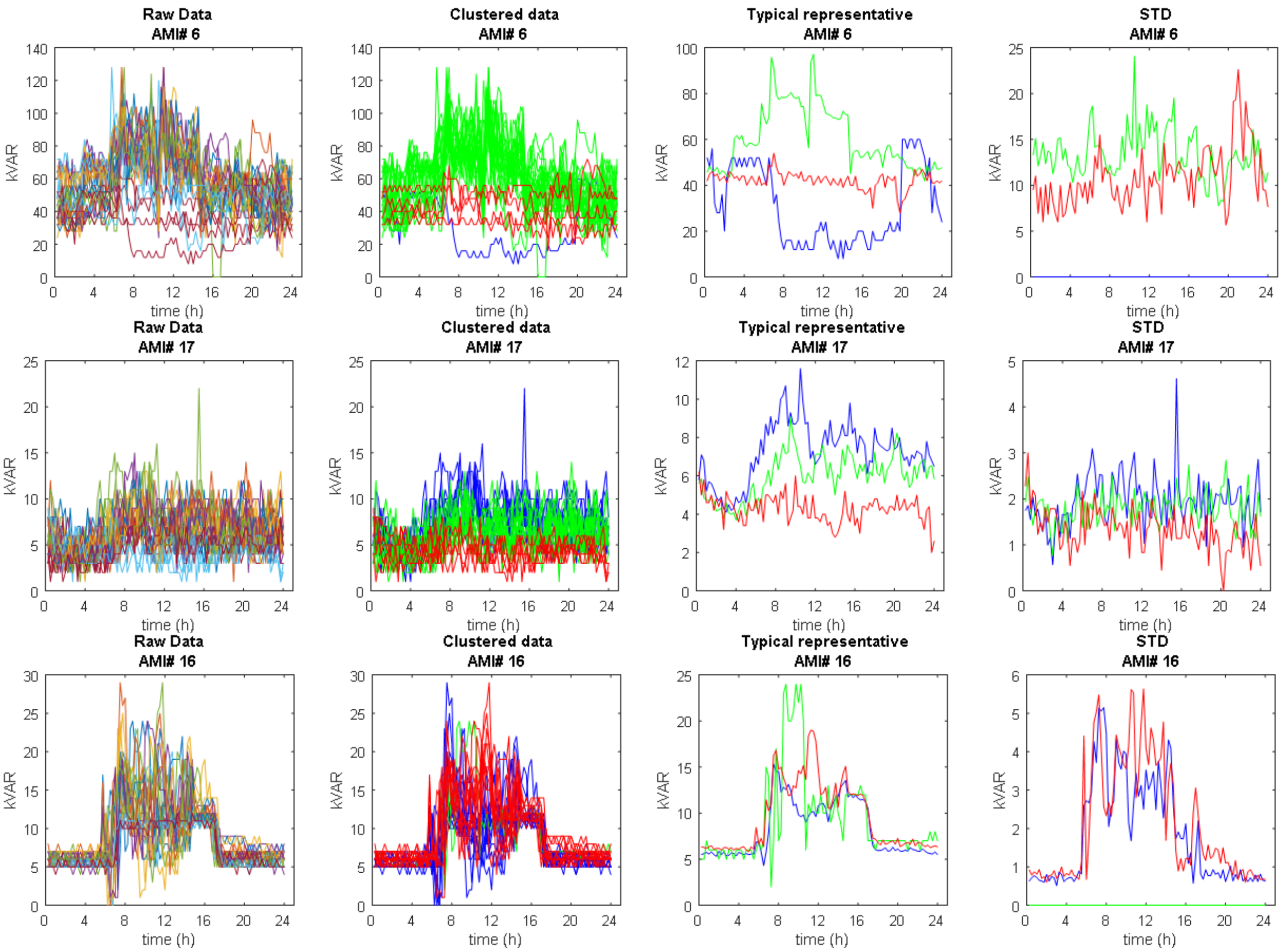

Fig. 8: Example of clustering of an industry consumer (high load).

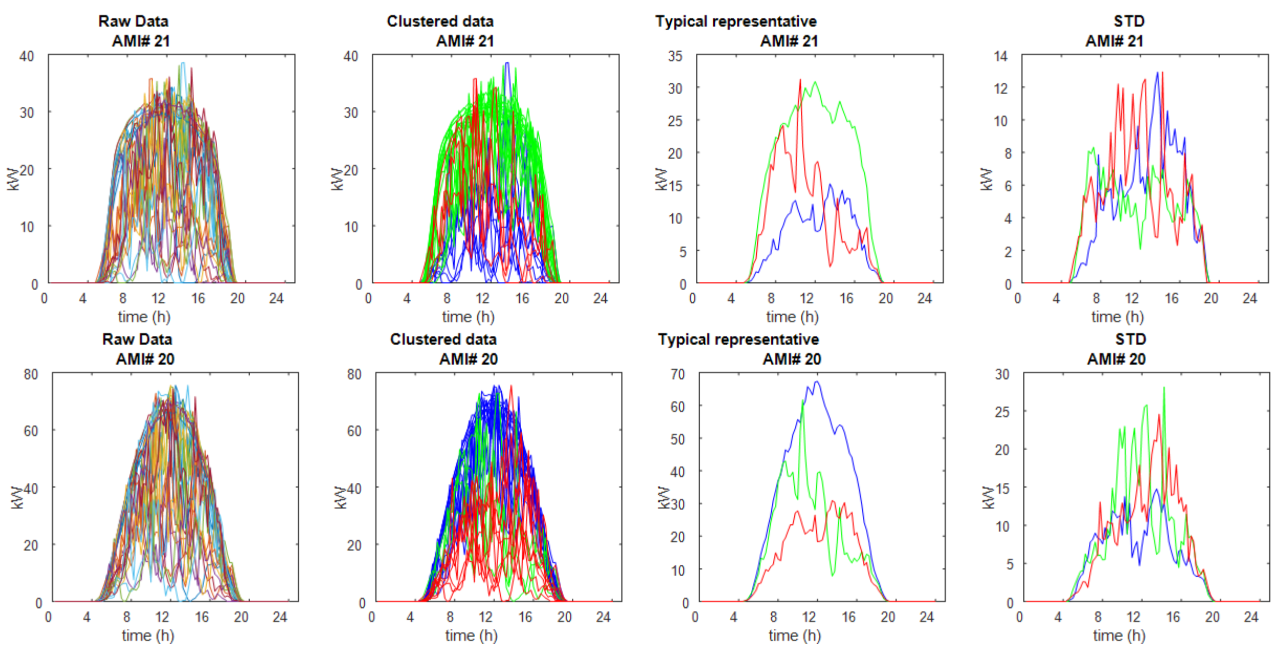

Fig. 9: Clustering of a solar plant

ISSN: 1998-4464

III. Pseudo measurement calculation

While measurements and monitoring of power systems transmission grids is well established and pervasive, the same cannot be said for distribution grids. Here measurements nodes are very sparse, often limited to primary substations, and absent in some portions of the grid, particularly in the low voltage (LV) network. In the light of implementing of state estimation more broadly in the distribution grid, it is of paramount importance that by lacking dedicated measurements it should be augmented with all available measurements also from AMI smart meters (domestic and industrial), and most of all with pseudo measurements.

Historical data collected over the years are used to build load profiles (mainly to be used in planning) that are then also used as sources for near real time measurement estimation and calculation of pseudo measurements, as discussed in previous section. In addition they can be used also for load forecasting purposes.

The quality of state estimation is highly dependent on modelling of loads. This is in particular true in the absence of any real measurement of loads which are highly distributed and diverse. Because of this they are usually treated as random variables with appropriate mean and variances [13]. Thus the accurate load models are critical for state estimation [13].

In the distribution network the load modelling at each injection node is done based on different measurements, which are (are not) available at particular node. In general, we can determine three “classes” of measurements:

- Real time measurement node: the consumed/produced P, Q power or even phasors are provided in real time. In this case the information can be used directly in SE.
- Near real time measurement node: the consumed/produced P, Q power is provided in near real time. Typical sources of such measurement are smart meters that reports the measurements P, Q power in intervals (typically 15 minutes) which are than the aggregated and measurements are send out delayed (typically from 15 minutes to 1 day)
- Unmeasured nodes: in this case usually the load profiles are developed for each type of customer (such as residential, industrial, etc.), based on some monitoring and energy bill data. Historical samples obtained for different seasons, days and times, are stored separately for different load types (residential, industrial etc.).

In the SUNSEED system [8] for the real time measurements we use PMU like devices at the “important” nodes, while the AMI smart meters are installed at each prosumer, generating the near real time measurements (P, Q power) which are sent to central node with approximately 15 min delay. Based on measurements information and historical data the estimated load is calculated as described further in this section. In our case we do not have unmeasured nodes with prosumers, thus this case will not be further evaluated.

In the following based on historical data from the smart meters we evaluate different load estimation methods. The particular attention is given on “quality” of such estimates, which serves as one of the parameters into DSSE.

For the comparison between different load estimation algorithms used, two ordinary statistical measures are used to quantify the difference between estimated and actual values: Relative Root Mean Square Error (RRMSE) and Mean Absolute Percentage Error (MAPE), given in (4) and (5) respectively.

$$RRMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(P_{i,act} - P_{i,est}\right)^2}{\sum_{i=1}^{n} {P_{i,act}}^2}} \tag{4}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{P_{i,act} - P_{i,est}}{P_{i,act}}\right| \tag{5}$$

First we evaluate the algorithm described in [15] (denoted as *Prev_day_based*), which uses near real- time information from AMI along with previous day’s data. Their estimates are done in one day old data. It deploys a simple time series model, given as.

$$P_{i,today} = \frac{P_{i-1,today}}{P_{i-1,prevday}} P_{i,prevday} \tag{6}$$

where:

i, i-1 load estimation time interval and previous time interval, respectively (i.e. 15 min interval in our case)

today, prevday day of load estimation and previous day (weekday or weekend depending on day of estimation, (e.g. *prevday* of Saturday (Sunday) is Saturday (Sunday) a week before), while for working day *prevday* is previous working day).

P_i corresponding electricity consumptions/productions.

The results for some AMI meters and different estimation algorithms are summarized in Table I. The results are based on historical data and are based on about 1000 consecutive estimates (100 days, each 15min interval). The results for *Prev_day_based* algorithm *RRMSE* are between 12% and 58% while for the *MAPE* it is between 10% and 63%. This results are not satisfying, thus we investigate two different approaches. However, as we are working on 15 min scale, we try to use for the estimate just the same value as it was the previous one, according to equation (7) denoted as *Prev_measurement_based*.

$$P_i = P_{i-1} \tag{7}$$

where: *i, i-1* load estimation time interval and previous time interval, respectively (i.e. 15 min interval in our case)

ISSN: 1998-4464

The results for *Prev_measurement_based* algorithm are summarized in Table I (5 selected consumers). On so short scale (15 min) it outperforms significantly the *Prev_day_based* algorithm. RRMSE is lower for up to 28 percentage points while MAPE is lower up to 15 percentage points.

TABLE I: RRMSE MAPE FOR DIFFERENT ESTIMATION ALGORITHMS FOR SELECTED AMI METERS (P).

	Prev_day_based		*Prev_measure-ment_based*		*Clustering_based*	
AMI	RRMSE	MAPE	RRMSE	MAPE	RRMSE	MAPE
6	24.9%	10.4%	16.8%	9.5%	14.5%	8.7%
9	52.1%	63.5%	30.3%	41.4%	29.7%	41.7%
11	58.4%	17.2%	30.0%	9.8%	29.5%	10.1%
16	23.7%	18.4%	15.5%	12.9%	14.9%	12.7%
17	12.8%	10.0%	9.1%	7.1%	9.0%	7.0%

The further in depth analysis of the errors as depicted in Fig. 10 (example for AMI #6 for few days) show that the highest errors is observed when there are high changes in the Load (P) as we expected. The red line presents the error for *Prev_measurement_based* algorithm while the green one represents the error for *Clustering_based* algorithm discussed below.

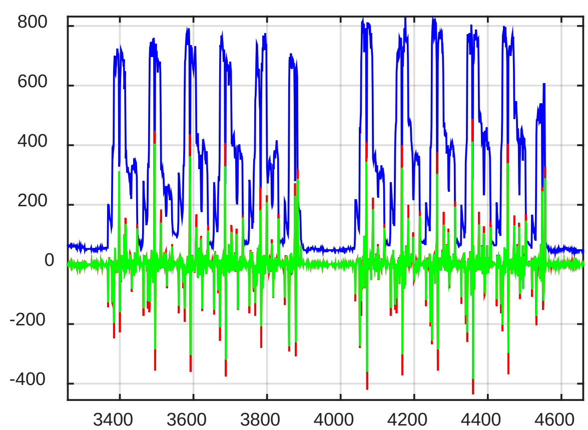

Fig. 10: Actual measurements (blue) and error (green, red) of estimated value (kW) for AMI# 6

Thus, we propose a clustering based algorithm (*Clustering_based*), which refines the previous value on the cluster based method (for clustering see also section 2) as explained in the following Flow chart (Fig. 11) and shown in (8).

$$P_i = P_{i-1} + \alpha\left(C_{k,i} - C_{k,i-1}\right) \tag{8}$$

where:

i, i-1 load estimation time interval and previous time interval, respectively (i.e. 15 min interval in our case)

C_k the closest Cluster k to last m measurements

α the impact of the cluster

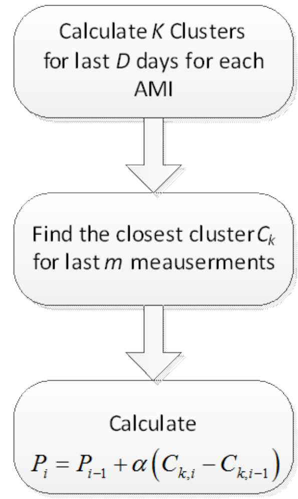

Fig. 11: Flow chart of Clustering based algorithm

The following parameters (giving the optimal results) are used in the evaluation: number of clusters K = 5; based calculated on last D = 38 days. m = 96 (i.e. 1 day), α is set to 0.2. The results summarized in Table I show that the estimated values can be further improved. Even though the average improvement is not significant the in depth investigation reveals (see Fig. 10) that the pick errors are smaller in the case of the proposed clustering based method (green). In Table II we present results that have RRMSE and MAPE on per sample bases are higher than 5% (i.e. as denoted in (9)).

$$\frac{\left|P_{i,act} - P_{i,est}\right|}{\max\left(\left|P_{i,act} - P_{i,est}\right|\right)} > 0.05 \tag{9}$$

TABLE II: RRMSE MAPE FOR DIFFERENT ESTIMATION ALGORITHMS FOR SELECTED AMI METERS (P).

	Prev_measurement_based		*Clustering_based*	
AMI	RRMSE	MAPE	RRMSE	MAPE
6	27.2%	26.1%	22.0%	21.1%
9	37.2%	60.8%	36.2%	59.6%
11	47.4%	54.6%	47.1%	52.1%
16	21.5%	27.8%	19.6%	24.5%
17	9.9%	8.3%	9.8%	8.2%

The graphical representation is shown in Fig. 12, where we can observe that in the case of clustering based method the errors are up to 390 kWh (right) while in the *Prev_measurement_based* case are up to 470 kWh (left).

Similar results are achieved also for reactive power Q as summarized in Table III. However, as AMI # 9 and #11, has very small reactive power Q the results are not representative.

We apply the same algorithms to solar plants and the results are summarized in Table IV, however, results for *Prev_day_based* calculation is not relevant due to infinitive numbers (division by zero)

ISSN: 1998-4464

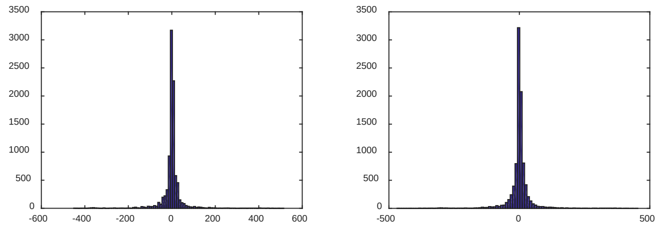

Fig. 12: Histogram of errors (kWh) for AMI #6mapping nonlinear data to a higher dimensional feature space

TABLE III: RRMSE MAPE FOR DIFFERENT ESTIMATION ALGORITHMS FOR SELECTED AMI METERS (Q).

	Prev_day_based		*Prev_measure-ment_based*		*Clustering_based*	
AMI	RRMSE	MAPE	RRMSE	MAPE	RRMSE	MAPE
6	28.8%	20.1%	19.7%	12.9%	19.0%	12.8%
9	N/A	N/A	116%	30.3%	114.5%	30.2%
11	N/A	N/A	37.3%	30.2%	36.3%	29.7%
16	28.3%	19.6%	20.9%	14.4%	20.6%	14.5%
17	47.8%	41.5%	29.3%	28.0%	29.2%	28.2%

TABLE IV: RRMSE MAPE FOR DIFFERENT ESTIMATION ALGORITHMS FOR SELECTED AMI METERS (P) SOLAR PLANT.

	Prev day based		*Prev measurement based*		*Clustering based*	
AMI	RRMSE	MAPE	RRMSE	MAPE	RRMSE	MAPE
20	N/A	N/A	16.1%	16.3%	15.9%	15.5%
21	N/A	N/A	17.3%	20.4%	17.0%	19.4%

Load estimation which serves as an injection parameter in SE is only one of the parameters, as each load estimation value should be accompanied by the measure of how good the estimation is. Thus we are analyzing the mean value and standard deviation of errors of estimated values. For all the consecutive samples they are summarized in Table V.

TABLE V: MEAN AND STD OF ERROR FOR DIFFERENT ESTIMATION ALGORITHMS FOR SELECTED AMI METERS (P).

	Prev_measurement_based		*Clustering_based*	
AMI	mean	std	mean	std
6	0.01	64.34	0	55.75
9	0	5.46	0	5.4
11	0	6.25	0,01	6.13
16	0	2.3	0	2.2
17	-0.01	13.73	-0.02	13.63

However, in depth investigation reveals that standard deviation is heavily dependent on time of the day (for all estimation algorithms). As time of day influence the STD significantly it should be taken into account. Thus, we divide the day in 24 hours and calculate the mean and STD for each hour (4 measurement on 15 min) for *N_days* (i.e in our case *N_days* = 100). The results for a typical AMI meters are shown in Fig. 13.

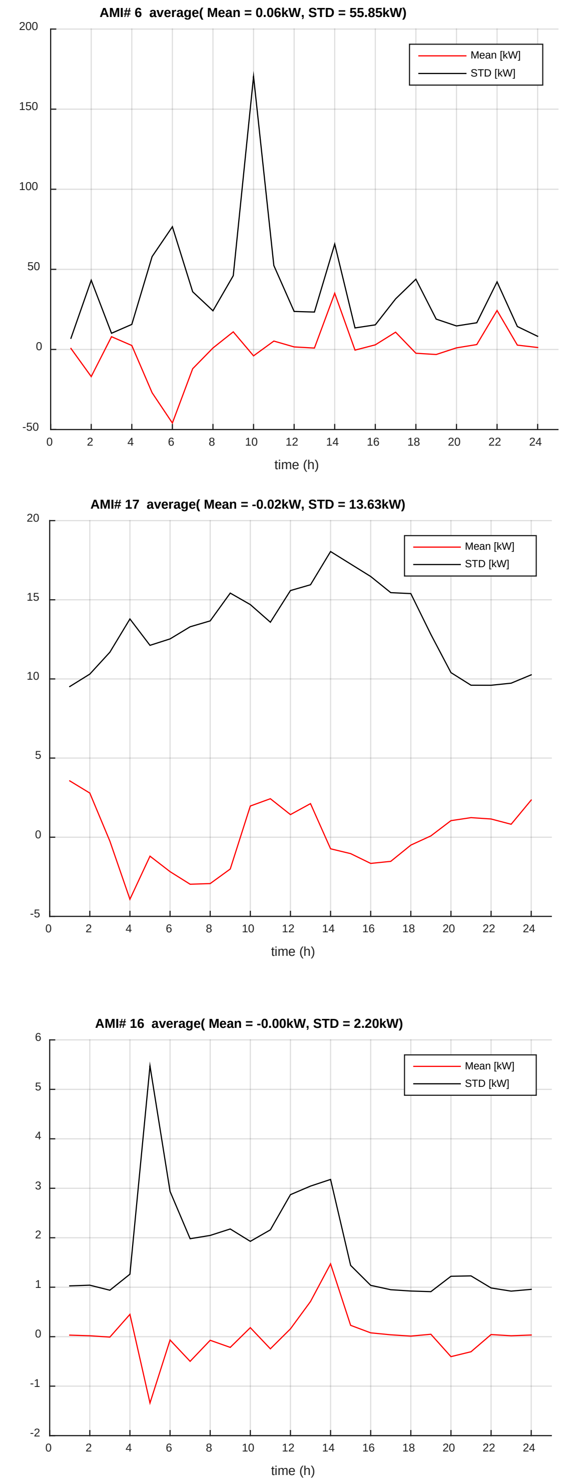

Fig. 13: Mean and STD for AMI #6, #17 and #16

Thus, as Load estimation input for SE we use P_i and Q_i, calculated using *Clustering_based* algorithm and the corresponding STD (based on historical values) for the particular time slot (i.e. 1 hour). In such a manner the "reliability" of Load estimation does not depend only on particular AMI meter, but also on time of the day. The pseudo

ISSN: 1998-4464

measurements based on load estimation are than calculated using load flow calculation and used in the state estimation algorithm.

IV. Conclusion

Many utilities are turning to Smart Grid solutions such as distributed energy resources (DERs) (i.e. small scale renewable energy sources and energy storage to balance load and capacity without building large-scale generation). Due to the increasing penetration of DER, the smart grid needs more and deeper monitoring and control to maintain stable operation The distribution system state estimation is seen as the key technology for providing the full observability of distribution grid. In this paper we focused on the pseudo measurements calculations from AMI which are one of the important inputs in distribution system state estimation paramount importance. Thus in this paper, we propose the pseudo measurements calculation based on prosumers characterization from the smart meters in the real network used within FP 7 SUNSEED project. The results were evaluated in terms of RRMSE and MAPE. In addition, we also evaluated how the errors are dependent on time of the day. The results show that the proposed *Clustering_based* algorithm outperforms both *Prev_measurement_based* and *Prev_day_based alghorithms*. Thus, we proposed it for the real testbed state estimation calculation.

References

[1] P. Miguel, P. da Rocha Almeida, "Impact of Vehicle to Grid in The Power System Dynamic Behaviour," PhD Dissertation, Faculty of Engineering, University of Porto, November 2011.

[2] D. Poli, M. Giuntoli et al., "The PRIME Project: the assessment of the impact of electric vehicles on the power system and on the environment", *WSEAS Advances in Power and Energy Systems*, Prague, 2012.

[3] P. Chrobák, J. Skovajsa, M. Zálešák, "Production of Electricity Using Photovoltaic Panels and Effects of Cloudiness", *WSEAS Transactions on Power Systems*, Vol. 12, 2017, pp. 346-35.

[4] A. Meier, D. Culler, A. McEachern, R. Arghandeh, "Micro-Synchrophasors for Distribution Systems", *IEEE PES Innovative Smart Grid Technologies Conference, (ISGT 2014)*, 2014.

[5] D. Santis, G. Abbatantuono, S.Bruno, M. ScalaLow, "Voltage Grid Control through Electrical Vehicles Charging Stations", *WSEAS Transactions on Power Systems*, Vol. 11, 2016, pp. 283-288.

[6] J. Nielsen, U. Kuhar, A. Svigelj, et al. "Secure real-time monitoring and management of smart distribution grid using shared cellular networks", *IEEE wireless communications*, 2017, vol. 24, no. 2, pp. 10-17.

[7] U. Kuhar, G. Kosec. M. Pantos, A. Svigelj, "The Impact of Model and Measurement Uncertainties on a State Estimation in Three-Phase Distribution Networks", *IEEE Transactions on Smart Grid*, In Press 2018, DOI: 10.1109/TSG.2018.2823398.

[8] EC FP7 SUNSEED (Sustainable and robust networking for smart electricity distribution), https://sunseed-fp7.eu/ , accessed 2018.

[9] U. Kuhar, J. Jurše, K. Alič, G. Kandus, A. Svigelj, "A unified three-phase branch model for a distribution-system state estimation", *In proc.of IEEE PES innovative smart grid technologies, Europe*, (ISGT-Europe) Ljubljana, Slovenia, 2016.

[10] D. Della Giustina, M. Pau, P. A. Pegoraro, F. Ponci, S. Sulis, "Electrical distribution system state estimation: Measurement issues and challenges", *IEEE Instrumentation and Measurement Magazine*, vol.1094, no. 6969/14, 2014.

[11] S. Ramos, J. Duarte, J. Soares, Z. Vale, F. Duarte, "Typical Load Profiles in the Smart Grid Context – A Clustering Methods Comparison", *IEEE Power and Energy Society General Meeting 2012*, San Diego, California, USA, July 2012.

[12] F. R. Islam, A. Lallu, K. A. Mamun, K. Prakash, A. A. Rattan Impact of Wind Generators in Power System Stability WSEAS Transactions on Power Systems, ISSN / E-ISSN: 1790-5060 / 2224-350X, Volume 13, 2018, Art. #25, pp. 249-257.

[13] R. Singh, B. C. Pal, and R. A. Jabr, "Choice of estimator for distribution system state estimation," *Generation, Transmission & Distribution, IET*, vol. 3, no. 7, 2009, pp. 666–678.

[14] J. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations", *5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1967, pp. 281-297.

[15] T. C. Xygkis, G. D. Karlis, I. K. Siderakis and G. N. Korres, "Use of near real-time and delayed smart meter data for distribution system load and state estimation," *MedPower 2014*, Athens, 2014, pp. 1-6.

ISSN: 1998-4464