Detection of Anomalies in Power Profiles using Data Analytics

Grigore Stamatescu
Automation and Industrial Informatics
University Politehnica of Bucharest
Bucharest, Romania
grigore.stamatescu@upb.ro

Radu Plamanescu
Faculty of Electrical Engineering
University Politehnica of Bucharest
Bucharest, Romania
radu.plamanescu@upb.ro

Irina Ciornei
MicroDER Lab, University Politehnica of Bucharest
KIOS Centre of Excellence, University of Cyprus
ciornei.irina@ucy.ac.cy and irina.ciornei@upb.ro

Mihaela Albu
Faculty of Electrical Engineering
University Politehnica of Bucharest
Bucharest, Romania
mihaela.albu@upb.ro

***Abstract*—Deployment of high reporting rate smart metering infrastructure together with a multitude of sensors for automation and control are an increasing trend among energy communities and prosumers. These systems provide useful information for data-driven prediction and classification models for micro-loads and local power generation. Matrix Profile is a promising general purpose data mining technique for time series data, such as electrical measurements from advanced smart meters. In this work, we first describe the measurement context that provides rich data availability for current advanced energy analytics applications. We target power profiles for both generation and load to highlight salient and complementary characteristics thereof, which can be leveraged in applications involving data-driven analytics for enhancing observability in distribution grids. A sensitivity analysis investigating the chosen method under various input noise assumptions is presented using Monte Carlo simulation. The comparative results indicate the relative robustness of the Matrix Profile approach for anomaly detection tasks in energy measurements traces.**

***Index Terms*—energy analytics, power measurements, time series, matrix profile, anomaly detection**

I. Introduction

Reliable and secure operation of modern power systems, the so called smart grids vision, is an increasingly challenging task due to several reasons: exponential penetration of variable renewable energy sources (RES), the ever-increasing demand for electricity, expansion and heterogeneity in terms of grid interconnections, deregulated energy market conditions that are revised continuously, among many others [1]. The greater challenge comes from the lack of sufficient observability and awareness on the dynamics of the power system especially bellow the MV level [2].

Energy communities and prosumers in general are leading the society efforts for the decarbonisation process of the energy industry [3]. This trend is sustained by ever increasing automation and modern measurement equipment deployed at prosumer level [1], [3]. Among the deployed intelligent sensors and metering equipment, smart metering infrastructure ensures the most popular source of data for applications targeting especially home energy management systems [4], for prediction models for load power profiles and/or power profiles for local generation [5], or to enhance the distribution system operator (DSO)'s awareness about the condition of the part of grid under its operation [6], [7], among many others.

Advanced data analytics methodologies and tools applied for RES assessment in distribution grids operation often involve the application of statistical techniques, predictive modelling, machine learning and deep learning, among others [8]. However, they rely on the availability of large and diverse datasets that can be processed and visualised, usually at centralized locations, in order to provide adequate information for improving or optimizing the system behaviour. Furthermore, besides availability of data, increased requirements for communication and computing resources are needed for selection and processing of this data such that to extract useful information [9], [10].

The application of data analytics in an engineering context for measurements in power systems can include time series data mining for pre-processing, forecasting and anomaly detection considering both statistical accuracy and computational efficiency [11]. This can be highly relevant in the case of increased reporting rates featured by modern measurement systems where the data acquisition and real time processing of the data streams should be tightly correlated in an adaptive manner [12]. A suitable illustration can be done through micro-transient regime classification where the algorithm should be able to label the incoming data within the control system sampling period to achieve closed loop behaviour [13].

By running the analysis in parallel for both generation, of a photovoltaic (PV) system in this case, and load-side power profiles, the goal is to highlight both the similarities, as common characteristics, and the complementary components

This work was supported by a grant of the Romanian Ministry of Education and Research, UEFISCDI no. PN-III-P4-ID-PCE-2020-2876 "Advanced Measurement Framework for Emerging Electric Power Systems" (EMERGE).

that can be exploited in a joint analysis framework [**?**], [14]. To be noted that so far the literature investigating anomaly detection in data time series for distribution grids focused either on the load [15] or on the generation [16] only, or the noise on the data was neglected. By emulating different levels of noise in the input data, the aim is to evaluate the robustness of a particular data analytics method under realistic assumptions that are tied to the measurement uncertainty of commercial grade energy smart meters. This is to be addressed in this work.

We have investigated the performance of deep learning, recurrent neural network, models for energy forecasting at various reporting rates in [17] under the paradigm of multi-scale analytics. The Matrix Profile (MP) [18] represents a general purpose and computationally efficient time series data mining technique which generates a companion time series of distance metrics that characterise a vector of values. A detalied description of the algorithm is presented in Section III. Investigation of MP for smart meter data information extraction has been previously discussed in [19]. In [20] we have applied this to residential active power measurements, to evaluate the effect of averaging and aggregation on the analytical results. Using labeled anomaly patterns for building energy data to infer the dominant usage was proposed in [21]. The contributions of the current work can be listed as two-fold:

- Parallel application of the Matrix Profile (MP) on complementary power profiles of (PV) generation and micro-loads for anomaly detection;
- Evaluating the robustness of the above-mentioned method under varying noise/uncertainty assumptions to quantify its robustness against the reported anomaly (discord) patterns.

The paper is subsequently structured as follows. Section II describes in detail the measurement context, including smart meter design for collection of load - and PV power profiles, that provides the quality datasets used in the analytics approach. Section III presents the matrix profile time series data mining technique for pre-processing and anomaly (discord) identification under varying noise assumptions. Results are illustrated in-depth in Section IV for the two use cases of choice: identification of anomalies and their distributions, as sensitivity analysis. Section V summarises the findings and presents the outlook for generalisation of the approach in future work.

II. Information Layer based on high-reporting rate measurements - Methods and Datasets

The information used in this paper is for load and generation models derived from extensive local measurements using Unbundled Smart Meters (USM) [22], set on 1 frame/s and 0.5 frames/s reporting rates, respectively. The USM consists of a Smart Metrology Meter (SMM) and a Smart Meter eXtension (SMX). SMM is the measurement part (metrology-tested) with real-time functionalities, while the SMX is able to extract, process and communicate the instrumentation values provided by the SMM, with additional versatile features to be implemented during the energy-meter lifetime towards future energy services and smart grid functionalities. In addition, the power profiles extracted were enhanced with synthetic white noise of an amplitude ranging from 2% to 5% of the recorded power values. The range was chosen in accordance with the accuracy class of the smart meters as to realistically consider the uncertainties in the measurement chain. The smart meters providing the information analyzed in this paper, are installed in the campus of University Politehnica of Bucharest. A key aspect for choosing the campus for the measurement campaign resorts to the students living in the buildings, as they are a relevant category of end-users. A simplified grid topology of the UPB campus with associated energy meters is highlighted in Figure 1.

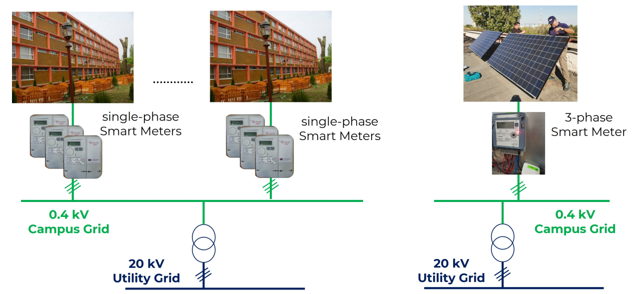


Fig. 1: Simplified UPB campus grid topology

The PV power profiles correspond to a roof-system active in the campus.

A. Highly-variable, high-time-granularity load power profiles

To derive the load profiles, the energy meters were installed in a 5-stories student building with 30 rooms and 60 students per floor. The metering infrastructure consists in 15 single-phase SLAMs (Smart Low-Cost Advance Meter) [23]. SLAM is the new generation of smart meters which takes advantages of new technologies in ICT and it is based on the smart meter Unbundled concept, differentiating the two modules (SMM and SMX). The SLAM meter is an advanced multi-function digital single-phase smart meter Class B in active energy and Class 2 in reactive energy, which complies with European legislation related to energy meters (MID) EN 50470-1 and EN 50470-3. It includes the SMX module that allows development of business related applications while allowing a multi user, multi- protocol communications with the grid actors. The communication of the smart meter with the exterior environment is done while preserving on the user side a strong control of data content and considering privacy and security aspects for data exchange to support data protection regulation rules. The micro-load on each floor is composed by specific profiles for office appliances such as personal computers (PCs), TV, refrigerator, air-conditioning, internet routers and lights. The granularity of the load power profiles is set at a 2s reporting time. An example of a daily load power profile, for one floor and one phase (extracted from one of the SLAMs) can be observed in Figure 2.

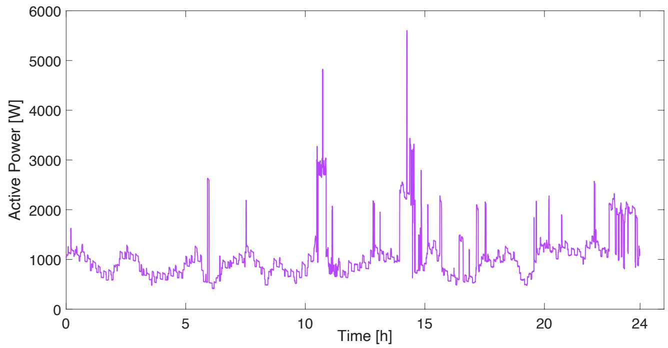


Fig. 2: Daily power profile of one floor and one-phase, extracted from one of the SLAM (2s resolution), in June

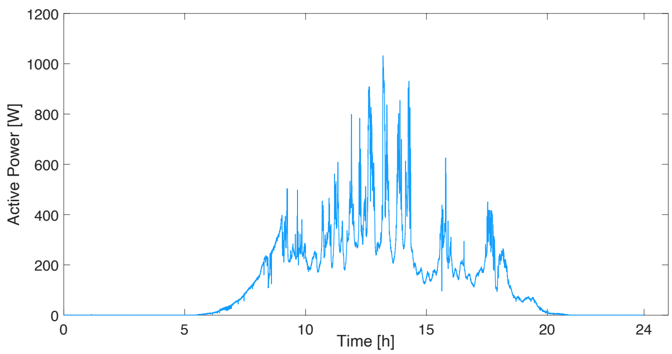


Fig. 3: Daily power profile of a PV in UPB campus, 1kW peak power (1s resolution), in May

B. Highly-variable PV power profile (1s-reporting time)

Local PV-generation has the power profiles based on daily real information extracted from a USM-reported measurements. The information is extracted from a high-reporting rate measurement equipment consisting of a SMX connected to a LandisGyr three phase meter [24]. with a metering accuracy Class B for active energy and Class 2 for reactive energy. The profiles were made available with 1 frame/s reporting rate, and correspond to UPB campus location in Bucharest. An example of a 1kW (peak power) PV system power profile with 1s resolution is depicted in Figure 3. Given the short distance between the 2 buildings, and similar shadowing, the solar irradiance was considered the same as the PV system was installed on the roof the student building we are studying the load power profiles.

III. METHODOLOGY

Matrix Profile (MP) technique for time series data mining was initially proposed in [18]. It is computed as a vector of values, obtained by sliding a window of size m over a time series T of size n. The subsequence length m represents the single parameter for tuning the method and can be derived either automatically, by using domain knowledge, e.g. periodicity of the underlying generative process or phenomenon, or through visual inspection of the input time series. Each value in the vector stores the minimum z-normalised Euclidean distance to its neighbors. The Euclidean distance d:

$$d\left(T_a, T_b\right) = \sqrt{\sum_{i=1}^{n} \left(T_{a,i} - T_{b,i}\right)^2} \tag{1}$$

where T_a and T_b are two subsequences of equal length and $T_{a,i}$ and $T_{b,i}$ are the i elements in the respective subsequence, leads to efficient computation. Z-normalisation provides comparative results across time series. For implementation one can choose between open software packages in python: matrixprofile and stumpy and also MATLAB and the Julia language. Each package offers a choice of computational algorithms for deriving the matrix profile, such as scrimp, scrimp++, stomp, mpx, can be used to compute it for both offline and online usage on streamed data.

For time-series T, two subsequences of length m, $\{T_{a,m}$, $T_{b,m}\}$ are considered a $motif$ pair [19] if:

$$dist(T_{a,m}, T_{b,m}) \leq dist(T_{i,m}, T_{j,m}), \tag{2}$$

$\forall i, j \in [1, 2, ..., n - m + 1]$ with $a \neq b$, $i \neq j$.

The subsequence with the maximum distance to its nearest non-self match neighbor can be interpreted as an unusual subsequence or anomaly and is denoted as a *discord*. Given the time-series subsequence $T_{c,m}$ of length m non-self matched with $T_{d,m}$ and the subsequence $T_{p,m}$, non-self matched with $T_{q,m}$, $T_{c,m}$ we label a *discord* if:

$$min(d(T_{c,m}, T_{d,m})) > min(d(T_{p,m}, T_{q,m})), \tag{3}$$

with $c \neq d$, $p \neq q$ and d a z-normalized Euclidean distance function.

To check the sensitivity of the method concerning the identification of anomalies (discords), we quantify its robustness by adding synthetic noise traces to the original input time series and evaluating the resulting profiles. This can be achieved using Monte Carlo-like simulations [25] where random-variability is inserted in the input according to predefined probability thresholds and their effect and ranges on the output is evaluated. In our case we apply Additive White Gaussian Noise (AWGN) to the original time series with zero mean and the standard deviation:

$$\sigma = \sqrt{1/n \sum \left(x_i - \bar{x}\right)^2} \tag{4}$$

with x_i the individual values, $\bar{x}$ the sample mean and n the sample size. The selection of Gaussian Distribution is motivated by the source of data: non-calibrated instrumentation values extracted from calibrated energy meters for a given accuracy class. Because no distribution of errors is available for the active power values, we considered the most favorable case for the SMM. The robustness for varying noise levels is evaluated based on descriptive statistics of the resulting profile vectors and through the positioning of the time series discords across the data. An additional pre-processing function that includes an 'add noise' parametrisation option is available natively in the matrix profile library but has not been used for this study given limited documentation available.

IV. Results

To illustrate our approach we first run multiple simulations on the micro-load and PV data described in Section II. For each daily dataset 100 runs were performed. Implementation was done using the Python programming language and suitable libraries in Colab [26] a cloud hosted environment for code notebooks. The code and data to replicate this analysis is available online [1]. The goal was to evaluate the stability of the top discord identification as the salient anomaly identified in the input data. We also quantify resulting profiles using statistical indicators. The parametrisation of the matrix profile computation is based on the optimal window size determined beforehand using the *mp.analyze* function.

The sample visualisations are presented in Figures 4 and 5, corresponding to the load and generation scenarios respectively. The figures both include multi-color line plots for each of the MP run results on the noisy data series. The red star markers pinpoint the top discord locations for each of the line series. It can be seen how, in the first case, the top discord position does not change significantly in the presence of noise in the input time series. This is pointing to a strong discord. For the case of PV data, there are two segments with high matrix profile values: at the end of the time series and around the middle; the added noise connects to a switch of the top discord in a few cases only.

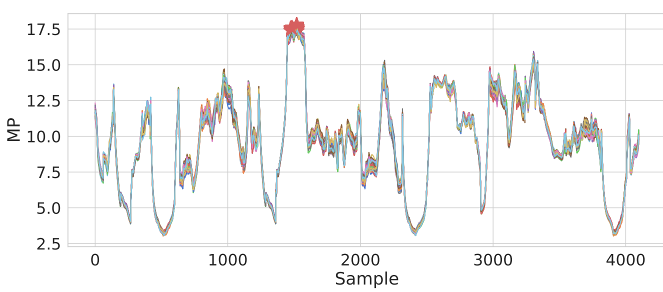


Fig. 4: Matrix profile variability with top discord identification - Load

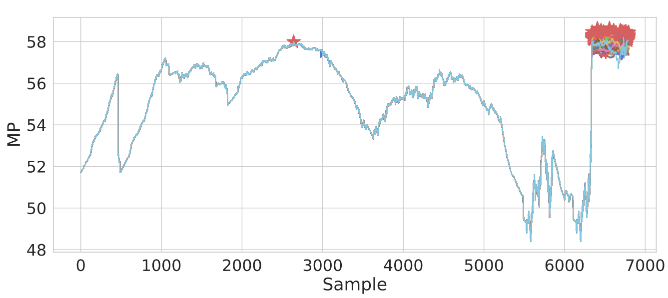


Fig. 5: Matrix profile variability with top discord identification - PV data

[1]https://github.com/grig101/amps22

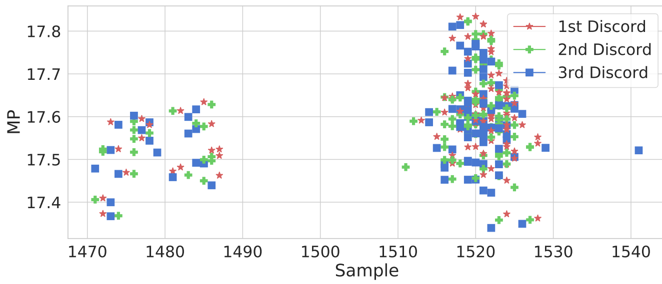
We subsequently report descriptive statistics for both the original matrix profiles and the averaged noisy profiles in both situations with the goal of reporting to what extent is the input noise replicated at the (MP) output. For the load data, given a lower analysis window (222 versus 1842 considered for the PV data), we observe a larger effect of the noise variability on the output. The quantitative metrics for the matrix profiles computed with this procedure are listed in Table I. The include the sample mean, minimum, maximum, standard deviation, as well as the 25% - 50% - 75 % quantile values.

TABLE I: Matrix profile statistics for load - and PV data

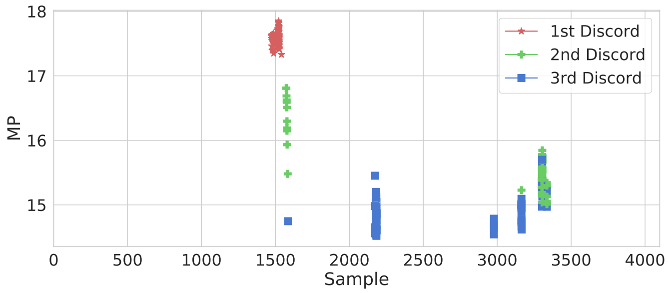
Statistic	Load MP	Load MP (noise)	PV MP	PV MP (noise)
Mean	9.19	9.56	55.02	55.06
Std	3.23	3.24	2.31	2.33
Min	3.05	3.22	48.39	48.39
25%	7.09	7.62	53.72	53.75
50%	9.33	9.81	55.61	55.65
75%	11.18	11.62	56.71	56.74
Max	17.6	17.55	58.07	58.06

For the anomaly sensitivity, we extend the analysis top three discords for both cases, illustrated in Figure 6. The software package also the setting of an exclusion zone, defined as the minimum sample distance to an identified discord, where a subsequent discord can be labeled. In the absence of the *exclusion_zone* parameter (Figure 6a), the discords tend to be grouped together in a certain area of the data: the algorithm identifies both the salient discord (the subsequence with the minimum distance to its nearest neighbor) and its trivial counterparts stemming from overlapping window subsequences. In this case all the top three discords appear in a narrow segment of the matrix profile around the area that was previously observed for the top discord in Figure 4. By defining the exclusion zone at the value of 100 for the load data we obtain the result in Figure 6b with a better delimitation of the first, second and third discord in different areas of the time series. This can be useful for discriminating secondary anomalies that might be hidden behind a dominant anomaly. The exclusion zone value can be defined based on the window size input parameter of the analysis.

Next, we apply various noise levels on the load and PV data according to the limits of errors associated with an assumed SMM accuracy class. The results are reported in Figures 7 and 8 and show the histograms of MP values for load 0% - 2% - 5% noise levels and PV data at 0% - 1% - 2% noise levels. The effects are mostly observed as a rightward shift of the distribution of the values, with larger effects for smaller analysis windows where the noise in the data can affect the z-normalised euclidean distance metric to a larger extent. The proportion of overlapping between the three empirical distributions can serve as a quantitative metrics in the data analytics evaluation pipeline. Higher noise levels are possible in order to challenge the stability of the MP algorithm. Also, other types of noise can be generated to assess the validity of alternative hypotheses.



(a) *exclusion_zone=None*



(b) *exclusion_zone=100*

Fig. 6: Distribution of the top three discords - Load

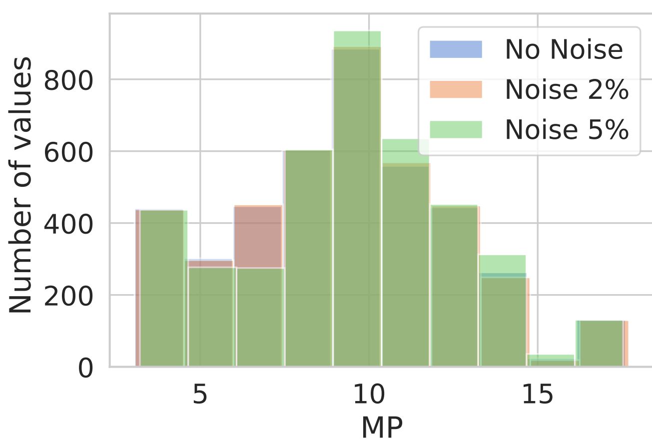


Fig. 7: MP distribution for varying noise levels - Load

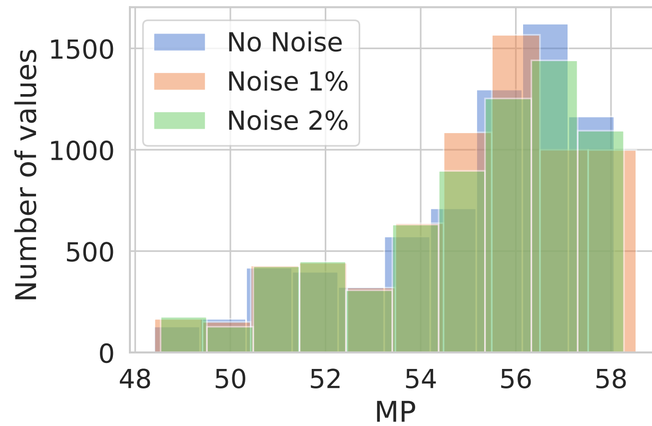
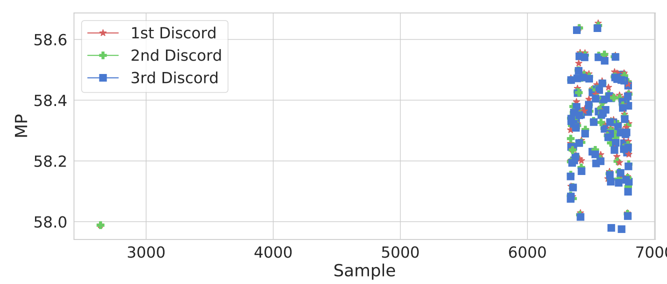
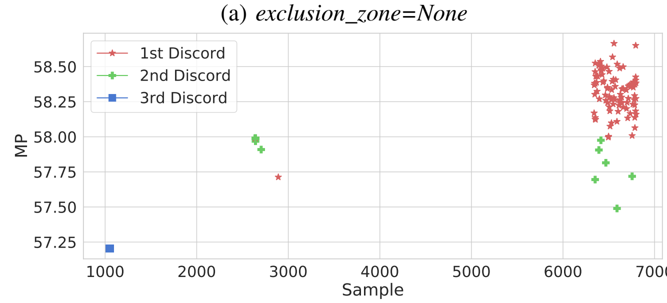


Fig. 8: MP distribution for varying noise levels - PV



(a) *exclusion_zone=None*



(b) *exclusion_zone=600*

Fig. 9: Distribution of the top three discords - PV

Finally the discord analysis results are replicated for the PV data in Figure 9. In this case the exclusion zone parameter is increased proportionally to the different window sizes to identify the non-trivial discord matches. The stability of the third discord (blue square) is emphasized across the multiple simulation runs. In this case the clustering of the top three discords can be better observed than in the previous case.

V. Conclusion

The article focuses on extracting anomaly patterns, in parallel for micro-load and PV data, from active power information extracted as instrumentation values from a SMX using a high reporting rate. The Matrix Profile (MP) time series data mining technique has been deployed together with various parametrisation options. The stability of the anomaly patterns was illustrated for varying underlying noise levels associated with assumed measurement quality. One main finding is that the method is robust to such input noise in the case of larger window sizes that mitigate the effect of the noise on the computed distance metric for the profile computation.

Future work will focus on extending and generalising the approach over multiple measurement databases, including aggregated net power profiles which are specific for prosumers. The potential for computationally efficient and real time embedded deployment into next generation smart meters will be investigated. This would involve the provision of real-time guarantees for the control loop execution in a hierarchical power system with increased uncertainty from renewable generation.

References

[1] Y. P. Agalgaonkar and D. J. Hammerstrom, "Evaluation of smart grid technologies employed for system reliability improvement: Pacific northwest smart grid demonstration experience," *IEEE Power and Energy Technology Systems Journal*, vol. 4, no. 2, pp. 24–31, 2017.

[2] Y. Yuan and Z. Wang, "Mining smart meter data to enhance distribution grid observability for behind-the-meter load control: Significantly improving system situational awareness and providing valuable insights," *IEEE Electrification Magazine*, vol. 9, no. 3, pp. 92–103, 2021.

[3] V. Caballero, D. Vernet, and A. Zaballos, "Social internet of energy—a new paradigm for demand side management," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9853–9867, 2019.

[4] H. Shareef, M. S. Ahmed, A. Mohamed, and E. Al Hassan, "Review on home energy management system considering demand responses, smart technologies, and intelligent controllers," *IEEE Access*, vol. 6, pp. 24 498–24 509, 2018.

[5] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125–3148, 2019.

[6] B. P. Hayes, J. K. Gruber, and M. Prodanovic, "A closed-loop state estimation tool for mv network monitoring and operation," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 2116–2125, 2015.

[7] M. Sănduleac, I. Ciornei, L. Toma, R. Plămnescu, A.-M. Dumitrescu, and M. M. Albu, "High reporting rate smart metering data for enhanced grid monitoring and services for energy communities," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 6, pp. 4039–4048, 2022.

[8] F. P. García Márquez, "Advanced analytics in renewable energy," *Energies*, vol. 15, no. 10, 2022. [Online]. Available: https://www.mdpi.com/1996-1073/15/10/3561

[9] A. Lang, Y. Wang, C. Feng, E. Stai, and G. Hug, "Data aggregation point placement for smart meters in the smart grid," *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 541–554, 2022.

[10] M. Ghorbanian, S. H. Dolatabadi, and P. Siano, "Big data issues in smart grids: A survey," *IEEE Systems Journal*, vol. 13, no. 4, pp. 4158–4168, 2019.

[11] R. Chiosa, M. S. Piscitelli, and A. Capozzoli, "A data analytics-based energy information system (eis) tool to perform meter-level anomaly detection and diagnosis in buildings," *Energies*, vol. 14, no. 1, 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/1/237

[12] K. R. Sahasranand, F. C. Joseph, H. Tyagi, G. Gurrala, and A. Joglekar, "Anomaly-aware adaptive sampling for electrical signal compression," *IEEE Transactions on Smart Grid*, vol. 13, no. 3, pp. 2185–2196, 2022.

[13] Y. Zhang, Y. Xu, Z. Y. Dong, and R. Zhang, "A hierarchical self-adaptive data-analytics method for real-time power system short-term voltage stability assessment," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 74–84, 2019.

[14] A. K. Marvasti, Y. Fu, S. DorMohammadi, and M. Rais-Rohani, "Optimal operation of active distribution grids: A system of systems framework," *IEEE Transactions on Smart Grid*, vol. 5, no. 3, pp. 1228–1237, 2014.

[15] D. Alahakoon and X. Yu, "Smart electricity meter data intelligence for future energy systems: A survey," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 1, pp. 425–436, 2016.

[16] Y. Seyedi, H. Karimi, and S. Grijalva, "Irregularity detection in output power of distributed energy resources using pmu data analytics in smart grids," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2222–2232, 2019.

[17] G. Stamatescu, I. Ciornei, R. Plamanescu, A.-M. Dumitrescu, and M. Albu, "Reporting interval impact on deep residential energy measurement prediction," in *2021 IEEE 11th International Workshop on Applied Measurements for Power Systems (AMPS)*, 2021, pp. 1–6.

[18] C.-C. M. Yeh, Y. Zhu, L. Ulanova, N. Begum, Y. Ding, A. Dau, D. Silva, A. Mueen, and E. Keogh, "Matrix profile i: All pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets," 12 2016, pp. 1317–1322.

[19] C. Nichiforov, I. Stancu, I. Stamatescu, and G. Stamatescu, "Information extraction approach for energy time series modelling," in *2020 24th International Conference on System Theory, Control and Computing (ICSTCC)*, 2020, pp. 886–891.

[20] G. Stamatescu, I. Ciornei, R. Plamanescu, A.-M. Dumitrescu, and M. Albu, "Multiscale data analytics for residential active power measurements through time series data mining," in *2022 IEEE International Energy Conference (ENERGYCON)*, 2022, pp. 1–6.

[21] C. Nichiforov, G. Stamatescu, I. Stamatescu, and I. Făgărăşan, "Learning dominant usage from anomaly patterns in building energy traces," in *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*, 2020, pp. 548–553.

[22] M. Sanduleac, L. Pons, G. Fiorentino, R. Pop, and M. Albu, "The unbundled smart meter concept in a synchro-scada framework," in *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, 2016, pp. 1–5.

[23] NOBEL GRID Project, "D4.1 specification of the unbundled smart meter concept based on commercially existing smart meters," 2015. [Online]. Available: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5a5d135e9&appId=PPGMS

[24] Landis+Gyr E550, "Transformer connected commercial electricity meter suitable for low and medium voltage applications," 2022. [Online]. Available: https://www.landisgyr.eu/product/landisgyr-e550/

[25] A. Ferrero and S. Salicone, "A monte carlo-like approach to uncertainty estimation in electric power quality measurements," *COMPEL-The international journal for computation and mathematics in electrical and electronic engineering*, vol. 23, no. 1, pp. 119–132, 2004.

[26] M. J. Nelson and A. K. Hoover, "Notes on using google colaboratory in ai education," in *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education*, ser. ITiCSE '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 533–534. [Online]. Available: https://doi.org/10.1145/3341525.3393997