APPLICATION OF THE SAVITZKY-GOLAY METHOD FOR PV POWER OUTPUT DATA SET

SLAWOMIR GULKOWSKI

Lublin University of Technology
Nadbystrzycka 38, 20-618 Lublin, Poland

(received: 25 November 2014; revised: 10 January 2015; accepted: 15 January 2015; published online: 13 February 2015)

Abstract: The power output of a PV system changes in time during the day and strongly depends on the location and orientation of the photovoltaic module as well as on seasonal conditions. Clouds occurring during a partly cloudy day are the reason why this data is very irregular and difficult to analyze in terms of obtaining energy. The Savitzky-Golay method was applied for the power output data obtained for sunny, cloudy and partly cloudy days in order to determine the average level of power produced by a PV system at a given location. The total amount of energy was analyzed for each case.

Keywords: photovoltaics, solar energy, time series, PV data analysis

1. Introduction

One of the most important problems to solve in the 21st century is the problem of energy consumption which still increases. Fossil fuels are running out and they become expensive [1, 2]. It is estimated that fossil fuels will be used up in about fifty years. On the other hand, preservation of the environment during the process of energy production is crucial. For these reasons research for finding an alternative form of energy has become a very important issue. Promising alternatives are renewable energy sources such as solar, wind, water or geothermal energy [3].

One of the most direct methods of energy conversion is the photovoltaic effect used in solar systems which produces electricity from the energy of the sun [4]. Due to the absence of moving parts in the photovoltaic system the cost of maintenance is very small and the total investment cost is almost equal to the installation cost. Generating energy from photovoltaic sources is a common solution in Europe. The most popular material in the photovoltaic technology is silicon [5]. Unfortunately there is still an insufficient level of investments in solar farms or rooftop installations in Poland mainly due to the lack of support in this area [6] and the common conviction that there are not sufficiently many sun hours in this geographical location. However, compared with Germany, the level of solar insolation in Poland is similar. On the basis of this fact it is possible to conclude that there is a high potential in the development of photovoltaics in this region of Europe. Especially in the south-east region, where the research has been carried out and which is characterized by one of the highest insolation ratios compared to the rest of the country. For this reason it is necessary to find a useful tool to estimate the amount of energy which can be obtained in this region.

This paper presents the application of the Savitzky-Golay method to the power output data set obtained for different weather conditions. This solution leads to the possibility of predicting the amount of energy produced by the system for a particular location without the influence of changes of the local weather condition. The solution was implemented in the Matlab-Simulink environment.

2. The experimental set-up

The experimental photovoltaic system consists of a 100 Wp polycrystalline roof-mounted module, an optimal resistance and control unit to measure and register the PV data output. The module is faced south and it is connected to the load. The current and the voltage generated by the module were measured for 12 hours at every 30 s starting from 6 am. A more detailed description of the research system can be found in [7, 8]. The computer system registers the values of the current and voltage in the circuit. Then, power is obtained on the basis of Equation (1).

$$P\,[\mathrm{W}] = U\,[\mathrm{V}] \cdot I\,[\mathrm{A}] \tag{1}$$

The power produced by the PV module as a function of time depends on the technology of the module, orientation, location and it is very sensitive to the local weather conditions. For these reasons the output data generated from the system can be irregular and difficult to analyze. An example of the power output for sunny and cloudy days as well as for a sunny sky partially covered by clouds is shown in Figures 1–3, respectively.

As can be seen in Figure 1 the power plot for a sunny day changes regularly in time. It means that the weather conditions stayed unchanged during the whole day. A similar situation can be observed in Figure 2 for a totally cloudy day. However, these situations are rare in many locations. Most of the daily data outputs have irregular shapes such as presented in Figure 3. For this kind of output data, using the method presented in this paper, an average level of power generated by a PV module through the whole day can be estimated.

3. Savitzky-Golay method

The basic idea of employing an average method is to replace the data points by some kind of a local average of surrounding data points [9]. In this paper the

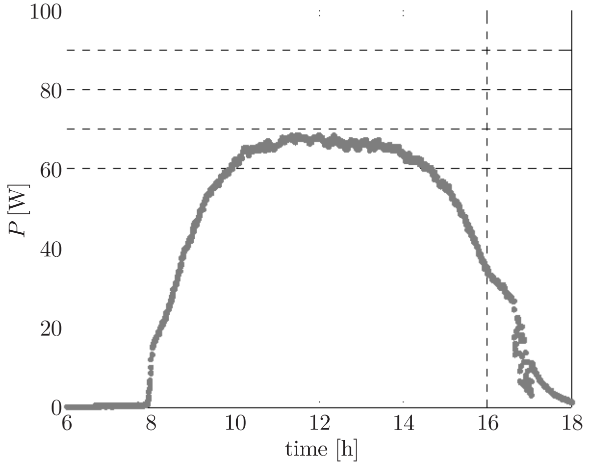

Figure 1. Measured power output data of PV module for cloudless day

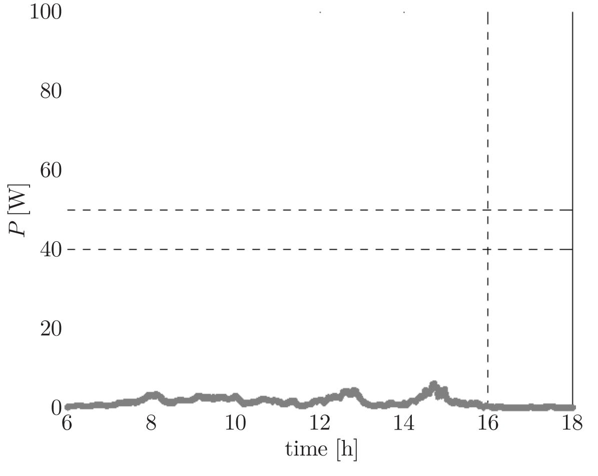

Figure 2. Measured power output data of PV module on clouded day

main goal of using the method is to obtain 1-h or 2-h average data from the PV system, especially in respect of the data which is presented as the example in Figure 3. It follows from the analysis that the application of this method allows predicting the average solar power output and energy produced by a PV system at a specific location.

The PV data output can be considered as a series of equally spaced data values in the function of time $f(t_i)$, where $t_i = t_{i-1} + \Delta t$ and Δt is the time interval between measurements. Following this method, each data value f_i is replaced by linear combination u_i of itself and a number of nearby neighbors:

$$u_i = \sum_{n=-k}^{n=k} a_n f_{i+n} \tag{2}$$

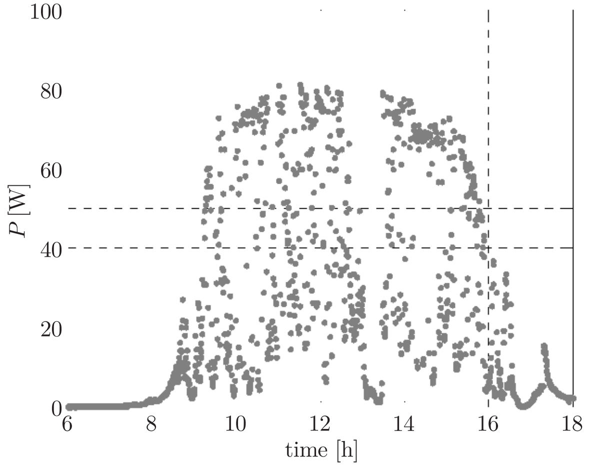

Figure 3. Measured power output data for sunny day partially covered by clouds

In Equation (2) k is the size of the window, which means the number of points used to the left and right of data point. The expression for calculating u_i can be written as:

$$u_i = \mathbf{a}f(i) = a_1 f_{i-k} + a_2 f_{i-k+1} + \ldots + a_{2k+1} f_{i+k} \tag{3}$$

It can be seen from Equation (3) that the idea of the Savitzky-Golay method is to find a vector of the coefficients $a_1, a_2, a_3, \ldots, a_{k+1}$. To derive such coefficients for each point, polynomial of l degree is fitted to all the $2k+1$ points in the moving window by means of the least-square method and then is set as the value of that polynomial at position i. The whole procedure is repeated when the window moves to the next data point f_{i+1} [9].

The set of coefficients can be calculated from the first row of the matrix given by the following equation [10]:

$$\mathbf{a} = -(A^T A)^{-1} A^T \tag{4}$$

where A is the matrix of dimension $(2k+1) \times (l+1)$:

$$A = -\begin{bmatrix} 1 & -k & (-k)^2 & \cdots & (-k)^l \\ 1 & -k+1 & (-k+1)^2 & \cdots & (-k+1)^l \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & k & (k)^2 & \cdots & (k1)^l \end{bmatrix} \tag{5}$$

It should be emphasized that the vector of coefficients $\mathbf{a}$ depends only on the polynomial degree l and the number k of the data points included inside the window. The values l and k should fulfill the following expression (5) [10]:

$$l < 2k+1 \tag{6}$$

The Savitzky-Golay method presented in this paper was implemented in Matlab. A listing of the Matlab script is given in Figure 4. Parameters k and l are taken as the input. The function returns a set of coefficients $\mathbf{a}$ as the output.

```
function [a]=moving_average(k,l)

%Calculation of matrix A:
for i=1:1:2*k+1
    for j=1:1:l+1
        if j==1
            A(i,j)=-1;
        else
            A(i,j)=A(i,j-1)*(i-k-1);
        end
    end
end

%Calculation vector of coefficients a:
AT=A';
ATA=AT*A;
invATA=inv(ATA);
a=-invATA*AT;

end
```

Figure 4. Matlab script file used to calculate vector of coefficients **a**

4. Results and discussion

The method described above was applied to the PV data output presented in Figures 1–3 with different values of parameters l and k. For each case the energy was calculated as an area under the fitted curve and compared to the total energy measured from the system. A relative error for the calculated energy was obtained from Equation (7):

$$\delta = \frac{E - E_{\mathrm{SG}}}{E} \cdot 100\,[\%] \tag{7}$$

where

E – the energy measured from the PV output data

E_{SG} – the energy calculated from the PV fitted data after using the Savitzky-Golay method.

Figures 5–8 show the measurements and estimated PV power after Savitzky-Golay fit into the data for a partly cloudy day. Calculations were performed out with parameter l equal to 1, and the k set as 30, 60, 90 and 120.

As can be seen from Figure 5 to Figure 8 the increasing number of points in the window (parameter k) leads to a more smooth curve fitted into the data. Single output data reach the level of 80 W. However, due to clouds which decrease the output power to almost at a zero level, the maximum peak of the fitted curve is between 50 W and 60 W. It depends on the number of points taken into account. The energy relative error for these parameters varies from 0.2 to 1.7 percent (see Table 1).

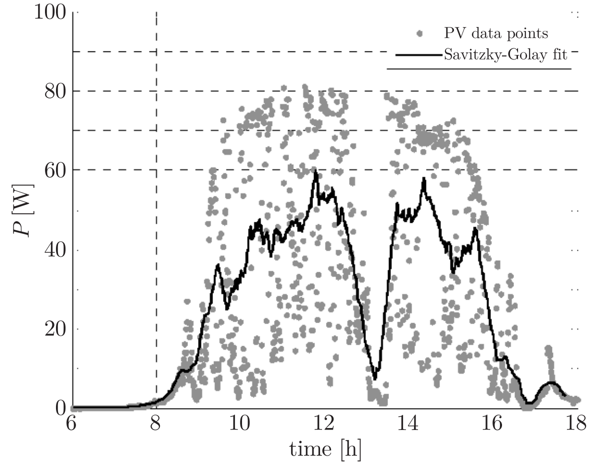

Figure 5. Measurements and estimated power PV output; $l=1$, $k=30$

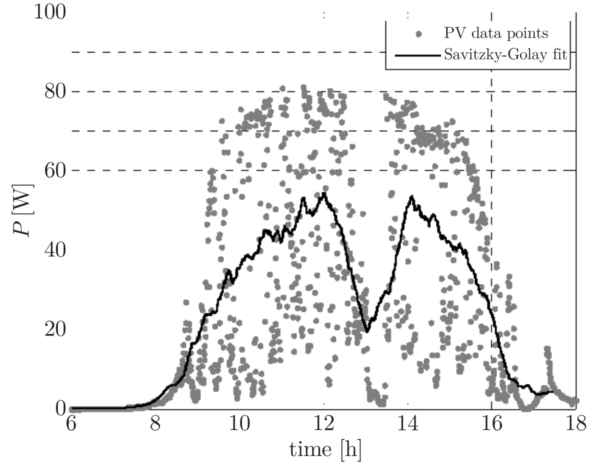

Figure 6. Measurements and estimated power PV output; $l=1$, $k=60$

Figures 9–10 present the results of the averaged power output for a higher polynomial degree. In this case the Savitzky-Golay method becomes more sensitive on changes of the power output. A period in which the power output is very low for relatively long time (approx. 1 hour) can be seen in Figure 9. For this reason the fitted curve is also low. The influence of this gap on the fitted curve is decreasing with the increasing parameter k (Figure 10).

After the change of the polynomial degree to the quadratic order the accuracy of the estimation also improved. It can be seen from the error analysis presented in Table 2. The maximum relative error of the calculated energy is decreasing from 1.66% for $l=1$ and $k=120$ to 1.37% for $l=2$ and the same value of k. However, further increasing of the polynomial degree did not improve the relative error at all. It seems that the degree of the polynomial fitted into the data equaling 2 is sufficient.

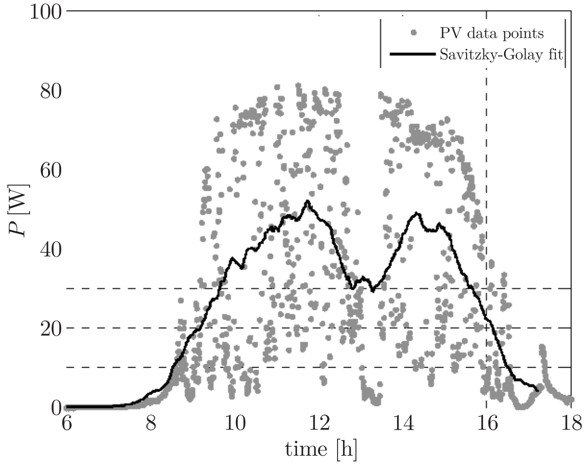

Figure 7. Measurements and estimated power PV output; $l = 1$, $k = 90$

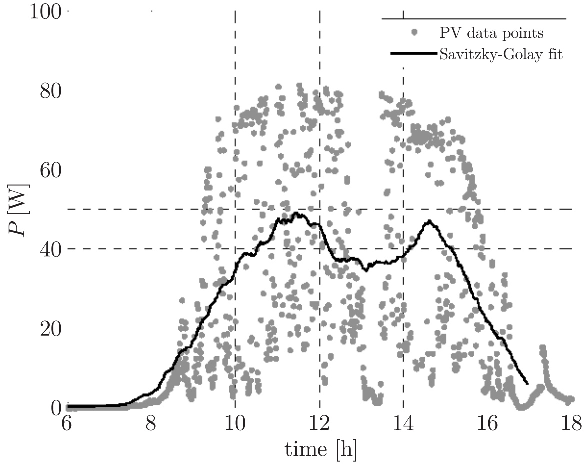

Figure 8. Measurements and estimated power PV output; $l = 1$, $k = 120$

Figures 11 and 12 present the fitting results for sunny and cloudy days, respectively. A very good agreement of the estimated power curve with the regular data can be seen in both cases.

The error analysis presented for a sunny day in Table 3 and for a cloudy day in Table 4 shows that energy calculated on the basis of the fitted curve is similar to the measured one. As one can expect, the relative error for the clear-sky day is smaller than for the sky partially covered by clouds. However, the relative error values for a cloudy day are higher. It is caused by the low value of the total energy obtained on this day.

5. Conclusion

This paper presents application of the Savitzky-Golay method for the PV data output obtained in different weather conditions: sunny, partly cloudy and

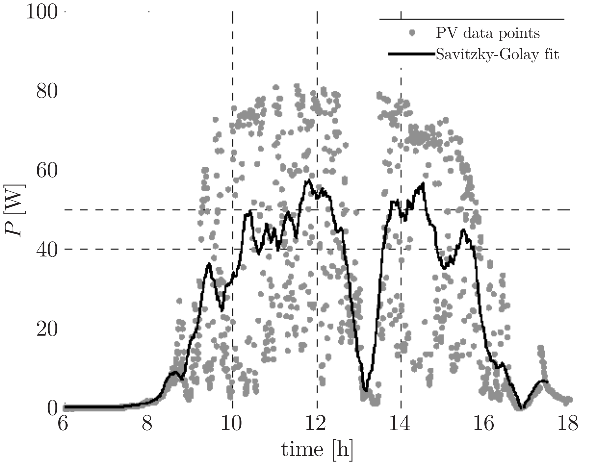

Figure 9. Measurements and estimated power PV output; $l = 2$, $k = 60$

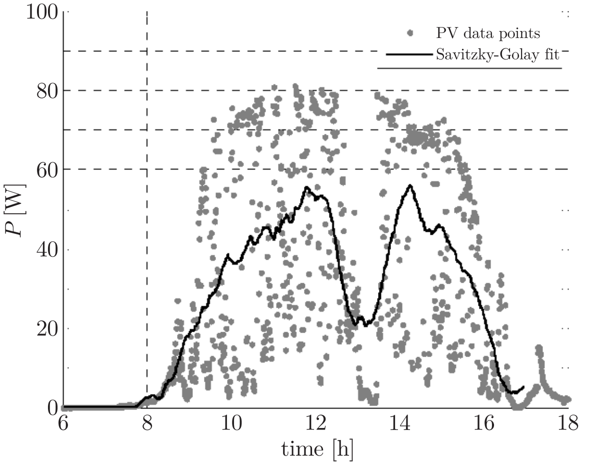

Figure 10. Measurements and estimated power PV output; $l = 2$, $k = 120$

Table 1. Relative error analysis of the energy obtained from the PV system for $l = 1$ and different parameter k

k	Energy measured from the system [Wh]	Energy calculated after S-G fitting [Wh]	Relative error of calculated energy [%]
30	281.50	280.84	0.23
60		279.50	0.71
90		278.57	1.04
120		276.81	1.66

fully clouded days. It has been shown that the quadratic function is sufficient for fitting into the PV data. Energy was calculated on the basis of the averaged data. The results of this calculation and the error analysis show that the presented

Table 2. Relative error analysis of the energy obtained from the PV system for $l = 2$ and different parameter k

k	Energy measured from the system [Wh]	Energy calculated after S-G fitting [Wh]	Relative error of calculated energy [%]
30	281.50	280.90	0.21
60		279.69	0.64
90		278.33	1.12
120		277.64	1.37

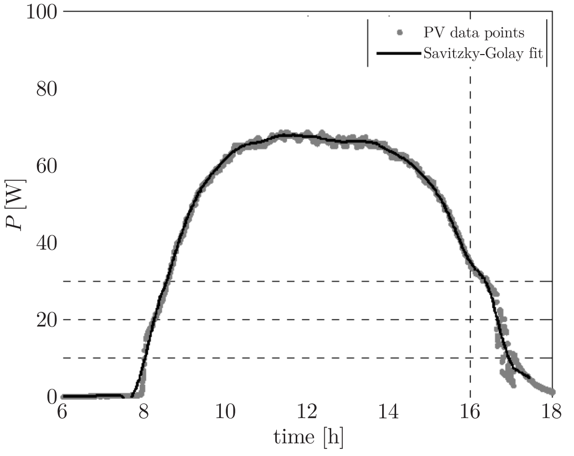

Figure 11. Measurements and estimated power PV output for clear-sky day; $l = 2$, $k = 60$

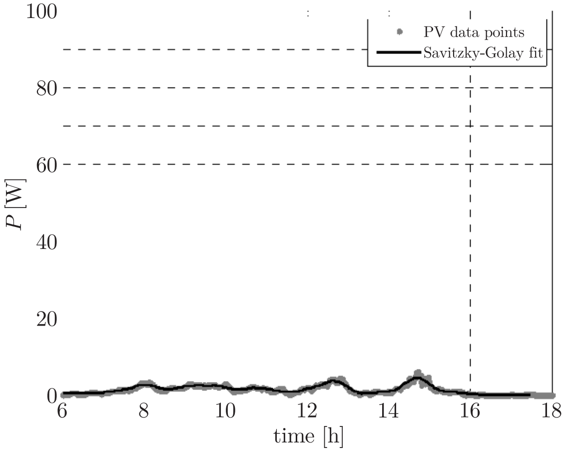

Figure 12. Measurements and estimated power PV output for clouded day; $l = 2$, $k = 60$

method gives an averaged power output for which the calculated energy is in good agreement with the energy measured in the system in changing weather conditions. For this reason the presented method is supposed to be a good tool to predict the energy produced by a PV system at a given location.

Table 3. Relative error analysis of the energy obtained from the PV system for a sunny day; $l = 2$

k	Energy measured from the system [Wh]	Energy calculated after S-G fitting [Wh]	Relative error of calculated energy [%]
30	463	462.51	0.11
60		461.62	0.30
90		460.33	0.58
120		458.01	1.08

Table 4. Relative error analysis of the energy obtained from the PV system for a cloudy day; $l = 2$

k	Energy measured from the system [Wh]	Energy calculated after S-G fitting [Wh]	Relative error of calculated energy [%]
30	18.1	17.97	0.91
60		17.76	2.03
90		17.55	3.21
120		17.32	4.47

References

[1] *http://webarchive.iiasa.ac.at/Research/TNT/WEB/Publications/publications.html*

[2] Carl-Jochen W 2009 *International Journal of Hydrogen Energy* **34** 1

[3] Scheer H 2005 *Photon International* 80

[4] Green M A 2002 *Physica E* **14** 11

[5] Goetzberg A, Hebling Ch 2000 *Solar Energy, Materials & Solar Cells* **62** 1

[6] Olchowik J M, Dragan P, Gembarzewski O, Gulkowski S, Szymczuk D, Tomaszewski R 2013 *The reasons of the dalays in introducing in Poland law regulations favorable for photovoltaics*, Proc of the 28th European Photovoltaic solar Energy Conference and Exhibition

[7] Olchowik J M, Gulkowski S, Cieslak K, Jozwik I, Banas J, Olchowik S, Zdyb A, Szymczuk D, Adamczyk J, Tomaszewski R, Zabielski K, Mucha J, Cieplak T 2006 *Comperative study of the solar modules performance in the hybrid system in south-easterly Poland during first two years of exploitation*, Procceedings 21st European Photovoltaic Solar Energy Conference 3049

[8] Olchowik J M, Gulkowski S, Cieslak K J, Olchowik S, Zabielski K, Jozwik I, Szymczuk D, Adamczyk J, Tomaszewski R 2007 *Analysis of the influence of diffuse solar radiation on the efficiency of photovoltaic conversion at south-eastern Poland condition*, 22nd European Photovoltaic Solar Energy Conference

[9] Press W H, Vetterling W T, Teukolsky S A, Flannery B P 2007 *Numerical Recipes in C: The Art of Scientific Computing*, Third Edition, Cambridge University Press

[10] Brandt S 1997 *Statistical and computational Methods in Data Analysis*, Springer Verlag