FULLY AUTOMATED PHOTOVOLTAIC SYSTEM MODELLING FOR LOW COST ENERGY MANAGEMENT APPLICATIONS BASED ON POWER MEASUREMENT DATA

Benedikt Hanke [a*], Mariano Bottega [a], Dorothee Peters [a], Nailya Maitanova [a], Jan-Simon Telle [a], Matthias Grottke [b], Karsten von Maydell [a], Carsten Agert [a]

[a] DLR Institute of Networked Energy Systems, Carl-von-Ossietzky-Str. 15, 26129 Oldenburg, Germany
[b] Hammer Real GmbH, Sylvensteinstr. 2, 81369 München / Munich, Germany
[*] Corresponding author: benedikt.hanke@dlr.de

ABSTRACT: Within this work an automated algorithm for the modelling of photovoltaic systems solely based on smart meter power readings of the system is proposed assuming low cost, stationary rooftop photovoltaic systems. As an example system, data from a nearly horizontal (7° tilt) thin film system is used for method validation. The photovoltaic power output modelling results are compared to an individually engineered INSEL® model and a Linear Regression (LR) Model. The resulting auto model output is discussed in the context of day ahead and intraday photovoltaic power forecast using no-cost / low-cost irradiation information. This work presumes that the characteristics of a photovoltaic system can be extracted from its clear sky /best system power curve. The extracted best system curve can be used directly as a model by scaling the clear sky power output with the ratio of the radiation to the clear sky radiation. The clear sky/best system curves are constructed by identifying the power maximum for a given time of the day for a data set of a sample number of previous days.
The auto model is comparable to the fitted INSEL® model and the linear regression (LR). The yearly error metrics for estimation of accuracy of the simulation, concerning 5-minute power values, are between -0.3% for the LR model, 3.7% for the INSEL® model and 4.7% for the auto model. Also the annual RMSE and MAE of the auto model (74.8 W and 27.4 W) are comparable to the LR (66.8 W and 26.2 W) and the INSEL® fit (68.1 W and 25.5 W). The auto model operates with less input data (generated energy only) and no knowledge about the engineering, design or surroundings of the PV system while yielding similar simulation results as conventional methods.
Keywords: PV System, Modelling, Prosumer

1 INTRODUCTION AND MOTIVATION

Energy systems in buildings and districts become more complex and diverse due to the available technologies and the ongoing sector coupling between electricity, mobility and heat. Additionally energy efficiency regulations encourage or even enforce local energy generation using renewable energy technologies [1][2]. To optimize the local electric infrastructure expenditure, load management becomes an integral part of future electricity grid designs, especially in the private sector targeting electrical installations behind the grid connection point. To integrate local photovoltaic production with flexible loads from sector coupling applications in the building, a low cost photovoltaic modelling and prediction is needed to assist the local energy and load management system. To ensure a low cost installation of the energy management system, operation and adaption due to degradation of the solar cells and the system, the modelling of the photovoltaic system needs to be highly automated and based on as few parameters as possible. These need to be configured without special photovoltaic knowledge by electricians without any additional training.

Various commercial (e.g. PV*SOL[1], INSEL®[2], PVsyst[3], HOMER Pro®[4]) and open source tools (e.g. PV_LIB[5] Toolbox) are available to model the behaviour and energy generation of a photovoltaic system mainly before system installation. This allows for building business cases and scaling renewable energy generation systems (e.g. mini grids or battery storage systems). Across these tools, a detailed knowledge about the photovoltaic system and the geographic lcation is required for the modelling. The pre-installation models are usually not fit to be used for operational optimisation, as installation changes need to be manually changed within the photovoltaic system model.

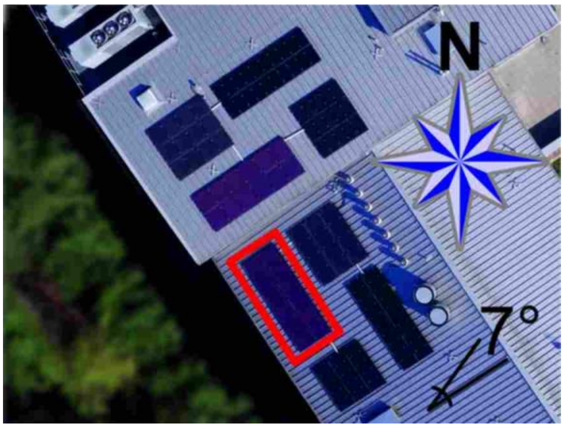

Figure 1: Roof top photovoltaic system at DLR Institute of Networked Energy Systems in Oldenburg

2 DATA

Data from December 1st 2016 until December 31st 2017 is used for evaluation and training purposes. Due to measurement errors in the photovoltaic system, the period from September 2nd 2017 until September 17th 2017 is replaced with data from 2011 of the same dates to achieve a full year functional dataset. The December 2016 time frame is only used for the automated modelling as a pre-training phase. All available data has been checked for plausibility.

[1] https://www.valentin-software.com/
[2] http://www.insel.eu/
[3] http://www.pvsyst.com/en/
[4] https://www.homerenergy.com/
[5] https://pvpmc.sandia.gov/applications/pv_lib-toolbox/

2.1 Photovoltaic System

The reference photovoltaic system used in this work is located at the DLR Institute of Networked Energy Systems in Oldenburg, Germany[6] (see authors contact for address and Figure 1). The roof top photovoltaic system is facing south-west (237°) with a 7° tilt.

The system uses 12 Schott ASi95 modules for energy generation with a nominal total system power generation of 1140 W_P. The modules are rated with a nominal cell efficiency of 6.6% at standard test conditions (STC). The module size is 1.108 m x 1.308 m = 1.499 m² resulting in an estimated total module area of 17.4 m² - ignoring additional installation gaps between modules. The resulting calculated module efficiency is 6.55%. For the year 2017 782.9 kWh of electricity are generated.

A SMA Sunny Boy 1100[7] is used as inverter for grid connection. A SMA Sunny Sensorbox is mounted in the plane of the photovoltaic modules to collect additional temperature and radiation information. The data is collected using a SMA Sunny Webbox configured to 5 min data interval.

For the manual system modelling, original system component data sheets are available for all system components.

2.2 Weather Measurements

Publically available reference weather measurements from the University of Oldenburg[8] are used as reference weather data for the system modelling. The weather station is located 32 m above sea level atop the main university building[9]. The weather station and the photovoltaic system are 1000 m apart. Weather data is publicly available in 30 min intervals since July 2003.

2.3 Data Resume

The plotted solar radiation of the given data set for the year 2017 versus the measured power of the photovoltaic system is shown in figure 2. The data points fit the simple linear regression of P = 0.0474 G with a coefficient of determination of 85% - resulting in a system efficiency of 4.74%.

The correlation between the measured radiation and the photovoltaic system power is given most of the time. The distance between the radiation measurement and the photovoltaic system locations is never the less visible in the wide scattering of the data points.

3 METHODOLOGY

For the Linear Regression Model (LR) the linear regression from section 2.3 is used.

Based on the described data set and system information, a static INSEL® model (manual model) of the photovoltaic system is designed. The manual model is benchmarked against the training data. The received benchmark is therefore reflecting the best case model performance for a full year model scenario.

The algorithm based automatic model (auto-model) is using only power and location information to create a model of the photovoltaic system (details see below). Thirty days of data from December 2016 are used to pre-train the auto-model, allowing for a full evaluation of the 2017 dataset. The same benchmark parameters are calculated as for the manual model. The dataset had some missing values (2nd – 18th of September and 11th – 13th of December) which were replaced by values from the year 2011 from the same PV system.

To describe the performance of the models the mean absolute error (*MAE*, eq. (1)), the root mean square error (*RMSE*, eq. (2)) and the simulation error (*ΔE*, eq. (3), [3]) are calculated using power values from the measurement P_M, the power values from the model simulation P_S and the number of data points n.

$$MAE = \sum_{i=1}^{n} |P_{Mi} - P_{Si}| \,/n \quad (1)$$
$$RMSE = \sqrt{\sum_{i=1}^{n} (P_{Mi} - P_{Si})^2 \,/n} \quad (2)$$
$$\Delta E = \frac{E_{model} - E_{measured}}{E_{measured}} \quad (3)$$

The results from the linear regression model and the manual model are used as a reference for the evaluation of the performance of the auto-model. The modelling results of both models are created by independent research teams to reduce the chance of result-biasing.

4 RESULTS

4.1 INSEL® Model

Two INSEL® models where created for comparison: the first solely build on database values and the second by matching the power measurements with the radiation measurements thus analysing the system performance. The block diagram of the created INSEL® models is shown in Figure 3.

- Block FROM contains a reference to the input data with global horizontal radiation (W/m²) and ambient air temperature (°C) for the whole year 2017.
- Block PVAI returns electric current and temperature of the PV modules depending on climate data and PV voltage. Calculation of the electric current is based on a modified one-diode model. The parameters of the block are given in Table 1.

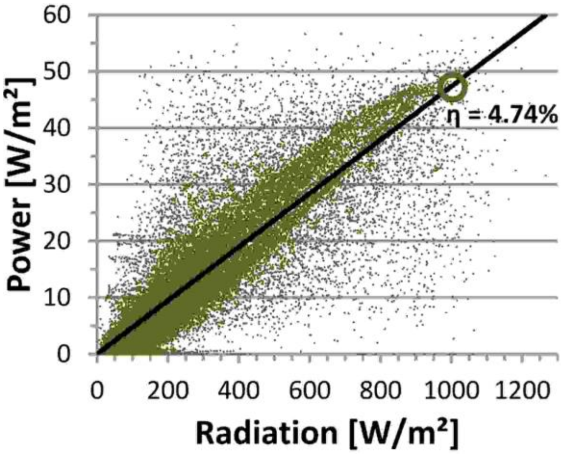

Figure 2: Measured solar radiation versus measured PV system power. The plot shows 105120 data points in 2017 at a resulting time resolution of five minutes.

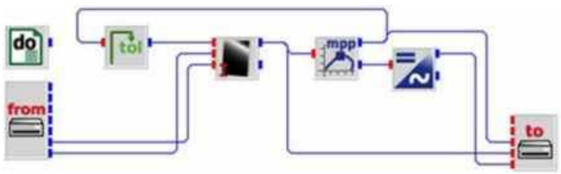

Figure 3: Block diagram of INSEL® model

[6] 53°09'03.8"N 8°09'59.5"E, https://goo.gl/maps/ewzJ1s5zcXy
[7] Serial number 2000689124
[8] http://www.uni-oldenburg.de/wetter/
[9] 53°08'51.4"N 8°10'48.4"E, https://goo.gl/maps/xiDm3Pgxa3m

Table 1: Characterization table of the block PVAI with the fitted parameters

Fitted Parameters	
Number of cells in series per module	24
Number of cells in parallel per module	3
Single cell area	0.0201
Single module area	1.449
Coefficient of short-circuit current density	0.113206+00
Temperature coefficient of short-circuit	0.568527E-04
Merten parameter	0.547033E+01
Built-in voltage	0.120437E+07
Coefficient of saturation current density	0.446160E+01
Temperature exponent of saturation	0.457965E+01
Band gap	0.152050E+01
Diode ideality factor	1.791466
Series resistance	0.44343E-03
Parallel resistance	0.110715E+01
Module tolerance plus	5.0
Module tolerance minus	-5.0
Characteristic module length	1.308
Module weight	18.000
Absorption coefficient	0.70
Emission actor	0.85
Specific heat of a module	900.0
Nominal operating cell temperature	47.0
Single cell voltage error tolerance	0.100000E-02

- Block MPP calculates maximum power points of the PV system under different climate conditions. This block operates in a combination with a block TOL (top of loop)
- Block IVP simulates the operation of the inverter and its losses and returns ac power produced by the PV system.
- Block TO gathers selected results after running the simulation and saves them into a file.

The block PVAI contains standard parameters, which need to be fitted for amorphous silicon solar modules and also for PV systems that have already been taken into operation in order to take a degradation of the modules into account.

The parameters of the block PVAI presented in the characterization table above were fitted on measured data: electric current and voltage of the real PV system, global horizontal radiation and air temperature measured for the whole year 2017. Then this fitted block was integrated in a model (see the block diagram above) to simulate a time series of the PV power over the year. In order to define the effect of the fitting, the block PVAI with the standard values of the parameters and values only from the datasheet, i.e. without fitting, was also be used to simulate PV power over the year.

4.2 Auto Model

The approach of the auto model [8] deals with the assumption to forecast and model the power generation of a PV system without knowledge about the orientation or any parameters of the modules and the inverter. The model only works with historical measured PV power data. The flow chart Figure 4 shows how the auto model works.

To predict/model the generated power of the next day the auto model only uses measured PV power data from the last few days. Therefore, PV data from the 1st December 2016 until the 31st December 2017 were used from the PV system which was described in chapter 2.1.

To find the best system power curve for the next day, the algorithm takes a sample number of PV data from previous days (PD) and calculates the maximum value for every time step (288, 5-minutes resolution) of all chosen previous days. Thereof results the best system power (P_{BS}) as shown, for the example of six previous days, in Figure 5.

To model and scale the PV power (P_{MO}) and make the auto model comparable to the INSEL model and the linear regression, the P_{BS} was scaled by the ratio of the measured irradiance GHI (GHI_{MS}) (Global Horizontal Irradiance) and the clear sky GHI (GHI_{CS}) irradiance. The comparison of the P_{MS} curve, the P_{BS} curve and the P_{MO} curve is presented in Figure 6 for the example of 6 previous days.

To find the optimum of how many days to look back to find the ideal P_{BS}, all calculation steps were repeated thirty times previous days and the calculation of 365 P_{BS}

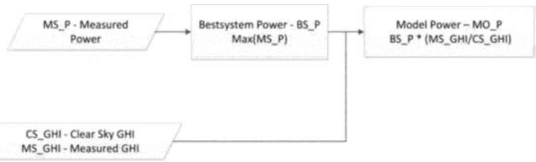

Figure 4: Flow chart of the auto model

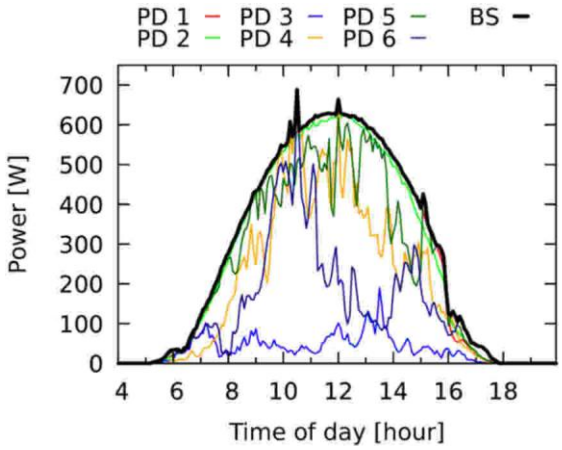

Figure 5: example of building the best system power curve of six previous days

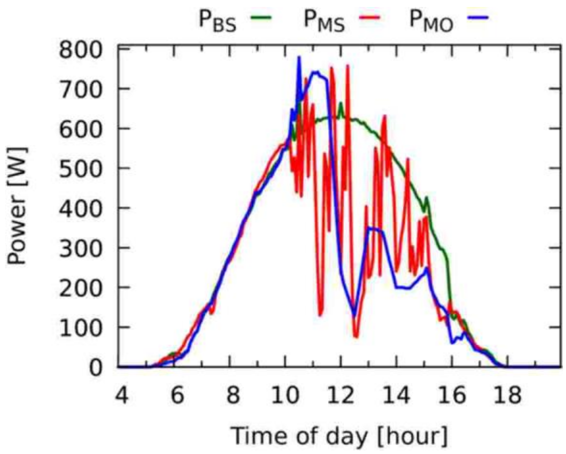

Figure 6: Comparison of the measured power P_{MS} and the auto modelled best system power curves P_{BS} and model power curve P_{MO}.

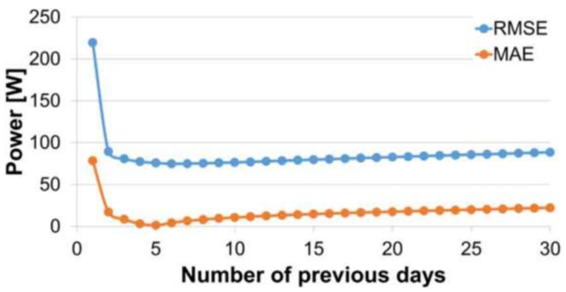

Figure 7: Simulation error from the auto model simulation of one year as selection criteria for the number of PD.

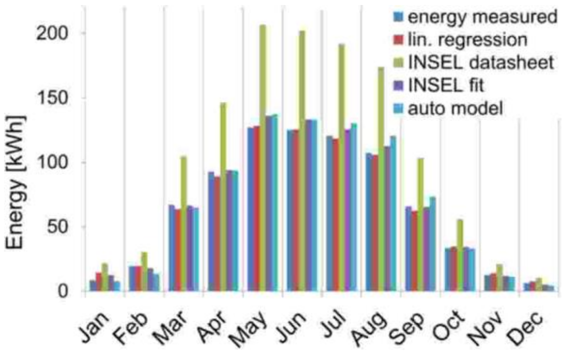

Figure 8: monthly energy output of the PV system simulated with different models in comparison with the measured energy output from PV system for year 2017.

curves. The selection criteria for the optimal number of previous days were the minimum simulation error or rather the *RMSE* (as seen in eq. 1) and *MAE* (eq. 2) between the modelled power and the measured power as shown in Figure 7.

From these results it was suggested that the optimal number of previous days is 6. Another important point is to choose PD > 1 so that persistence should exclude and also not to go too far in the past. This have a negative effect with regards on seasonal impacts of time-dependents increasing or decreasing irradiance/power.

4.3 Linear Regression Model

The linear regression model described in section 2.3 is used as a most simple model for the photovoltaic system. This model requires radiation and output power as input parameters and cannot be created independently of weather data. The model is forced to 0 W power output P_{LR} at 0 W/m² of solar radiation G by omitting a constant factor. The resulting linear regression is shown in eq. (4).

$$P_{LR} = 17.4\ m^2 * 0.0474 * G \qquad (4)$$

The *MAE* for this model is 26.3 W and the *RMSE* is 67.2 W based on the training data set of 2017 values.

4.4 Simulation results

The results of the linear regression approach, INSEL simulations with and without fitting of the block PVAI and auto model were evaluated and compared with the measured power of the PV system (see Fig. 8).

When the block PVAI was not fitted, the INSEL software overestimated the power of the PV system dramatically. INSEL didn’t take the features of the amorphous Si and the degradation effect into account. A 7 year old PV-system requires manual adaption of the datasheet values to reflect current system behavior.

Table 2: Statistical metrics for estimation of the accuracy of the simulation (concerned power values)

	MAE	RMS	Simulation
Linear	2	66.8	-0.3 %
INSEL	6	130.4	61.5 %
INSEL fit	2	68.1	3.7 %
Auto	2	74.8	4.7 %

Table 3: Statistical metrics for estimation of the accuracy of the simulation (concerned energy values)

	Lin.	INS	IN	Au
Simulation error [%]				
Mean	41,01	115,	27,	14,
Minim	-	5,11	-	-
Maxim	2784,	4610	28	17
MAE [Wh]				
Mean	2,18	5,08	2,1	2,2
Minim	0,14	0,09	0,0	0,0
Maxim	8,53	16,5	9,0	9,7
RMSE [Wh]				
Mean	4,33	8,77	4,3	4,8
Minim	0,22	0,22	0,2	0,1
Maxim	15,59	25,5	16,	18,

An accuracy of the considered simulation approaches was measured with the help of three statistical metrics: mean absolute error (MAE), root mean square error (RMSE) and simulation error ΔE (difference between simulated and measured energy outputs of the PV system in percent).

The values of the simulation error can be positive and negative: positive values mean overestimation and negative values – underestimation. The calculated values of the MAE, RMSE and simulation error ΔE are demonstrated in the Table 2. These results are comparable and better referring to the simulation results of [4] with a mean RMSE of 80.97 W and a MAE of 95.17 W.

The statistical metrics also indicated that simulation values of the INSEL model with non-fitted PVAI block diverged significantly from the measured power of PV system. The fitting of the model improved the simulation results and increased the accuracy of the simulation. Despite the fact that the PVAI block was fitted, the INSEL model continued to overestimate the measured values, but to a far lesser extent. The datasheet based INSEL® model was discarded as not being suitable for benchmarking the auto model.

The annual energy output calculated with the linear regression approach deviated from the measurements o by only 0.3 %. The deviation of the energy outputs delivered by the fitted INSEL® model and auto model amounted to 3.7 % and 4.7 % respectively.

The fitted INSEL® model and auto model had one significant commonality: these two models underestimated the PV energy output in almost all winter months and overestimated it in summer months. The simulation error per day is shown in Figure 9 and per month is shown in Figure 11. During the summer the simulation error is in a much lower range as during the winter month. This can be accounted to the much higher generation. The error distribution is presented in Figure 10. In 73% of the days, the errors are within ±20% and in 90% of the days the error is within ±40% of the generated energy. As a result, 37 days per year have a power prediction error of more than ±40%. Taking the annual distribution (Figure 9) into account, these outliers happen during times of low solar activity and are therefore representing only a small part of the energy generation.

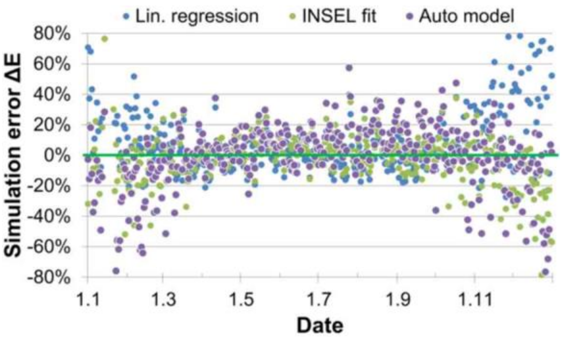

Figure 9: Values of the simulation error ΔE in % per day

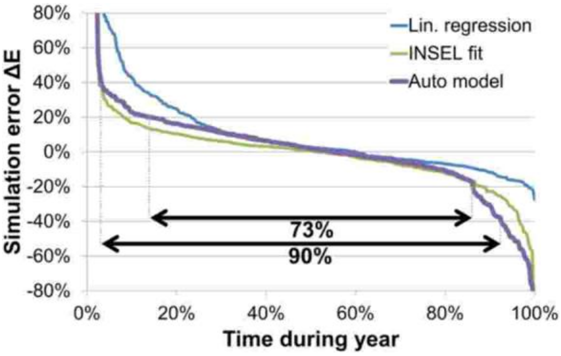

Figure 10: Distribution of daily simulation error ΔE in % over one year (365 days).

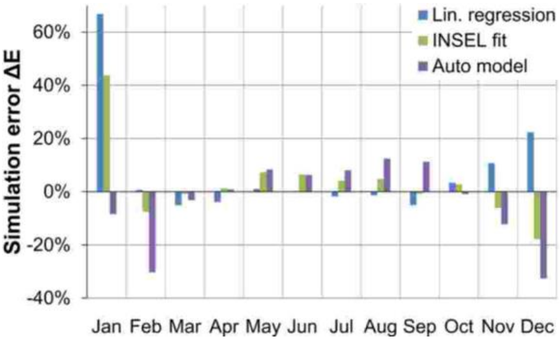

Figure 11: Simulation error in percent ΔE of the fitted INSEL model and auto model

Table 4: Monthly simulation error ΔE

Month	Lin.	INSEL fit	Auto
January	66,9%	43,7%	-8,4%
February	0,7%	-7,5%	-30,3%
March	-5,1%	-0,8%	-3,2%
April	-3,9%	1,4%	0,9%
May	1,1%	7,3%	8,3%
June	0,3%	6,5%	6,3%
July	-1,8%	4,1%	8,1%
August	-1,3%	4,8%	12,5%
September	-5,0%	-0,7%	11,3%
October	3,4%	2,8%	-0,9%
November	10,8%	66,3%	-12,2%
December	22,4%	64,5%	-32,6%

Although the annual simulation error of the auto model was not much higher than the annual error of the fitted INSEL model, the auto model underestimated the PV energy output in winter months up to 30 %.

The simulation error shown in Fig. 11 and Table 3 are comparable to [5] and [6] with simulation errors between 3 % and 5.2 % of simulation and measured PV power data in Germany. The course of the monthly simulation errors are consistent referring to [5] and also the great deviation of the monthly simulation error between summer and winter months are comparable referring to [7].

5 DISCUSSION AND OUTLOOK

The presented approach proved suitable to predict the behavior of a photovoltaic system for an upcoming day, by providing at least two days of measured photovoltaic output. The number of previous has a large influence on the quality of the systems model. Especially in the second half of the year, the better performance of the system is expected for the earliest days in the dataset – therefore ignoring seasonal variations for longer periods of time. In a further optimization of the algorithm, the historic data timeframe used for model creation can be obtained dynamically by analyzing the prediction error and modifying the data frame time span used for model creation. Further the historical data could be scaled by its individual clear sky G_{HI} curve, to reduce seasonal effects on the input data.

Figure 12 plots the PV systems power output, irradiation and the modeled results of 19 days in January 2017. Due to snow on the modules the power generation was far below the expectations from the radiation measurements. Starting Jan. 13th the auto model starts to adapt – as it is not using the radiation readings and only reacting to the changes in power output - and the predictions of the 19th and 20th of January represent very well the electric generation despite high levels of solar radiation. Starting on the 22nd the snow started to melt and the generation got back to normal. The auto model however needed the number of previous days (six) to adapt back to normal operation. This demonstrates how the algorithm is able to react to introduced disturbances (e.g. shadow of growing trees or newly build structures in the area of the generation site, failing strings, dust and dirt) without any interaction of an operator. In case of the snow the very short period of disturbance is not bringing any advantage, as the other techniques do not require re-adaption after the disturbance is gone. The behavior of the fitted INSEL® model is visible in Figure 13 and the one of the linear regression model in Figure 14 respectively. As the output power has no feedback loop to the model, the radiation and temperature measurements are the only indicator for the model calculation. Therefor beginning with the 21st the calculated power is instantly a very good approximation of the generated power. By comparison the linear regression model is working better in low irradiance situations in comparison to the fitted INSEL® model. A dynamic evaluation of the prediction error could reduce the size of the historic data frame after a radical change in the photovoltaic generation, allowing for faster adaption times.

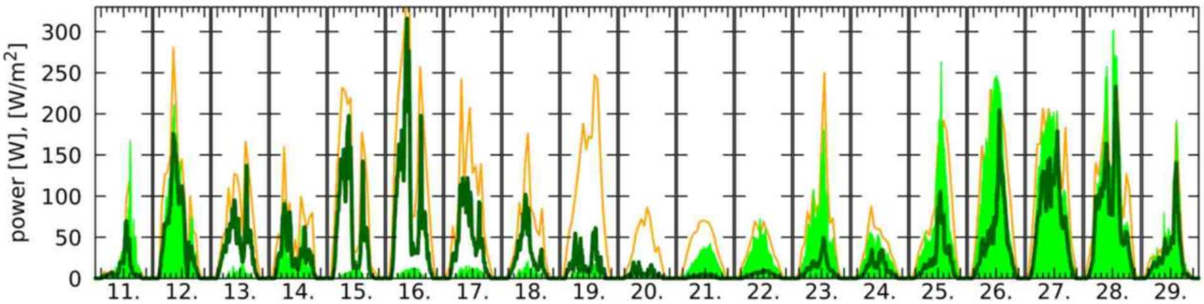

Figure 12: Irradiation (orange), generated power (light green, area) and auto model prediction (dark green) for 19 days in January 2017 (days are given below subgraphs) from 7:00 to 17:00.

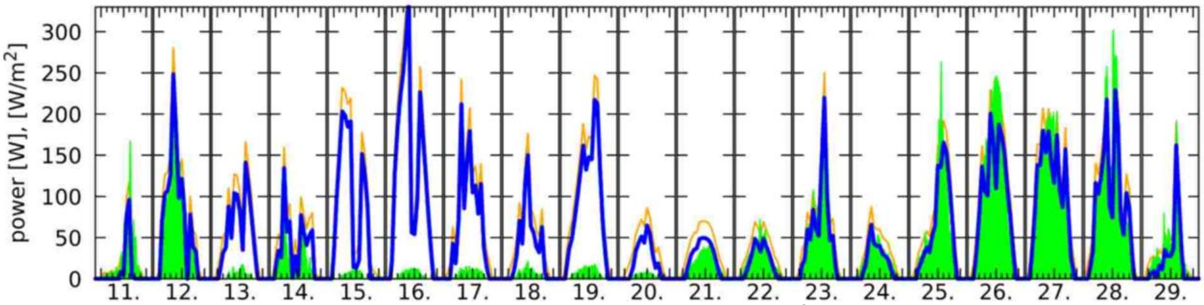

Figure 13: Irradiation (orange), generated power (light green, area) and INSEL® model prediction (blue) for 19 days in January 2017 (days are given below subgraphs) from 7:00 to 17:00.

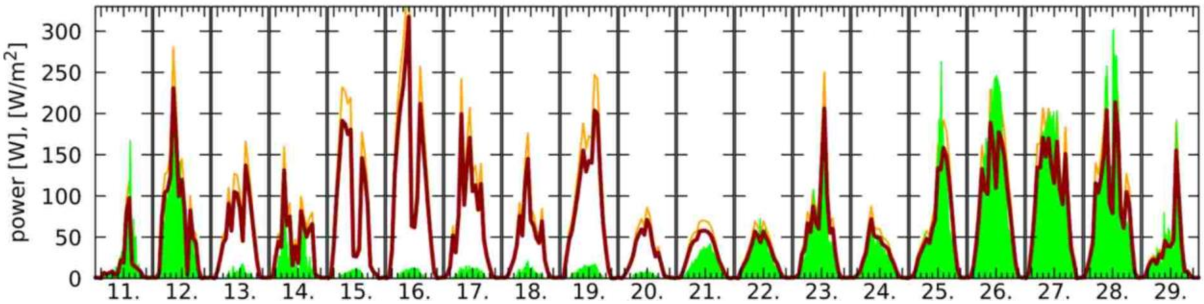

Figure 14: Irradiation (orange), generated power (light green, area) and linear regression model prediction (red) for 19 days in January 2017 (days are given below subgraphs) from 7:00 to 17:00.

6 REFERENCES

[1] die tageszeitung (taz), July 5th 2018, p. 08, "Sonne darf nicht mehr nur so scheinen"

[2] „Verordnung über energiesparenden Wärmeschutz und energiesparende Anlagentechnik bei Gebäuden (Energieeinsparverordnung - EnEV)", status as of October 24th 2015, §5 „Anrechnung von Strom aus erneuerbaren Energien"

[3] Zinßer, B; Dissertation (2010); Shaker Verlag Aachen; „Jahresenergieerträge unterschiedlicher Photovoltaik-Technologien bei verschiedenen klimatischen Bedingungen"

[4] Chouder, A. et al.; Solar Energy (2012); "Monitoring, modelling and simulation of PV systems using LabVIEW"

[5] Becker, G.; et al.; "Energy Yields of PV Systems - Comparison of Simulation and Reality"; Bavarian Association for the Promotion of Solar Energy; Stuttgart University of Applied Sciences, Munich University of Applied Sciences, BEC Engineering GmbH

[6] Matthiss, B., et al. (2015). PV Performance Modelling Workshop; "Comparison and Validation of System and Irradiance Models for Yield estimation"

[7] Schweiger, P. Oberpertinger, R., Petz, A.. Forschungsforum der österreichischen Fachhochschulen (2017); „Performance-Monitoring von PV-Anlagen auf der Basis von Wetterdaten. Ertragssimulation und potenzielle Fehlerquellen"

[8] Bottega, M., Master thesis (2017); "Management of Local Energy Availability and Customer Charging Demand for Battery Electric Vehicles Parking Areas"; University of Oldenburg

7 ACKNOWLEDGEMENTS

The authors acknowledge the financial support of the "Federal Ministry for Economic Affairs and Energy" of the "Federal Republic of Germany" for the project "EG2050: EMGIMO[10]: Neue Energieversorgungskon-zepte für Mehr-Mieter-Gewerbeimmobilien" (03EGB0004G). For more details see: emgimo.eu. The presented work is created as part of this project.

Finally the authors acknowledge the paper size limitations for the conference proceedings of the EUPVSEC and regret the figure quality and jpeg artefacts.

[10] https://www.emgimo.eu/