УДК 004.62

DIGITAL TWIN IN THE ANALYSIS OF A BIG DATA

R.M. Asimov, PhD
CEO of Senso-tronica Ltd

S.V. Chernoshey
Engineer of Senso-tronica Ltd

I. Kruse,
CEO of SunSniffer GmbH

V.S. Osipovich, PhD
Associate professor of department of Human Engineering and Ergonomics

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus
E-mail: v.osipovich@bsuir.by

Abstract. The concept of monitoring of solar power plants` condition by creation of digital twins of solar panels including Photovoltaic Mathematical model, Big data analytics engine and Artificial Intelligence (AI) Knowledge engine is offered. PV Mathematical model is created and her primary approbation is carried out.

Keywords: Digital twin, Mathematical model, Photovoltaic, Analytics engine, Knowledge engine.

Solar power became an increasingly popular and economically viable energy source. More than 100 GW of solar power plants were put into operation in the world in 2017. At the same time, maintaining high efficiency of energy generation, identification and prevention became an increasingly important problem.

As it is shown at [1] approximately 2% of the solar panels are predicted to fail after 11-12 years. At the same time, the losses caused by accumulation of dust (soiling) can be an order of magnitude larger than losses due to cell degradation [2]. So, importance of determining required cleaning action cannot be underestimated. Growing vegetation can also cause losses due to shading. These and many other problems may be avoided or greatly reduced through early detection and proper maintenance actions.

Common approach to keep track on solar plant health is to install data acquisition system that monitor various parameters of Photovoltaic (PV) equipment. Typically this analysis is made top-down instead of bottom-up, which would be more useful as it is the modules which are most error-prone. The ultimate monitoring setup allows to collect data and control every solar panel [3, 4] as well as make this data available [5] for the analysis and forecasting.

However indirect nature of collected information in conjunction with permanently growing amount of time-based data (time series) poses great challenge for the analytics.

Typical data set describing operation of a solar panel includes: current, voltage, temperature, level of solar irradiation, date and time of measurement. When all these values polled at 5-minute intervals, just for the solar plants put into operation in 2017 it will generate 10 Gbyte daily.

Another factor that complicates the analysis is variability of weather conditions. While performance of the solar panels is specified for standard test conditions (STC) the real-world application shows dramatic variations of ambient conditions during the day. For example, solar irradiation heavily depends on atmospheric conditions like clouds, rain, snow or mist.

Considering all above, one can formulate two major problems to get business use from PV monitoring data:

– Collected data reflects system health only in indirect way (example: system delivers less power, but the reason is not directly clear) this shows one of the weaknesses of top-down models;

– Amount of data continuously growing over the time, making analysis more and more complex.

To overcome complexity of the analysis, concept of PV Digital twin was introduced. The Digital twin is a digital replica of real physical installation. It offers following main features:

– Check for monitoring data consistency;

– Perform data mining to detect existing and forecast upcoming problems;

– Use of AI Knowledge engine to support effective business decisions.

The core component of Digital twin is living mathematical model that acquire knowledge and learn from the monitoring data. The model converts virtually endless time series data in to a fixed set of the model variables (see Figure 1).

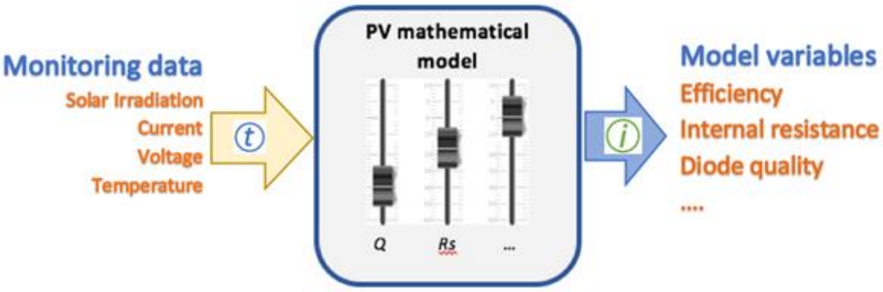

Figure 1. PV Mathematical model

Mathematical model calculates the voltage (V) values at the terminals of a particular solar panel based on the measured values of current (A), temperature (K) and solar radiation (W/m2). The learning process consists of tuning internal parameters of the model to the maximum convergence of the calculated and the measured voltage values. Once the tuning process is complete, internal parameters of the model (model variables) can be used to access efficiency of the solar panel. Figure 2 shows process of converting solar irradiation in to electricity that takes place inside the panel.

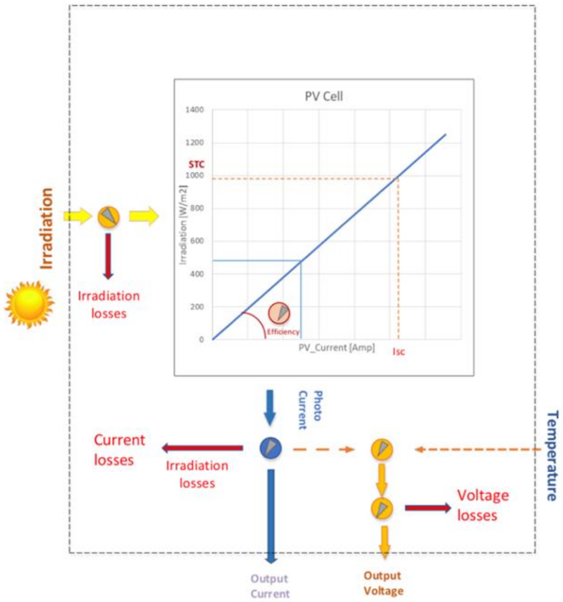

Figure 2. Model variables and assessment of solar panel efficiency

It should be noted, that the model variables (shown as circles) can be attributed to various losses that occurs during conversion process. This offers following benefits:

– model variables provide direct information about solar panel state and performance.

– use of the model variables instead of 'endless' time series from monitoring equipment, greatly reduces amount of data for further analysis.

Another application of the model is validation of monitoring data. The model can be applied to every measured data set in order to verify if such constellation of the values is physically possible for a device.

Another important component of the system is Big data analytics engine (see Figure 3). It performs classification task using values of model variables and provide useful indicators of the system health.

Figure 3. Big data analytics engine

The analytics engine uses SVM classifier technique to determine problematic solar cells and to assign their defect type. The outcome of data mining across the digital replica of PV installation is a list of problematic equipment together with health indicators for every its part.

In order to convert analytics results in to meaningful business decision, Digital twin use AI knowledge engine (see Figure 4). It applies knowledge base rules to the health indicators to create system status reports and make decisions about required maintenance actions.

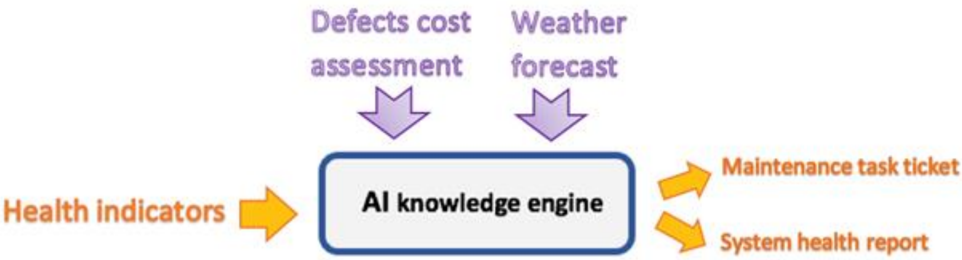

Figure 4. AI Knowledge engine

Here is summary of some key tasks that are performed by the engine:

– detect or forecast malfunctions and predict their economic consequences;

– estimate economic effectiveness and expediency of actions for elimination of the revealed problems;

– evaluate a payback period of advised actions;

– predict economic effect of preventive actions taking into account a weather forecast for the coming days;

– Evaluate quality of the carried-out maintenance activities.

At PV Digital twin system, all three components work together converting endless flow of monitoring data into set of the model variables which are transformed then into system health indicators and finally into human readable information to support effective business decisions.

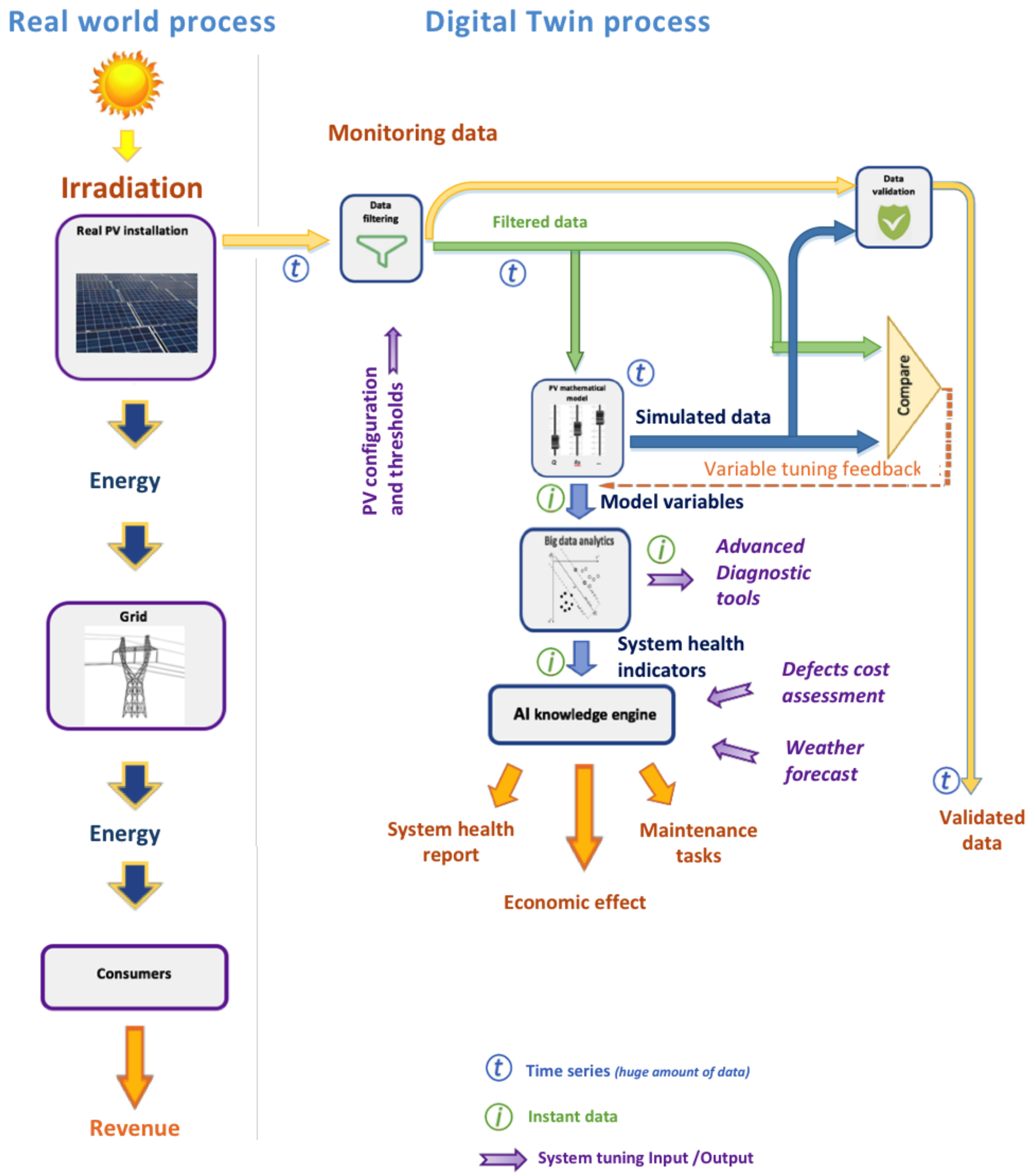

Figure 5. Digital twin data flow and processing

The block scheme of the system as a whole is submitted in the Figure 5.

1) The data collected during monitoring process pass through data reliability filter. The purpose of the filter is to extract the most reliable data that can be used for tweaking mathematical model. For example, all data points collected during weak or unstable irradiation condition will be dropped out.

2) Solar irradiation, current and temperature measurements are feed it to the model and module voltages are obtained as the result.

3) Simulated and Measured module voltages are compared, and tuning feedback supplied back to the model.

4) The process steps 2. and 3. Repeated several times until the difference between measured and simulated values became minimal.

5) Now the simulated data can be used to validate monitoring data through comparing them in data validation module. This module has two tasks:

–Verify if recorded data sample fits to the confidence limits obtained through simulation;

–Fill up the gaps in monitoring data if data samples are missed or found invalid.

6) The model variables tweaked through steps 2. and 3. are supplied then to Big data analytics engine. This engine performs data mining process searching for the system parts whose model variables shows abnormal or deviated values (health indicators).

7) The list of the such parts together with their health indicators supplied to AI knowledge engine. The engine applies its knowledge base rule sets in order to classify defects and analyze necessity and economic advisability of maintenance or repair tasks.

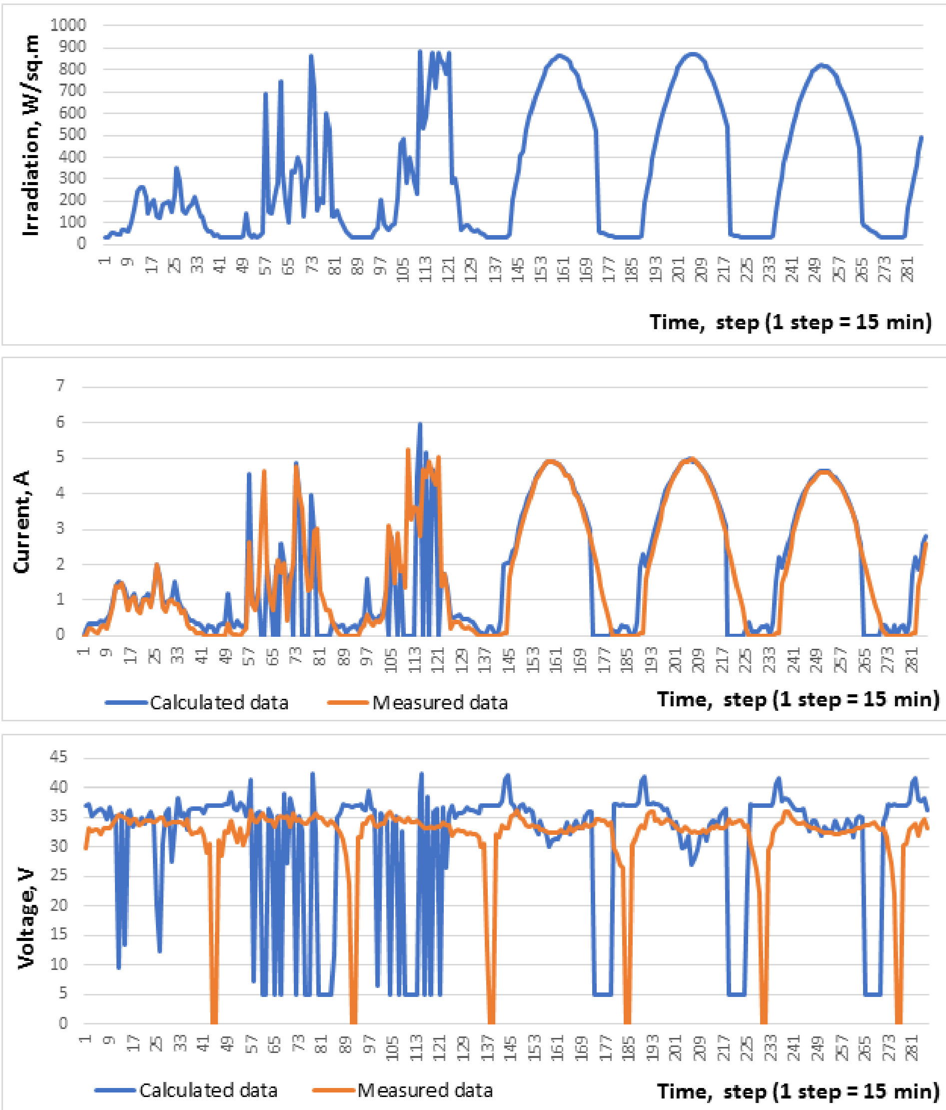

Figure 6. The modeling`s results in comparison with the measured values of current and voltage

At the first step of PV Digital twin development, feasibility study for living PV Mathematical Model was carried out. Initial approach was based on Single Diode Model described in [6] and sample data was recorded from real solar power plant. The figure 6 reflects results of the modeling.

By comparing curves of the real and calculated currents, one can see the mathematical models working capacity. Noticeable deviations (a mistake) of the calculated values from the real are clearly visible during morning and evening hours of the day. If to correlate the current curves to the curve of solar irradiation, it is possible to draw a conclusion that spatial orientation of the solar radiation`s sensor differs from orientation of the solar panel. Sharper recession of irradiation curve in the evening hints that solar sensor gets in to a shadow rapidly. At the same time the measured current is fading smoothly. This demonstrates that the sunlight continues to arrive on the panel. Such data constellation is inconsistent for the model and as a result the calculated current shows abrupt recession, while measured one has smooth slope in the evening. For the same reason there is a sharp recession to zero of the calculated voltage be end of the day.

These results demonstrate importance of preliminary filtration of acquired data that will be used for learning process of the model. It is clear that quality of model parameters tuning cannot be not better than quality of the data used for it. Besides this, improvement of model regarding better accuracy in voltages calculations needs to be done. Obtained mistake of 0,5-1,2V for voltage and about 0,02-0,1A for current calculations can be significantly reduced through use of filtered data and more precise mathematical description of the solar cell.

At the second step, two diodes model of the Solar cell was used (figure 7) [6] to model solar panel behavior. The software was implemented using JavaScript programming language and comprises several modules. Individual test and validation procedure was developed for every critical module in order to ensure consistent results.

Internal electrical parameters of solar panel where described by following model variables: diodes quality factor, (1 – perfect; bigger values – worse quality), resistance of parallel shunt R_p (inner isolation quality, bigger is better), series resistance R_s (contacts quality, smaller is better), was carried out on a data set has undergone preliminary filtration: current – 0–15 A, irradiation – 360–1500 W/m^2, derivative of current – -0,2–0,2, derivative of irradiation – -3,8–3,8, temperature – 25–400 K.

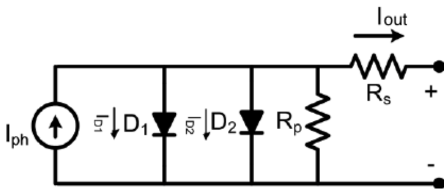

Figure 7. Two diodes model of the Solar cell

In addition to the electrical parameters the model uses set of correction factors applied to irradiation sensor readings. This is necessary to take in to account variations caused by geographical orientation, dust, shadow and other factors.

The learning process involves adjusting both electrical and irradiation parameters in a way to achieve maximum convergence between calculated and measured values.

For analysis we took week-long monitoring data for 12 modules connected in one string (fig 8). After processing the monitoring data, Digital Twin produced matrix with model parameters (table 1).

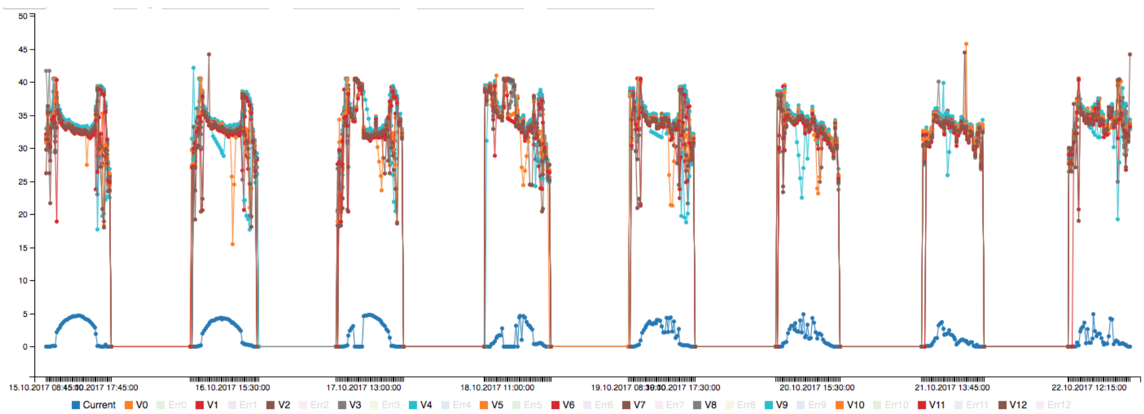

Figure 8. Voltages from 12 modules and string current (blue) during the week

Table 1

Digital twin output

PV panel number	Rp (Ohm)	Rs (Ohm)	Diode1 Qualty	Diode1 Qualty	Irradiation factor							
					10 AM	11 AM	12 AM	1 PM	2 PM	3 PM	4 PM	5 PM
0	999883	0,01	1,03	1,05	1,41	1,20	0,98	1,06	1,02	0,98	0,96	0,80
1	999883	0,01	1,03	1,03	1,68	1,20	0,98	1,03	1,00	0,98	0,95	0,75
2	999883	0,01	1,03	1,03	1,63	1,22	1,03	1,03	1,00	0,98	0,95	0,75
3	999971	0,01	1,05	1,05	1,44	1,22	0,97	1,02	1,00	0,96	0,96	0,50
4	999971	0,01	1,00	1,03	1,63	1,27	1,03	1,05	0,98	0,96	0,96	0,84
5	999971	0,01	1,03	1,05	1,52	1,19	1,00	1,05	0,98	0,98	0,95	0,79
6	999895	0,01	1,03	1,04	1,33	1,13	0,98	1,02	0,98	0,95	0,93	0,50
7	999895	0,01	1,03	1,04	1,39	1,14	0,95	1,00	0,98	0,97	0,94	0,70
8	999971	0,01	1,03	1,04	1,41	1,04	0,95	0,98	0,98	0,95	0,95	0,66
9	999883	0,01	1,05	1,05	1,34	1,05	0,98	1,04	0,98	0,96	0,93	0,53
10	999909	0,01	1,05	1,05	1,31	1,04	0,93	0,95	1,00	0,96	0,93	0,65
11	999883	0,01	1,05	1,05	1,49	1,11	0,94	0,97	0,98	0,96	0,96	0,58
12	**74**	**0,01**	**1,56**	**1,55**	1,47	1,25	0,98	1,08	1,08	1,08	1,10	0,77

Here each raw content model parameters for an individual solar panel that gives the best fit to the monitoring data. The table columns can be logically split on two groups: “Irradiation factors” and “electrical parameters”.

Irradiation factor describes correction that should be applied to irradiation sensor reading to get “useful” irradiation. Not all irradiation manages to penetrate inside the module, reach solar cell and produce current. Proportion of “useful” irradiation typically varied during the day, depending from module orientation, possible soiling and shadow conditions. This is why Digital twin uses set of the factors and each one applied for a particular hour of the day.

Electrical parameters describe quality of the processes that takes place inside Solar cells. The solar cell is described electrically as a circuit with source of photo current, two diodes and “parasite” resistors.

Let’s analyze separately different parameter groups from table 1. Irradiation factors shows similar behavior for all the panels over the day time. That indicates absence of shadow conditions. Also, for midday hours the value stays around 1 which means no irradiation losses from the surface (no soiling).

However, analysis of the electrical parameters gives different result. Module 12 have very low value or parallel shunt R_p and bigger values of Diodes quality. That can indicate high internal current loses and worse behavior of the solar cell.

Close look on Module 12 voltage in comparison to other modules reveals that it has significant voltage drops during morning and evening hours (fig. 9).

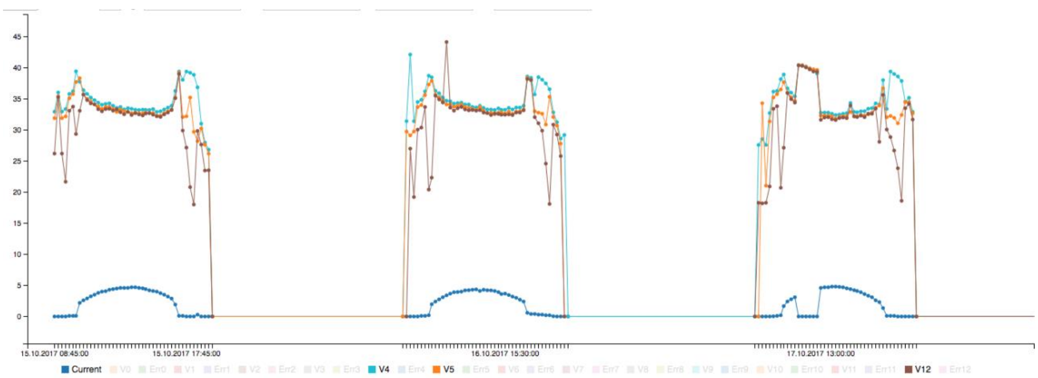

Figure 9. Voltage on Module 12 (brown), Module 4 (light blue) and Module 5 (orange)

That can be explained by combination of two factors.

1) Low R_p value induces high internal current losses which can be roughly estimated using formula:

$$I_{loss} = V_{mpp}/R_p\,, \tag{1}$$

where: V_{mpp} is maximum power point voltage.

For this type of the modules V_{mpp}=36,5 V. The Digital twin result (table 1) gives R_p=74 Ohm. I_{loss} = 0,495 Amp.

So, we can expect internal current loss about 0,5 Amp. This should certainly affect the module current-voltage characteristic (IV curve, see fig. 10). Let's estimate what would be the effect.

2) Additional current loss adds up with current drown by maximum power point module of PV inverter (Fig. 10).

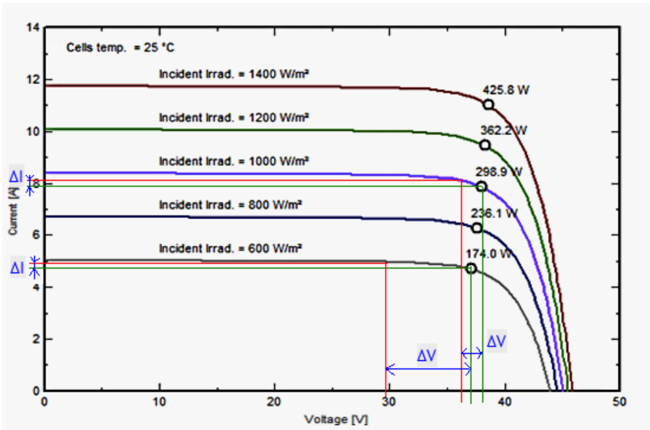

Figure 10. Example IV curves at different solar irradiation levels

What is interesting, at lower solar irradiation, IV curve shifts toward lower currents. And it can be clearly seen that same current increase will result on noticeable bigger voltage drop compare to high irradiation conditions. In other words, current increase through internal losses produces big voltage drops at low irradiation (morning and evening) and just very small voltage change when irradiation is high. This result correlates well with observations shown at Figure 9.

The Digital twin concept allows simple and direct method for data validation. The mathematical model of PV module can be also expressed as:

$$I_{ph} - I_{out} - I_{loss} = 0, \tag{2}$$

where: I_{ph} – photocurrent produced by solar irradiation, I_{out} – output current of the module (consumed by PV inverter), I_{loss} – current losses inside the module.

After Digital twin algorithm finds the model parameters they can be used to validate data recorded by monitoring system. The formula (2) fulfills Kirchhoff law that states: "sum of currents meeting at a point is zero". However, if we put in to this formula wrong data, result will obviously differ from zero. We can use this effect in order to validate every data sample recorder by the monitoring system. Let's rewrite (2) as:

$$I_{ph} - I_{out} - I_{loss} = E_{rror}, \tag{3}$$

For good samples we can expect Error → 0, but for wrong ones we obtain either Error>>0 or

Error<<0

Example of Error function indicating inconsistent data sample shown at Figure 11. The spike in Error function around 13:45, 21.10.2017 corresponds to voltage spike that does not fit current irradiation, voltage, current and temperature condition.

Figure 12 shows conditions with Error → 0 that indicates both valid data sample and quite good fit on the model parameters by the Digital twin. Small spikes in the morning possibly caused by misalignment in time of voltage and current readings. When inverter suddenly starts consuming current, correspondent voltage reading comes with delay caused by power line communication from the sensor.

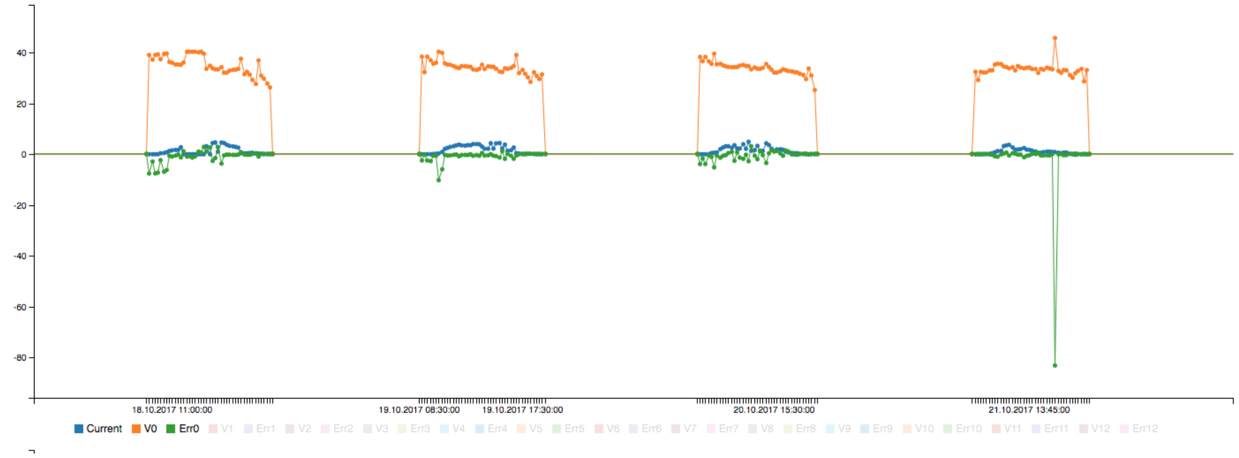

Figure 11. Error function (green), voltage (orange) and current (blue) Module 0

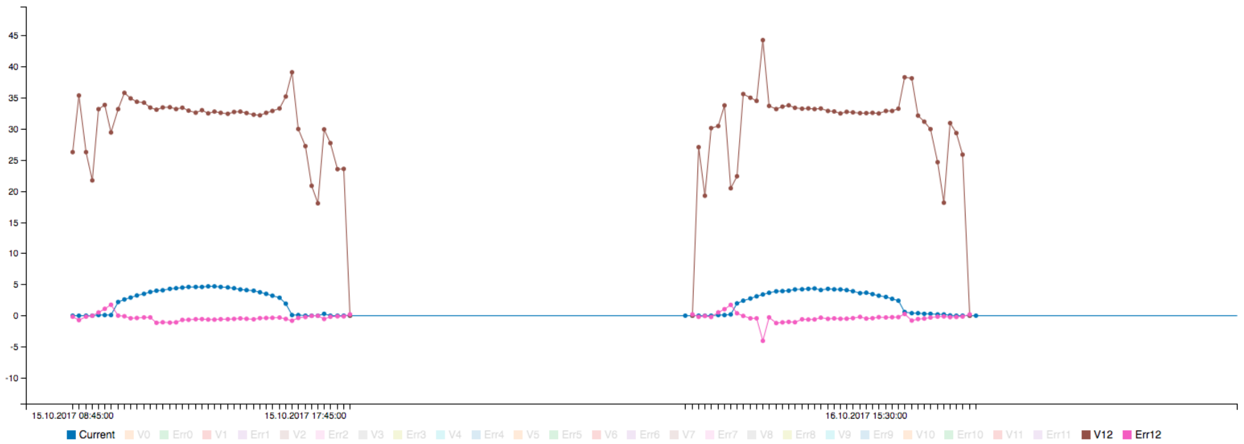

Figure 12. Error function (pink), voltage (brown) and current (blue) Module 12

So, the expediency of use in solar power engineering of the conceptual approaches offered by authors is proved in the theses. The essence of the concept consists in use of solar power plant`s Digital twin for analyzing efficiency of it work and search for faulty elements.

On the basis of the data collected by the SunSniffer company (Germany) approbation of the considered conceptual approaches in practice has been carried out. It was shown that living mathematical model of solar plant allows greatly reduce amount data required for analytics as well as provide direct indicators of the system health and performance.

Authors managed to confirm the efficiency of use of Digital Twin in searching for faulty solar panels.

References

[1]. Performance and Reliability of Photovoltaic Systems. IEA International Energy Agency, External final report IEA-PVPS March 2014, ISBN 978-3-906042-16-9, p. 5

[2]. F. Mejia et al. The effect of dust on solar photovoltaic systems / Energy Procedia 49 – 2014, p.2375

[3]. Method for monitoring individual photovoltaic modules in an arrangement that comprises several photovoltaic modules and device for performing said: pat. US20120197569 international G01K13/00, G01R19/00, G01R27/00, G06F15/00 / Ingmar Kruse, Roustam Asimov; Ingmar Kruse – US 13/379,319 at 28.06.2010; public 2.08.2012 //United States Patent and Trademark Office [Electronic resource] / USPTO Patent Full-Text and Image Database (WO 2011000505 A8, CN102473764A, CN102473764B, DE102009031839A1, DE102009031839B4, EP2449643A2, EP2449643B1, US9070281, US20120197569, WO2011000505A2, WO2011000505A3, WO2011000505A4).

[4]. Method for disconnecting a photovoltaic assembly and photovoltaic assembly: pat. US 20140311546 A1 international H01L31/18, H01L31/05 / Ingmar Kruse, Roustam Asimov; Ingmar Kruse – US 13/993,981 at 13.12.2011; public 23.10.2014 //United States Patent and Trademark Office [Electronic resource] / USPTO Patent Full-Text and Image Database (DE102010054354A1, EP2652857A1, WO2012079742A1).

[5]. SunSniffer [Electronic resource] / "Deep View" the functionality of each module is visible – 2017. – Mode of access http://www.sunsniffer.de/solution/what-is-sunsniffer.html. – Date of access: 5.02.2018

[6]. V. Tamrakar, S.C. Gupta, Y. Sawle. Single-diode PV cell modeling and study of characteristics of single and two-diode equivalent circuit // Electrical and Electronics Engineering: An International Journal. – Vol 4, No 3, – 2015, p. 13 – 24.