Online Embedded Analytics for Energy Time Series Pre-Processing in LVDC Microgrids

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

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

***Abstract*—AI-enhanced applications in low voltage direct current (LVDC) microgrids offer the potential for online embedded load forecasting and anomaly detection with tangible impact on improved energy management and real-time control. The paper presents a time series data mining approach for LVDC microgrids which can serve both as a method for labelling micro-transient phenomena and events and as pre-processing technique to incorporate high reporting rate electrical measurement in subsequent advanced machine learning pipelines. We test the Matrix Profile algorithm on a reference dataset of PV active power measurements and discuss the obtained results for domain-specific parameter selection and validation and anomaly event detection. Results show useful performance for future benchmarking and *in situ* deployment on open field level hardware, such as smart meters and embedded development kits.**

***Index Terms*—data analytics, embedded ai, microgrid measurements, load forecasting, anomaly detection.**

I. Introduction and Scientific Context

The emergence and accelerated adoption of high performance measurement systems and devices in electrical grids leads to significant quantities of data being generated and handled by ICT systems of increasing complexity. In order to extract usable information from high-reporting rate data flows, leading to optimised operation of the energy systems, efficient algorithms have to be implemented and validated in realistic operation scenarios. General purpose data mining and machine learning techniques also require specialised tuning through domain expertise in order to provide credible and robust outputs for integration in predictive control structures.

Both artificial intelligence and machine learning tasks such as prediction, e.g. load forecasting [1], and classification, e.g. outlier detection by labelling of particular events such as faults or operation mode changes, can be implemented using readily available open software libraries on general purpose or specialised hardware. As an example, modern smart meters offer the necessary computing and communication resources that can enable online inference of AI algorithms and collaborative real-time assessment of the LV grid state for energy system and microgrid operators. Higher level models, such as federated learning frameworks [2] can aggregate lower level local models for distributed learning and global optimisation under various privacy and data or parameter leakage constraints.

In particular, new DC microgrid architectures integrate multiple energy sources such as renewable energy production from photovoltaic installations and hydrogen fuel cells, together with electrochemical energy storage in batteries, as well as a diverse range of DC native consumers at multiple voltage levels and AC consumers connected to the DC bus via DC-AC inverters. Well studied challenges for the design, operation and protection of DC microgrids include higher currents and low fault clearance times. Cooperation and standardisation initiatives such as Current/OS [3] aim to jointly advance a set of technical rules, while promoting the application of DC technology in various residential and industrial scenarios. Given low fault clearance times and ultrafast transients, modern current protection solutions rely on semiconductor devices for switching and current limiting tasks, as compared to conventional electro-mechanical devices. Several initiaives, such as the NOVETROL project, aim to design customised current limiter devices that are applicable to several types of practical use cases. Online real time observability and information extraction from DC microgrids is thus critical and can be realised using efficient data processing and learning algorithms, correlated with the (low) time constants of the dynamic process. This can be supported through increased availability of high-quality datasets that reflect varied modes of DC microgrid operation and are suitable for time series pre-processing, feature extraction and machine learning tasks.

The main contribution of our work is the comparative evaluation of the Matrix Profile time series data mining technique on real-world PV measurement datasets that aims to improve LVDC microgrid real-time control through real-time prediction and anomaly detection.

The remainder of the paper is structured as follows. Section II briefly discusses the state-of-the-art with regard to the application of intelligent algorithms to energy data analytics. Section III presents the main method used, the Matrix Profile algorithm along with the testing dataset, collected from an existing PV installation from our university. Results are illustrated in Section IV together with the main highlights regarding the practical insights gained from such an applications and the use of the pre-processed data in machine learning pipelines e.g. LSTM recurrent neural networks for prediction. Section V concludes the paper with observations for future work.

II. Related Work

The scientific context of our work is defined through a growing interest in data analytics for monitoring, control and optimisation, that represent both adaptations of existing methods and tools as well as newly developed approaches that consider the specific requirements and operational constraints of these new energy systems. In [4] power quality phenomena in DC microgrids are studied in comparison to AC occurrences and several indicators are defined along with simulation-based correlation analysis. This can form the basis for specific definition of ground truth labels and context for ai-based monitoring and control systems. A combined, AC and DC, machine learning based approach for fault detection is presented in [5]. The authors apply their method on a simulation case study with a 500kW grid-connected photovoltaic plant using windowed voltage sensor data as input features. The multi-class classification problem is solved through classical machine learning algorithms such as decision trees and random forests, logistic regression, kNN and support vector machines.

Related energy data applications of the Matrix Profile technique can be found in [6] where the authors applied the method on a public building energy consumption dataset. The extracted features are then fed to a secondary machine learning classification algorithm to assess the possibility of inferring the type of utilisation of the building from the raw energy measurements. Classification accuracy based on MP features is reported for various algorithms such as k-Nearest Neighbors (KNN), Support Vector Machines (SVM) and Artificial Neural Networks (ANN). In [7] the application of Matrix Profile is focused on power system measurements from the ERCOT dataset in Texas. The objective of the investigation is to carry out time series aggregation under the influence of renewable energy sources in the grid, wind power in this specific case. The cyclic patterns, in the form of time series motifs are identified and characterised in a way that can benefit grid energy management. The auhtors of [8] discuss a multi-dimensional approach for heterogeneous datasets in transmission networks with the goal of accurately modelling high variability renewable energy penetration of the grid. In this case, the Matrix Profile is presented as an alternative to complex architectures and parametrisation issues for computationally intensive deep learning time series models.

Previous own contributions have focused on data analytics methodology [9] and automated machine learning tools [10] for information extraction for energy time series datasets. In the former, the behaviour of the algorithm was studied under various parametrisation and data granularity assumptions for residential consumption power traces. The latter study uses specific anomaly detection and automated machine learning instruments at scale for similar residential unit data in order to obtain actionable results and practical value from the resulting predictions. We subsequently apply a similar approach on PV energy production data in order to leverage the insights for LVDC microgrid analytics.

III. Methods and Data

The Matrix Profile time series data mining technique was initially introduced in [11] and it produces a vector of values, obtained by sliding a window of size m, as sole parameter, over a time series T of size n. Each value in the vector stores the minimum z-normalised Euclidean distance d to its neighbours:

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

A time-series subsequence $T_{c,m}$ of length m, excluding self-matches, is denoted a *discord*, considered also as unusual subsequence or anomaly, if:

$$min(d(T_{c,m}, T_{d,m})) > min(d(T_{p,m}, T_{q,m})), \tag{2}$$

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

For identifying recurrent patterns in energy time series, mostly associated to periodic production trends or circadian activities on behalf of the LVDC microgrid end-users, in a time series T, two subsequences of length m, $\{T_{a,m},\ T_{b,m}\}$ are considered a $motif$ pair if:

$$dist(T_{a,m}, T_{b,m}) \leq dist(T_{i,m}, T_{j,m}), \tag{3}$$

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

New versions of the method such as Matrix Profile MADRID [12] and MOMP enable parameter-free and fast assessment of input time series for unsupervised real-time embedded deployment. Though the parameter-free approach can serve as an useful tool for fast deployment, we currently consider that the ability of selecting the input subsequence length in the classical MP method allows for useful domain knowledge tuning and adaptation of the expected results. This can be related to the various known effects, production and consumption patterns that are observed in the particular LVDC microgrid over longer time spans.

We apply this technique to a real-world, freely available, PV generation dataset with 1s reporting rate stemming from a PV plant installed on the roof of the Faculty of Electrical Engineering at the University Politehnica of Bucharest, in Bucharest, Romania, in a continental temperate climate area. The data is collected using a custom RaspberryPi based data logger from a grid connected PV campus plant during northern hemisphere spring over the course of one month in May 2020. The data logging routines perform periodic reading of the energy meter of the PV plant and store the main active power readings, together with the system time stamps in plain tab-delimited text files. Data quality issues such as missing values or erroneous readings are corrected, mainly through usual imputation methods e.g. by replacement with the last valid value or the average of surrounding values.

IV. Results

We use the recent MATLAB R2024b implementation of the matrix profile algorithm from the Predictive Maintenance Toolbox. Alternative open source implementations are available, mainly through the Python libraries *matrixprofile*, *matrixprofile-ts* and *stumpy*.

Figure 1 presents a sample day from our dataset, as the active power PV generation in Watts on May 1st, 2020. With a reporting rate of 1s, this results in 86400 active power values for the daily profile and associated timestamps, which can be selectively decimated in order to speed up the computational process during periods of low variability of the measurements.

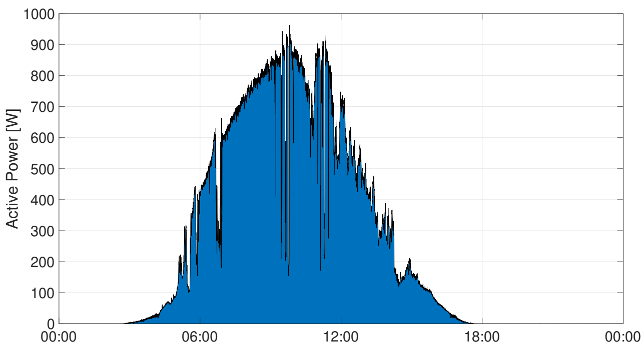

Fig. 1. Daily PV power profile

Figure 2 presents the actual computation of the matrix profile distance vector. Given the bell-shaped curve of the input power profile, we apply the MP algorithm by discarding the tails of the distribution in order to roughly consider only the sunlight period of the day. For our dataset this corresponds to the period roughly between 3am in the morning and 6pm in the afternoon and an index value from 1080 to 6840 in the data vector. The second subfigure represents the actual vector as minimum z-normalised Euclidean distance between the subsequence pairs. The subsequence length in this case is selected as one hour of data. Higher values in the profile indicate highly dissimilar subsequences, corresponding to discords, while lower values in the profile subsequences that are highly similar to other subsequences in the data, also denoted as time series motifs. The result also indicates the top discord of the time series, starting at index 2867, decimated by a factor of 10 in this case, which can be associated to a high variability of PV power generation in conjunction with local and daily weather conditions and clouding. The initial and final samples of the data are associated to recurring patterns, with lower variability which can also be observed on the original power profile.

Figure 3 presents a detailed illustration the top five discords, with the index pointer placed at the beginning of each subsequence, for the PV power profile, while considering a suitable exclusion zone to account for overlapping discord subsequences which tend to group together. This approach is helpful in order to classify and rank anomalous events and correlate them with additional contextual information in

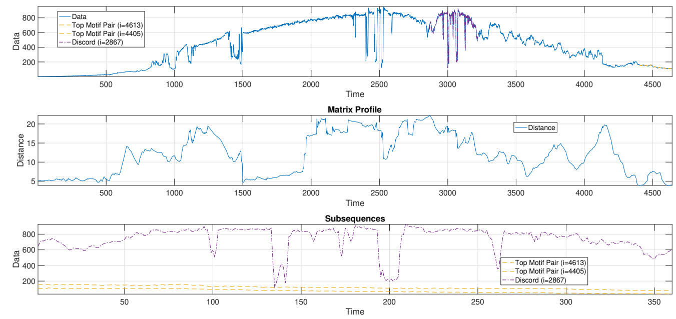

Fig. 2. Matrix profile daily PV power profile analysis

a machine learning approach. The top discord marked on this figure corresponds to the starting index of the discord sequenced marked in purple in the top subplot of Figure 2. More specifically, the automatically identified discords can serve as labels for a supervised learning algorithm such as neural networks or decision trees which associate the various local time series statistics indicators and derived features with the occurrence of an event that can impact the control algorithm behaviour at the LVDC microgrid level. This eventually leads to a probabilistic anticipation of anomalies and faults for correct operation of the microgrid.

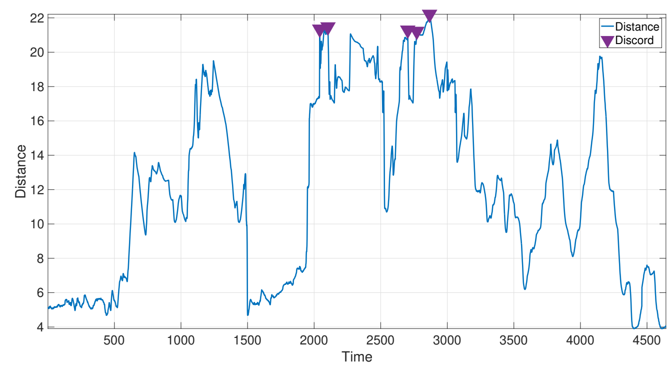

Fig. 3. Top five discords (anomaly subsequences)

For comparison purposes, we select a second day of data during the month of May 2020 and present in Figure 4 the results of the top three discord analysis. In this case we choose a coarser subsequence length of 720 samples, compared to 360 samples beforehand - approximately corresponding to a two hour sliding window on the decimated input time series. We observe how, in this case, the top anomaly/discord pattern is identified on the upwards trend of the PV panel production curve, around 7-8am in the morning. The trade-off for the selection of the sliding window size is represented by focusing either on micro-transient regimes at smaller time scales, which can also include noisy or perturbed measurements, or by aiming to identify major trends over the daily data that can be associated to weather patterns or other significant changes in the production context. For online operation a dynamic selection of the sliding window length can be performed thereby adjusting the granularity of the analysis to past or foreseen trends in the data series. The exclusion zone in this case has a value of 40 samples which allows to suitable differentiation of the discord samples. The absolute values of the distance metric are also higher in this case compared to the previous example.

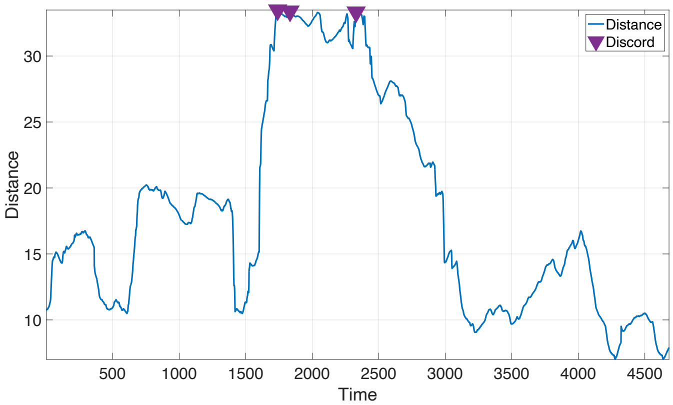

Fig. 4. Top three discords for alternative day with coarser subsequence length [d=720]

The resulting matrix profile vector, annotated with the respective discord indices, can be considered as the output of the data processing stage in a conventional machine learning pipeline for prediction (forecasting) or classification (anomaly detection) tasks. The already normalised MP vector along with the discord labels can be input in a recurrent neural network (RNN) architecture that natively handles temporal dependencies in multi-step training examples. In particular, using the Longs Short Term Memory units in RNN architectures has been shown to provide good forecasting results. Other options include 1D and hybrid Convolutional Neural Networks (CNN) models that have been initially designed for handling bidimensional image data but for which the input data can be reshaped in suitable formats for progressive feature extraction and prediction.

Systematic analysis of multiple PV power profiles over longer time periods can help derive robust estimates for the behaviour of the analysed energy subsystem under local conditions. Also, the impact of the parametrisation of the matrix profile method can be better understood at scale by adaptively adjusting the subsequence length based on automatic estimations combined with domain expertise for sanity checking. Expanding the analysis through automating of the data analysis routines over multiple data records for generalisable results can yield statistical relevant results for trend characterisation. Parallel, generation side and consumption side analysis, can also lead to mitigation of negative factors that affect optimised microgrid operation under uncertainty.

V. Conclusion

We present initial results for production-side analytics of PV energy data of LVDC microgrids. The method of choice, the Matrix Profile, has been used with good results in many applications domains for time series modelling and feature extraction, including energy data processing. Practical value of the work lays in the efficient computation time of the algorithm and various ready available implementation that support the deployment on suitable embedded platforms. Future work is concentrated on collecting of a high-quality dataset from a laboratory DC hybrid microgrid that will allow the data-driven characterisation of various operation modi using online and real-time data processing and learning algorithms. Combining MP features with machine learning methods can provide accurate predictions and detection of anomalies in stationary or transient regimes for improved energy management and protection of LVDC microgrids.

Acknowledgment

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101021936 (NOVETROL). This work was partially supported by the grants of the Ministry of Research, Innovation and Digitization, CNCS/CCCDI-UEFISCDI, project number COFUND-CETP-ECom4Future-1 and project number ERANET-ENERDIGIT-INFUSE, within PNCDI IV.

References

[1] Z. A. Khan and D. Jayaweera, "Smart meter data based load forecasting and demand side management in distribution networks with embedded pv systems," *IEEE Access*, vol. 8, pp. 2631–2644, 2020.
[2] X. Cheng, C. Li, and X. Liu, "A review of federated learning in energy systems," in *2022 IEEE IAS Industrial and Commercial Power System Asia (ICPS Asia)*, 2022, pp. 2089–2095.
[3] C. Foundation, "Current/os system public document," Version 2.1.0, Tech. Rep., 2024.
[4] J. Liao, Y. Liu, C. Guo, N. Zhou, Q. Wang, W. Kang, J. C. Vasquez, and J. M. Guerrero, "Power quality of dc microgrid: Index classification, definition, correlation analysis and cases study," *International Journal of Electrical Power and Energy Systems*, vol. 156, p. 109782, 2024.
[5] S. Naeiji, H. J. Ashtiani, A. Shabani, and Z. J. Shen, "Unified machine learning based fault detection strategy through voltage-sensing for both ac and dc side faults in photovoltaic farms," in *2024 IEEE Sixth International Conference on DC Microgrids (ICDCM)*, 2024, pp. 1–6.
[6] C. Nichiforov and M. Alamaniotis, "Learning matrix profile method for discord-based attribution of electricity consumption pattern behavior," *Cogent Engineering*, vol. 10, no. 1, p. 2199518, 2023.
[7] N. Sarajpoor, Z. Parvini, and A. Jahanbani, "Time series aggregation in power system studies in the presence of wind energy: A matrix-profile perspective," in *2023 North American Power Symposium (NAPS)*, 2023.
[8] T. Liu, L. Xie, H. Liu, R. Ye, H. Fan, and L. Ding, "Multidimensional matrix profile-based anomaly detection for power grid with high variable renewable energy penetration," in *2023 2nd Asia Power and Electrical Technology Conference (APET)*, 2023, pp. 605–609.
[9] G. Stamatescu, R. Plamanescu, I. Ciornei, and M. Albu, "Detection of anomalies in power profiles using data analytics," in *2022 IEEE 12th International Workshop on Applied Measurements for Power Systems (AMPS)*, 2022, pp. 1–6.
[10] G. Stamatescu, R. Plamanescu, and M. Albu, "Leveraging anomaly detection and automl for modelling residential measurement power traces," in *2023 IEEE 13th International Workshop on Applied Measurements for Power Systems (AMPS)*, 2023, pp. 1–5.
[11] C.-C. M. Yeh, Y. Zhu, L. Ulanova, N. Begum, Y. Ding, H. A. Dau, D. F. 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," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*, 2016, pp. 1317–1322.
[12] Y. Lu, T. V. A. Srinivas, T. Nakamura, M. Imamura, and E. Keogh, "Matrix profile xxx: Madrid: A hyper-anytime and parameter-free algorithm to find time series anomalies of all lengths," in *2023 IEEE International Conference on Data Mining (ICDM)*, 2023, pp. 1199–1204.