Statistical analysis of photovoltaic power plant production in second resolution

Jozef Novotný
Innovation Department
joint stock company sféra, a.s
Bratislava, Slovakia
jozef.novotny@sfera.sk

Abstract **- Nowadays, measured data at 15-minute resolution are used in the electric power industry mainly for commercial balancing or in various other business and technical processes. However, the temporal resolution of these data does not allow to study in detail the variability of generation, the rapid changes and fluctuations of the generated power of photovoltaic plants caused mainly by the passage of clouds and fog in a given location. However, the statistical analysis of measured high-frequency PV plant production data at a second resolution allows to investigate and analyze power output variations down to the level of transients.**

Keywords: high-frequency data; renewable energy; photovoltaic power plant output variability; statistical data analysis; high-frequency fluctuations; statistical software R

I. INTRODUCTION

The production of electricity from renewable energy sources has increased significantly over the last decade, mainly due to energy policy and environmental factors. In the conditions of the Slovak Republic, priority has been given to the construction of photovoltaic power plants, whose aggregate installed capacity reached 535 MW in 2020, according to the annual report of the short-term electricity market operator OKTE [1].

Renewable energy sources are generally known for their seasonality and fluctuations in performance on different time scales. Therefore, one of the biggest challenges is the technical management of these fluctuations. Power system analysis, modelling and research are in the vast majority of cases carried out on 15, 30 or 60 minute time scales. Therefore, this paper focuses on very short time resolution by analyzing high-frequency second-by-second PV generation data. The aim is to represent the importance of variability in modeling and simulations of power system operation, while providing a statistical basis for second-based analyses. Understanding the characteristics of renewable resources using high-frequency data is an important aspect in the design of power systems, for example microgrids in islanded operation, because the high variability of the generated power leads to voltage fluctuations as a local parameter and frequency fluctuations as a global parameter. For this reason, microgrids are challenging not only to control but also to size the electrical protections and to maintain the dynamic and static stability of their islanded operation.

In general, high-frequency fluctuations are minimally studied in academia, mainly because of the unavailability of such data and also because of the natural mutual damping of fluctuations that arises from the distribution of renewable resources in the geographic relief of the landscape. This distribution reduces variability, but does not eliminate it completely. Some research studies in the past have used various statistical techniques to transform from low frequency (common commercial measurements) to high frequency data or tools to generate synthetic high frequency data [2].

This paper presents numerical and graphical results of a time series analysis in the form of measured values of the active power output of a 16.2 kWp photovoltaic power plant connected to the electricity distribution network. The intention of the paper is to provide insight into the second order dynamics of PV generation and to justify its rationale in power system modelling and design.

II. HIGH-FREQUENCY DATA

The database of measured values contains high-frequency data in second resolution. All measured parameters are listed in Table 1. The time period from 1.7.2021 to 31.7.2021 of the total generated power *P* is analyzed. The day 4.7.2021, 5.7.2021 and 6.7.2021 are omitted from the analysis due to the malfunction of the metering system. In total there are 28 days analyzed. The location is Bratislava, Slovakia. The measurement system produces one text file in .txt format on a daily basis. Data processing and analysis is performed in the statistical software R. The total number of measured parameters is 19.

Table 1. Header of the processed .txt file

Parameter	**Label**
Timestamp	–
Voltage [V]	$V_{RMS\,1}$, $V_{RMS\,2}$, $V_{RMS\,3}$
Total Harmonic Voltage Distortion [%]	THD_U
Current [A]	I_1, I_2, I_3
Total Harmonic Current Distortion [%]	THD_I
Active power [W]	P_1, P_2, P_3, P
Power factor [–]	Cos φ_1, Cos φ_2, Cos φ_3, Cos φ
Electricity consumption [Wh]	E_{in}
Cumulative electricity produced [Wh]	E_{out}

The photovoltaic power plant contains 69 polycrystalline panels with an output of 235 Wp in three strings. The analysis is based on a comparison of the two days with the lowest and highest variability of production.

III. Methodology for selecting the two days with the lowest and highest production variability and their analysis

The selection methodology is based on quantifying the variability of the time series of power production using the sum of the absolute values of the first differences according to the formula

$$VAR_{Pout} = \sum_{2}^{86400} \left| \frac{P_{(t)} - P_{(t-\Delta t)}}{3600} \right| [Wh] \quad (1)$$

$P_{(t)}$ is the power at time t, $P_{(t-1)}$ is the power at time t - 1. Both quantities are measured in watts. The differentials are divided by the number of seconds in an hour so that the given value corresponds to the energy expressed in Wh. The value 86 400 corresponds to the number of seconds per day. Since this is the first differential, the time step in this case is $\Delta t = 1$. From the above formula, it follows that the higher the magnitude of the differentials and the more frequent the step changes, the higher the variability becomes. Table 2 shows the VAR_{Pout} for the days with the lowest and highest variability of electricity generation.

Table 2. Days with the lowest and highest production variability

Day	VAR_{Pout} [Wh]
16.7.2021	126.33
14.7.2021	814,56

For comparison, Figure 1 shows the pattern of second values of power production on the days with the lowest and days with the highest variability. A significant difference in production variability is evident. The timeline itself is expressed in hourly resolution for clarity.

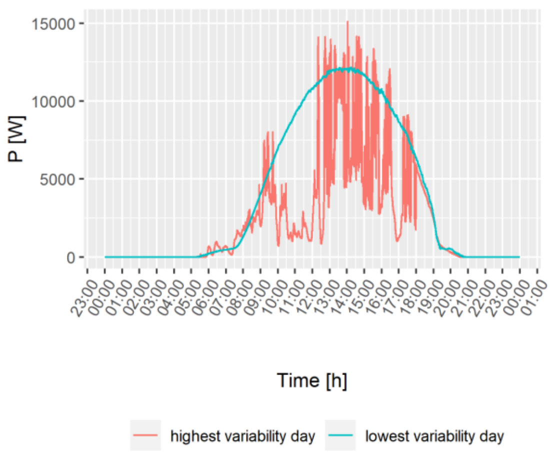

Figure 1. Profile comparison of days with the lowest and highest variability

Since the most commonly used data in analytical and simulation practice is 15-minute time resolution data, which involves expressing one quarter-hour with a single numerical value, Figure 2 is derived from Figure 1 with the time period from 14:00 to 14:15. Figure 2 thus provides a detailed view of the variability of production within a quarter-hour. The time axis in this case is expressed in minute resolution.

For interest, Table 3 shows the statistical characteristics of the selected quarter-hour. Means, medians and standard deviations are given to describe the width of the distribution of values in the set of measured data. Conventional measurement systems calculate the average value within quarter-hourly readings at predefined time slices within a quarter-hour. The resulting measured value is the average of these values. Thus, information about the intra-quarter-hour variability is not retained.

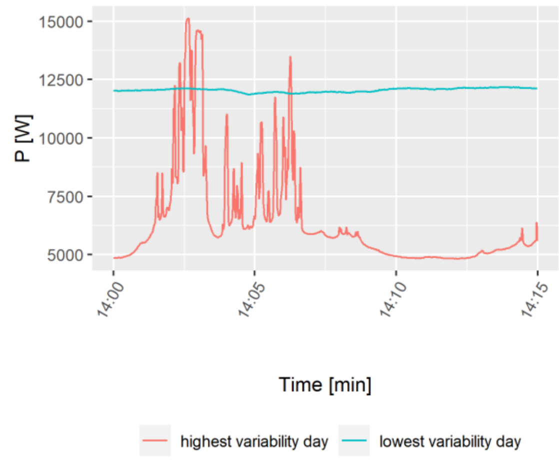

Figure 2. Time zoom of the selected quarter-hour in second resolution

Table 3. Statistical characteristics of quarter-hours for selected days

	The day with the lowest variability	**The day with the highest variability**
Minimum value	11 860	4 824
Median	12 071	5 722
Mean	12 051	6 398
Maximum value	12 184	15 130
Standard deviation	82,02	2 214,23

It is also useful to visualize the intra-quarter-hour power variability using a histogram, which interprets the measured power values by indicating the number of measured values that lie within a certain range. The higher the frequency of values within a given range, the higher the bar in the histogram. Figure 3 visualizes the histogram for selected quarters. The figure shows a significant difference in the distribution of values between the high and low variability quarters. The quarter-hour with low variability is represented by one high bar and one very low bar. Hence, in this quarter-hour, the concentration of measured values oscillating very close to the mean value of 12 051 W, which is shown in Table 2. On the contrary, at high variability, a concentration of values is evident especially in the range from 5 000 to 7 500 W. Values above 7 500 W are also present, but their frequency is low. The histogram also gives a rough indication of the underlying probability distribution of the values.

The table also shows another indicator of variability, namely the percentage difference between the mean and the median. Both characteristics belong to measures of central tendency. A quarter with low variability has a percentage difference of 0.17%, while a quarter with high variability has 11.81%. The difference can be explained by the non-uniform asymmetric distribution of the high variability values, with the mean, unlike the median, as an indicator of the central value being sensitive to outliers.

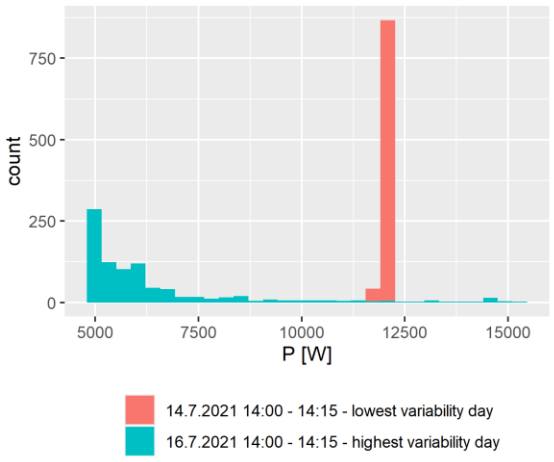

Figure 3. Visualization of measured second data in selected quarters by histogram

IV. STATISTICAL ANALYSIS OF FLUCTUATIONS

The power fluctuation $\Delta P_{\Delta t}(t)$ is defined at time t for a given time resolution period. The parameter Δt is the difference between two consecutive values of the produced power $P(t)$. The values are converted to the installed power P* [3]:

$$\Delta P_{\Delta t}(t)=\frac{[P_{(t)}-P_{(t-\Delta t)}]}{P^{*}}.100\ [\%] \quad (2)$$

The statistical analysis is performed for the measured values of 02.07.2021, when the largest positive and negative change in the differenced time series occurred simultaneously for $\Delta t = 1$. Figure 4 shows the high variability and irregularity of the produced power.

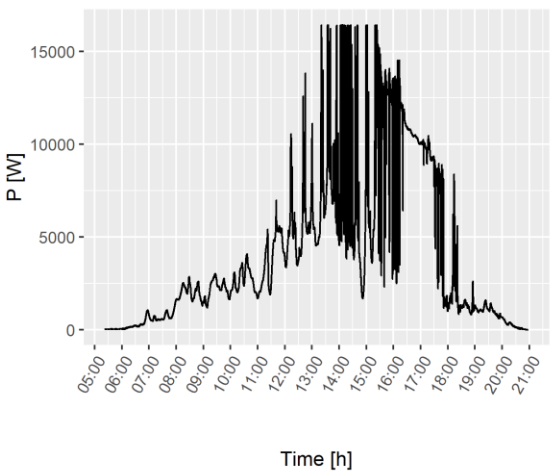

Figure 4. Visualisation of the active power for the day 02.07.2021

The percentage changes in performance according to (2) are also shown graphically in the time domain in Figure 5. Figure 6 shows a histogram of the percentage changes. The values when production was zero are removed from the analysis. The histogram clearly declares that the largest concentration of percentage changes is around zero. This fact is also expressed in Table 4, which shows that the maximum value of the negative fluctuation and therefore the decrease in output was more than 50% between two consecutive seconds. The increase in this case was 31.49%.

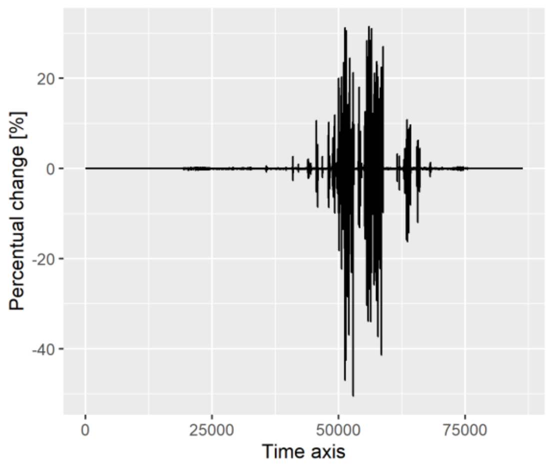

Figure 5. Visualisation of percentage changes in power on 02.07.2021 in the time domain

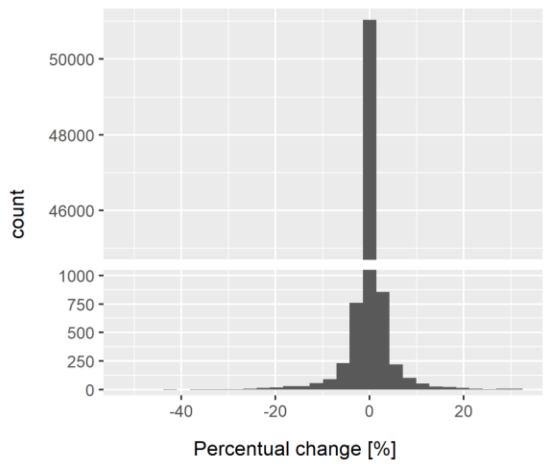

Figure 6. Visualisation of percentage changes in power on 02.07.2021 using histogram

Table 4. Statistical characteristics of fluctuations expressed as a percentage

	$\Delta P_{\Delta t}[\%]$
Minimum value	-50.50
Median	0.00
Mean	0.00
Maximum value	31.49
Standard deviation	1.59

Since the percentage of fluctuations is burdened by a dominant number of very low values oscillating around zero, all values falling in the interval (-5%, 5%) are removed for further analysis, thus creating two data sets with positive and negative values. The statistical characteristics of these data sets are expressed in Table 5.

Table 5. Statistical characteristics of the two datasets of fluctuations expressed as a percentage

	(min($\Delta P_{\Delta t}$),- 5 %)[%]	**(5 %, max($\Delta P_{\Delta t}$))[%]**
Minimum value	-50.50	5.01
Median	-8.62	8.34
Mean	-11.22	10.25
Maximum value	-5.02	31.46
Standard deviation	7.13	5.54

V. Statistical analysis of the rate of decline and increase in performance

The above analyses show that a step change in power drop can reach up to 50% of the installed power within one second, which in the case of microgrids or other isolated arrays with high PV penetration can have a negative impact on the quality of the delivered electricity.

From a power management perspective, it is therefore important to quantify the rate of decline and increase in power output given in MW/min in common practice, especially for distribution and transmission system operators. Since second values are available, this chapter works with W/s. The calculation is made according to [4]:

$$R_{RATEPV}=\frac{P_{(t)}-P_{(t-\Delta t)}}{\Delta t}\ [W/s] \quad (3)$$

Where R_{RATEPV} is the rate of change of power, $P_{(t)}$ is the power over time and Δt is the time step. The analysis is performed for a moving window with time difference $\Delta t = 1$, 2, 3 and 4, corresponding to time window widths of 2, 3, 4 and 5 seconds, respectively. Table 6 and Table 7 show the statistical characteristics for the power decrease and power increase, respectively. The values are given in W/s.

The tables show that as the width of the window used increases, the absolute value of the characteristics decreases and a smoothing effect occurs.

Both datasets for all time differences were tested for normality by the Shapiro-Wilk test. In practice, it is often assumed that the data come from a normal distribution. However, testing rejected the hypothesis that the data sets fit a normal distribution [5]. Hence, it is more appropriate to consider the median as a measure of central tendency. The tables also show a slight difference between the absolute values of the medians for the decrease and increase in power. The values suggest the conclusion that the decline in performance is generally faster. For example, for $\Delta t = 2$, the median decline is 10.13% faster than the increase. Based on this fact, the hypothesis was formulated that the difference in the rate of decline and increase is greater than 0 and hence statistically significant. Since the data did not come from a normal distribution, the Wilcoxon test was chosen for testing [6]. The test decides based on the calculated p-value whether the median difference between the two data sets is statistically significant or not. The results from the testing are presented in Table 8. In general, if the p-value is greater than 0.05, the hypothesis that the difference is not statistically significant is accepted. If the value is less, the alternative hypothesis is accepted and the difference of the two medians is statistically significant.

Table 6. Statistical characteristics of the decline in active power

	Δt = 1	**Δt = 2**	**Δt = 3**	**Δt = 4**
Minimum value	-8 182	-4 934	-3 635	-2 771
Median	-1 396	-1 413	-1 321	-1 272
Mean	1 818	-1 673	-1 506	-1 409
Maximum value	-813	-812	-810	-811
Standard deviation	1 155	847	641	497

Table 7. Statistical characteristics of the increase in active power

	Δt = 1	**Δt = 2**	**Δt = 3**	**Δt = 4**
Minimum value	812	813	812	811
Median	1 351	1 283	1 205	1 147
Mean	1 660	1 493	1 368	1 268
Maximum value	5 101	4 520	3 548	2 844
Standard deviation	898	677	517	405

Table 8. Evaluation of testing the statistical hypothesis of the difference in medians

time step	**p-value**	**conclusion**
Δt = 1	0.189	The difference is not statistically significant
Δt = 2	0.005	The difference is statistically significant
Δt = 3	0.014	The difference is not statistically significant
Δt = 4	0.000	The difference is statistically significant

VI. Conclusion

The analyses show a high degree of variability in the power produced by PV plants, which needs to be considered when designing microgrid control and understanding the dynamics of renewable energy sources. The maximum decrease between two consecutive seconds reached a magnitude of -50.50 %, which is 8 181 W, while the installed capacity is 16 200 W. The maximum increase in power between two consecutive seconds reached 31.49 %, which is 5 101 W.

Based on the median differences in the rate of decline and increase in power for different time windows, the hypothesis was formulated that there is a statistically significant difference between the rate and decline in generation. This was demonstrated at $\Delta t = 2$ and $\Delta t = 4$. Numerically, the rate of decline is generally slightly greater than the rate of power production ramp-up after the clouds depart.

ACKNOWLEDGMENT

This article has been produced thanks to the support under the Operational Programme Integrated Infrastructure for the project International Centre of Excellence for Research on Intelligent and Secure Information and Communication Technologies and Systems - Phase II, ITMS code: 313021W404, co-funded by the European Regional Development Fund.

REFERENCIE

[1] https://www.okte.sk/en/information/annual-reports/

[2] LAVE, Matthew, Matthew J. RENO a Robert J. BRODERICK. Characterizing local high-frequency solar variability and its impact to distribution studies. *Solar Energy* [online]. 2015, **118**, 327-337 [Date of citation: 2022-03-30]. ISSN 0038092X. doi:10.1016/j.solener.2015.05.028

[3] MARCOS, Javier, Iñigo DE LA PARRA, Miguel GARCÍA a Luis MARROYO. Control Strategies to Smooth Short-Term Power Fluctuations in Large Photovoltaic Plants Using Battery Storage Systems. *Energies* [online]. 2014, **7**(10), 6593-6619 [Date of citation: 2022-03-30]. ISSN 1996-1073. doi:10.3390/en7106593

[4] MALEKI, Sajad a Mehrdad TARAFDAR HAGH. Review of PV power ramp rate control methods and their requirements. *SVU-International Journal of Engineering Sciences and Applications* [online]. 2021, **2**(2), 14-25 [Date of citation: 2022-03-30]. ISSN 2735-4571. doi:10.21608/svusrc.2021.83458.1011

[5] Shapiro–Wilk Test in R Programming. *GeeksforGeeks* [online]. 16.7.2020 [Date of citation: 2022-03-31]. Web adress: https://www.geeksforgeeks.org/shapiro-wilk-test-in-r-programming/

[6] Unpaired Two-Samples Wilcoxon Test in R. *STHDA* [online]. [Date of citation: 2022-03-31]. Web adress: http://www.sthda.com/english/wiki/unpaired-two-samples-wilcoxon-test-in-r