25th IMEKO TC4 International Symposium

23rd International Workshop on ADC and DAC Modelling and Testing

IMEKO TC-4 2022

Brescia, Italy / September 12-14, 2022

On Data Compression Algorithms for Power Measurements in Distributed Energy Systems

Paolo Bellagente[1]

[1] *University of Brescia, Department of Information Engineering Via Branze 38, 25123 Brescia, Italy, paolo.bellagente@unibs.it*

Abstract **–The diffusion of renewable energy sources is heavily affecting all aspects of modern power grids equipped by distributed energy resources and Battery Energy Storage Systems. Suitable Key Performance Indicators (KPI) have been developed, to find optimal storage capacity and control algorithms, which are resulted to be really dependent on data recording time interval. The aim of this paper is to analyze, on real data, which effect the time resolution of measurement data has on energy KPIs and to investigate how to select a proper compression algorithm to enhance the efficiency of the data collection process. Results shows that the KPI will be underestimated by the 40% if the sampling time rises from 5 s to 10 s, making lossy algorithms not interesting. Furthermore, lossless compression algorithms are effective but the tuning of those algorithms is all but intuitive.**

I. INTRODUCTION

The increasing penetration of distributed generation from Renewable Energy Sources (RESs) is heavily affecting the design and operation of modern distribution networks [1]. In particular, the intermittent nature of renewables, such as Photovoltaic (PV) systems, is calling for improved monitoring and control functions over Distributed Energy Resources (DERs) [2]. The uncertainty of distributed RES generation has indeed a relevant impact on the active power profiles of prosumers, i.e., end-users equipped with RES generators [3], [4]. To mitigate this drawback, the adoption of distributed battery storage systems has been recently proposed, with the aim of increasing the self-consumption of intermittent PV generation [5]. Nevertheless, the adoption of storage capabilities at prosumers' premises is only half of the story, and advanced DER control algorithms must be adopted to reduce the impact of the uncertain generation of RESs [6], [7]. The choice of the optimal storage capacity and of the most suitable control algorithm is generally carried out by referring to specific Key Performance Indicators (KPIs). These KPIs, however, strongly depend on the prosumer's generation and consumption profiles [8], and require power data gathered from heterogeneous measuring devices, such as smart meters and inverters. Data recorded by such devices, however, are usually stored in time-series databases as average values with time intervals ranging from 15 min to 1 h. Even though this temporal aggregation is typically used in power system analyses, recent works in the literature demonstrated the need for measurements with higher time resolution, and that the selection of the proper measurement time resolution is crucial for the assessment of the performance of RES installations [9]. In particular, the study published in [10] demonstrated that, if compared to the original time series sampled every 5 s, the adoption of the 15 min aggregation time interval could lead to relevant errors in the computation of some energy KPIs. Even though it is apparent that increasing the time resolution of measurements would surely help to increase the accuracy energy KPIs, the trade-off between the associated costs and benefits is not obvious, and further work should be done to assess the trade-off between data storage and transmission costs and KPI accuracy. On the other side of the advantages of the extensive use of meters for power grid management, there is the massive amount of data generated which should be transmitted to supervisory systems and often stored for further analysis. To reduce the transmission pressure and storage overhead, various compression techniques for smart meter big data, both lossless as lossy, have been proposed in literature [11] showing their effectiveness. The aim of this research work is to analyze in detail which effect the time resolution of measurement data has on energy KPIs on real data and to investigate how to select a proper compression algorithm to enhance the efficiency of the data collection process. The paper is organized as follow: section II introduces the reference use case on which analyzed data has been collected. Among all the KPIs used in literature section III contains the definitions of the one selected as reference for this study and how variations in the aggregation time interval between 5 s and 15 min affect them. Suitable reference compression algorithms are introduced in section IV, how they have been tested is described in section V. Test results are reported in section VI. Finally section VII contains some conclusions.

II. THE REFERENCE USE-CASE

The power flow data analyzed in this study have been collected by the supervisory system of the eLUX laboratory [12] on a public building of the University of Brescia, Italy which houses classrooms and offices as well as student services (i.e., dorms, reading rooms, a cafeteria). The facility, equipped with a 64 kWp PV plant and a 13.8 kWp/25.2 kWh $LiFePO_4$ battery energy storage system (BESS), represents a significant scenario as its power demand is similar to that of a big residential building. To exchange power with the grid, the BESS Power Conversion System (PCS) uses three single-phase bidirectional inverters with one configured as master device. Charge and discharge of batteries are controlled with requests t to the BESS PCS by means of a BESS Controller. All requests are forged on the base of the PV active power generation and the active power measurement at the Point of Common Coupling (PCC) between PV, BESS and the grid. Modbus TCPI/IP is used to let communicate the BESS Controller with the BESS Master Inverter, for charge-discharge control, and the two meters (PCC Meter and PV Meter) to measure the power flow. Further details about the system architecture are available in [13].

III. KEY PERFORMANCE INDICATORS

Different sets of KPI for power plant management have been proposed in literature for the optimal operation of BESSs (including residential PV-BESS installations). Among those, in [8], a system-independent set of KPIs able to measure the effects of the PV and BESS operation at the PCC with the main grid have been proposed to evaluate: i) the impact of distributed PV generators on the uncertainty of prosumers' net active power flows with the distribution grid; ii) the ability of BESS control strategies in reducing the power flow uncertainties caused by distributed PV systems; iii) the ability of a BESS rule-based control approach in reducing the power flow uncertainties introduced by a PV system.

A. Absolute Active Power Flow Ramp

Commercial meters collect data with their own sampling time, usually in the orders of seconds. DSOs use to consider time intervals of 15 min for grid control and monitoring. So a day d is divided into i time intervals $T_{i,d}$ each having k measured samples. Let $P_{i,d}^{U}$ be the Power exchanged with Utility in the i-th time slot of day d, positive taken, negative injected into. The average of the absolute ramp of the active net power flow at the PCC during the i-th time interval of day d, with the BESS in operation is reported in (1).

$$PR_{i,d}^{U} = \frac{|P_{i,d}^{U} - P_{i-1,d}^{U}|}{T_{i,d}}, i = 2 \dots n \qquad (1)$$

Let be, now $P_{i,d}^{BESS}$ the Power exchanged with storage in the i-th time slot of day d, positive charge, negative discarge. The Power exchanged with Utility in the i-th time slot of day d without the contribution of the BESS is reported in (2).

$$P_{i,d}^{U*} = P_{i,d}^{U} - P_{i,d}^{BESS} \qquad (2)$$

The average of the absolute ramp of the active net power flow at the PCC during the i-th time interval of day d, without the BESS in operation is reported in (3).

$$PR_{i,d}^{U*} = \frac{|P_{i,d}^{U*} - P_{i-1,d}^{U*}|}{T_{i,d}}, i = 2 \dots n \qquad (3)$$

B. Impact of Time resolution on KPIs

Using the data collected by the same system of section II, for 24 consecutive hours on April the 15th, 2020, authors in [10] demonstrated that, when evaluating the effect of BESSs in reducing the absolute power ramps in prosumers' PV-BESS installations, the classical sampling time resolution of 15 minutes would lead to misleading results which reduce the average values of power ramps by up to 20% respect to the same KPI calculated using a sampling time of 5 s. Lower sampling time shall improve the phenomenon representation but, at the same time, shall increase the amount of data to be collected, transmitted and stored. In the case studied in [10], moving from a 15 min sampling time to 5 s raises the data amount 180 times.

To better estimate this phenomenon and see if it is possible to find a suitable compromise between accuracy and data volume, in this work 5 s samples between April the 8th 2020 and April the 13th 2020 have been considered. These samples have been aggregated to simulate sampling time resolutions between 5 s and 15 m in steps of 5 s. For each simulated sampling time both PR^{U} and PR^{U*} have been calculated. Results are shown in Fig. 1. The estimated PR^{U} drops from 110 W/s to 65 W/s using only a 10 s sampling time. Results normalized by PR^{U} calculated with sampling time of 5 s are shown in Fig. 2. The value of PR^{U} is underestimated by 40% with a sampling time of 10 s reaching an underestimation of 80% with sampling time over 35 s. Following these results, to be accurately monitor the power flow at the PCC samples should be collected using sampling time of 5 s.

IV. COMPRESSION ALGORITHMS

As the number of available compression algorithms is high, in this work a couple of them have been chosen as reference. In particular, lossy algorithms have been ignored as, accordingly with section III. B, errors in the measure representation have a major impact on the KPI effectiveness. Among lossless algorithms the LZ family is one the most common as it is used by the 7-Zip archiver, a program publicly available under the terms of the GNU Lesser General Public License and commonly embedded

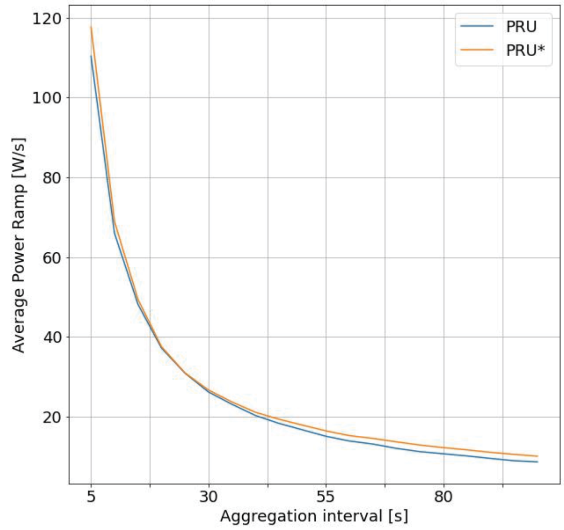

Fig. 1. Average Absolute Power Ramp.

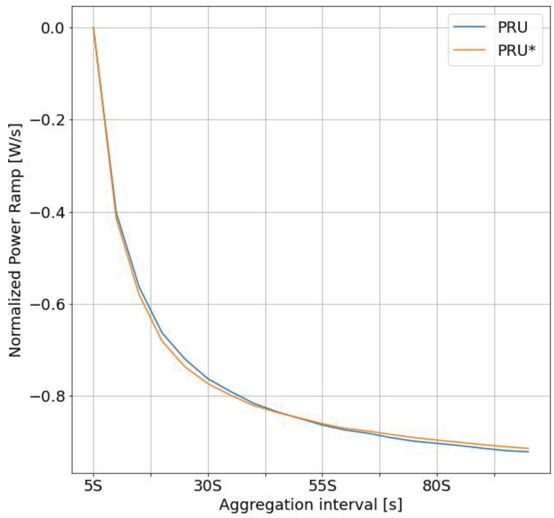

Fig. 2. Average Absolute Power Ramp normalized on its value in a 5s aggregation interval.

in many operating systems since 2009 [14]. Among the format supported by 7-Zip two interesting algorithms have been identified: i) DEFLATE: it is a combination of the LZ family progenitor LZ77 and Huffman coding. ii) LZMA: is a modified version of LZ77 already used in literature for smart meter data compression [15].

A. Huffman Coding and LZ77

The Huffman code maps each symbol to a variable number of bits. The number of bits depends on the frequency of occurrence of the considered symbol. The compression algorithm works creating a binary tree in which symbols result stored in leaves nodes. Links between nodes are labelled as '0' or '1' respectively for link connecting the left child or the right child. Following the tree allows to obtain the bit string representing the desired character [16]. LZ77 [17] compress a data stream mapping a repeated occurrences of data to a reference of the same data present earlier in the uncompressed stream. It stores a couple (length, distance) for each match. The stored match works as coordinates of the matched data from the position back to the actual data. The algorithm uses a sliding window of fixed, programmable, size to save the most recent data in which look for a new match.

B. DEFLATE Compressor and LZMA

DEFLATE creates a series of compressed blocks. Each block is composed by a 3 bit header, which defines the stream start/stop blocks and the encoding method, followed by a chunk of compressed data. The data is compressed in a two stage process first involving Huffing coding, giving the bit code for each symbol, and then LZ77 to further compress eventual duplicated series of bytes [18]. LZMA (Lempel-Ziv-Markov chain Algorithm) adds further compression steps to LZ77. In particular a LZ77 stage is followed by a Markov-Chain-based range encoder [19]. Both algorithms implementations can be tuned in run time by means of a specific parameter, which varies as integer number in the range 1-12 for DEFLATE and 0-9 for LZMA.

V. METHODS

To test the two compression algorithms, the system described in section III has been considered to collect data every 5 s while the control system needs them every 15 m as usual for DSO management. Thus, the measures of Active Power generated by the PV, power flow with the grid and power flow with the BESS, collected every 5 s between April the 8th 2020 and April the 13th 2020, have been aggregated in separate chunks of 15 min. Each chunk has been separately compressed and decompressed using both DEFLATE and LZMA algorithms with each value of their tuning parameters. For each run it has obtained the compression ratio in percentage, as defined in (4), and the compression speed expressed in MB/s.

$$C_r = \frac{CompresedSize}{OriginalSize} \cdot 100 \qquad (4)$$

All the tests have been performed using the lzbench in-memory benchmark of open-source LZ77/LZSS/LZMA compressors [20] using the provided official docker image. This approach has a big advantage of using the same compiler with the same optimizations for all compressors. The container has been hosted on a Dell Precision Tower 1700 MT Workstation equipped with quad-core intel Xeon E3-1220 v3 and 16GB RAM with CentOS 8 Stream. The container has been ran without the Docker daemon leveraging the deamonless, open-source, tool podman included in CentOS 8.

VI. RESULTS AND DISCUSSION

Results of all chunk of data, for each couple algorithm-parameter have been grouped. Fig. 3 and Fig. 4 show respectively mean compression ratio and the mean compression speed of all data chunks for each couple algorithm-parameter. Compression Ratio of chunks results to be between the 13,4% to the 19.7%. As saved space (SP) is defined as in (5):

$$SP = 1 - \frac{CompresedSize}{OriginalSize} = 1 - C_r \qquad (5)$$

Fig. 3 shows also SP between 86,6% and 80,3%. The compression speed lays within 0.78 MB/s and 82.5 MB/s. It is worth of noting, more than the absolute performance of the algorithms which could not be the best available, the role of the tuning parameter. It is clear from Fig. 3 that the algorithm with the best (lower) compression ratio depends on the tuning parameter as DEFLATE-11 compress more than LZMA-2 making difficult to choose the right algorithm without considering the tuning parameter. The tuning parameter have a huge effect also in the performance of the same algorithm as DEFLATE-8

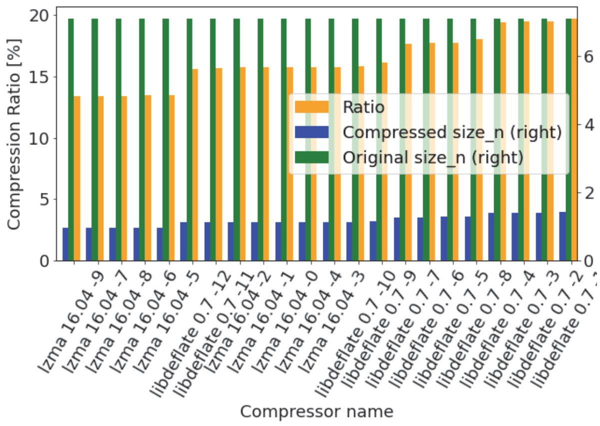

Fig. 3. Compression Ratio varying the algorithm compression parameter.

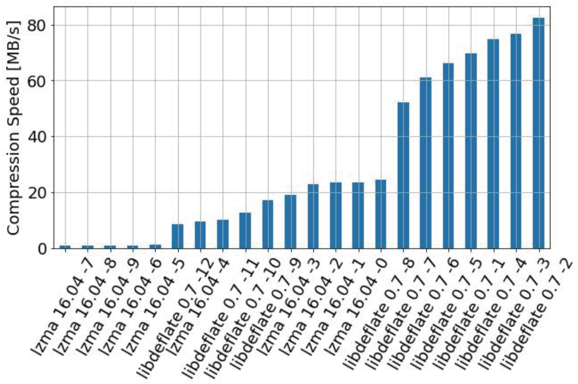

Fig. 4. Compression Speed varying the algorithm compression parameter.

compress less than DEFLATE-5 or LZMA-0 compress more than LZMA-4. A similar behavior could be observed also for Compression Speed in Fig. 4. The same results of Fig. 3 and Fig. 4 are reported together in Fig. 5 using circles to mark DEFLATE results while squares are used to mark LZMA results. The tuning parameter is represented by the color of the marker. It could be observed as LZMA obtains absolutes better results in compression ratio while DEFLATE has better absolute performance in compression speed. It is also evident as DEFLATE depends more on the tuning parameter as DEFLATE-5/6/7 compress more than DEFLATE-8. It is also interesting as LZMA with low parameter values performs really similar to DEFLATE with high parameter values

VII. CONCLUSIONS

The uncertainty of distributed RES generation on active power flows at the prosumer's side has a relevant impact on the operation of power grids and the classical sampling time resolution of 15 minutes would lead to misleading results. In this paper, using 6 days of data produced by a PV+BESS plant at the University of Brescia, Italy, the impact of the variation of the sampling time between 5 s and 15 min on the Absolute Average Power Rate has been evaluated. Results shows as the KPI will be underestimated by the 40% if the sampling time rises from 5 s to only 10 s, making necessary a 5 s sampling time and thus raising the overall amount of data generated. The use of lossless compression algorithm is a common proposed solution to try to reduce the data volume impact, but select the right algorithm could be difficult. For this reason two common lossless algorithms, DEFLATE and LZMA, have been selected and tested on the 6 days long, 5 s sampled dataset already used. Measures have been grouped in 15 min chunks as they would be transferred in a usual

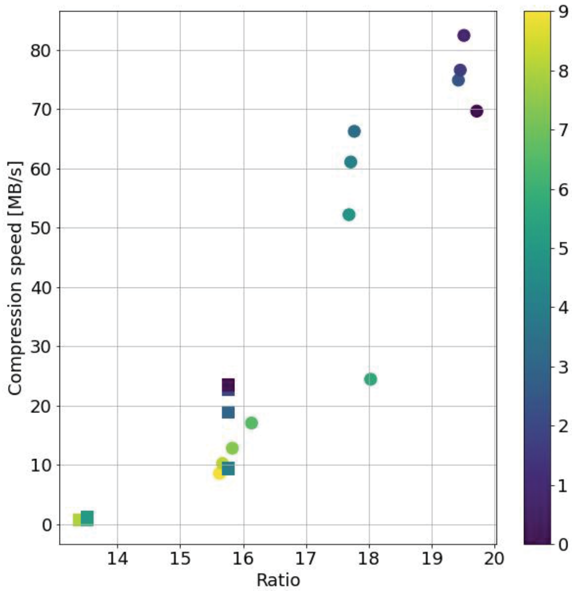

Fig. 5. Performance of couples algorithm-parameter. Circles are DEFLATE results while squares are LZMA results.

configurations. Results shows that as the compression algorithms are somehow effective as expected the tuning of the compression parameter is all but intuitive. This behavior makes try-and-error assessment or offline simulation needed to find the best compromise. In future works, has been already planned to experiment different algorithms aggregating data, sampled each 5 s, in chunk bigger than 15 m to find a suitable compromise between accuracy of KPIs and the overall size of the collected data.

ACKNOWLEDGEMENTS

Author should thanks Dr. Marco Pasetti for the valuable discussions about power grids.

REFERENCES

[1] S. Howell, Y. Rezgui, J. L. Hippolyte, B. Jayan, and H. Li, "Towards the next generation of smart grids: Semantic and holonic multi-agent management of distributed energy resources," *Renewable and Sustainable Energy Reviews*, vol. 77, no. September 2017, pp. 193–214, 2017, DOI: 10.1016/j.rser.2017.03.107.

[2] M. Pasetti, P. Ferrari, D. R. C. Silva, I. Silva, and E. Sisinni, "On the Use of LoRaWAN for the Monitoring and Control of Distributed Energy Resources in a Smart Campus," *Applied Sciences*, vol. 10, no. 320, pp. 1–27, 2020, DOI: 10.3390/app10010320.

[3] S. Liu, P. X. Liu, X. Wang, Z. Wang, and W. Meng, "Effects of correlated photovoltaic power and load uncertainties on grid-connected microgrid day-ahead scheduling," IET Generation, Transmission and Distribution, vol. 11, no. 14, pp. 3620–3627, 2017, DOI: 10.1049/iet-gtd.2017.0427.

[4] V. Sharma, S. M. Aziz, M. H. Haque, and T. Kauschke, "Effects of high solar photovoltaic penetration on distribution feeders and the economic impact," Renewable and Sustainable Energy Reviews, vol. 131, no. June, p. 110021, 2020, DOI: 10.1016/j.rser.2020.110021.

[5] D. Kucevic et al., "Standard battery energy storage system profiles: Analysis of various applications for stationary energy storage systems using a holistic simulation framework," Journal of Energy Storage, vol. 28, no. November 2019, p. 101077, 2020, DOI: 10.1016/j.est.2019.101077.

[6] M. Baumann, M. Weil, J. F. Peters, N. Chibeles-Martins, and A. B. Moniz, "A review of multi-criteria decision making approaches for evaluating energy storage systems for grid applications," Renewable and Sustainable Energy Reviews, vol. 107, no. January, pp. 516–534, 2019, DOI: 10.1016/j.rser.2019.02.016.

[7] C. Jankowiak, A. Zacharopoulos, C. Brandoni, P. Keatley, P. MacArtain, and N. Hewitt, "Assessing the benefits of decentralised residential batteries for load peak shaving," Journal of Energy Storage, vol. 32, p. 101779, 2020, DOI: 10.1016/j.est.2020.101779.

[8] M. Pasetti, "Assessing the Effectiveness of the Energy Storage Rule-Based Control in Reducing the Power Flow Uncertainties Caused by Distributed Photovoltaic Systems," *Energies*, vol. 14, no. 8, p. 2312, 2021, DOI: 10.3390/en14082312.

[9] V. Papadopoulos, J. Knockaert, C. Develder, and J. Desmet, "Investigating the need for real time measurements in industrial wind power systems combined with battery storage," Applied Energy, vol. 247, no. January, pp. 559–571, 2019, DOI: 10.1016/j.apenergy.2019.04.051.

[10] M. Pasetti *et al.*, "Impact of the Measurement Time Resolution on Energy Key Performance Indicators for Distributed Energy Resources: An Experimental Analysis," *AMPS 2021 - 2021 11th IEEE Int. Work. Appl. Meas. Power Syst. Proc.*, Sep. 2021.

[11] L. Wen, K. Zhou, S. Yang, and L. Li, "Compression of smart meter big data: A survey," *Renew. Sustain. Energy Rev.*, vol. 91, pp. 59–69, Aug. 2018.

[12] A. Flammini *et al.*, "A Living Lab and Testing Infrastructure for the Development of Innovative Smart Energy Solutions: the eLUX Laboratory of the University of Brescia," in *2018 AEIT International Annual Conference*, 2018, pp. 1–6.

[13] M. Pasetti, S. Rinaldi, and D. Manerba, "A Virtual Power Plant Architecture for the Demand-Side Management of Smart Prosumers," *Applied Sciences*, vol. 8, no. 3, p. 432, 2018, DOI: 10.3390/app8030432.

[14] I. Pavlov, "7-zip History." [Online]. Available: https://www.7-zip.org/history.txt. [Accessed: 22-Apr-2022].

[15] F. M. Dahunsi, O. A. Somefun, A. A. Ponnle, and K. B. Adedeji, "Compression Techniques of Electrical Energy Data for Load Monitoring: A Review," *Niger. J. Technol. Dev.*, vol. 18, no. 3, pp. 194–208, Nov. 2021.

[16] D. A. Huffman, "A Method for the Construction of Minimum-Redundancy Codes," *Proc. IRE*, vol. 40, no. 9, pp. 1098–1101, 1952.

[17] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Trans. Inf. Theory*, vol. 23, no. 3, pp. 337–343, May 1977.

[18] L. P. Deutsch, "DEFLATE Compressed Data Format Specification version 1.3," 1996. [Online]. Available: https://www.w3.org/Graphics/PNG/RFC-1951. [Accessed: 23-Apr-2022].

[19] I. Pavlov, "LZMA SDK." [Online]. Available: https://www.7-zip.org/sdk.html. [Accessed: 23-Apr-2022].

[20] "lzbench Github repository." [Online]. Available: https://github.com/dblalock/lzbench. [Accessed: 23-Apr-2022].