Hourly PV Power Output Predictions in Pre-Alpine Terrain in Switzerland

Dr. André Meyer, Dr. Peter Meyer, MIT Coaching GmbH, Version: July 2022, not peer-reviewed

1 Motivation

To reduce our reliance on fossil and nuclear energy, states and society are increasingly interested in and heavily investing in renewable energy. In particular focus is solar energy, due to its abundance of sunlight. However, the various meteorological and topological impacts make reliable operation challenging - which is why it is of great interest to energy companies and private consumers to have access to accurate predictions of their photovoltaic (PV) systems' power output. While their interests are aligned, the timespans of interest are different:

- **Intra-day (12-24h) predictions** allow private consumers to economically optimize the usage of their own or network power, by charging their EV, running the heat pump or using their whiteware at the optimal time.
- **Intra-hour (15-60min) predictions** allow utility companies and power grid operators to develop more efficient and sophisticated strategies for managing their grids and reduce costs by buying or selling power at the right time and avoid price fluctuations, by leveraging more robust knowledge about upcoming availability of power [e.g., Ayon2017, Zhou2017].

Finally, comparing PV power output predictions based on meteorological forecasts allows the introduction of a feedback loop to improve meteorological forecasting models based on a better understanding of factors impacting the local weather.

2 Related Work

Our extensive literature review showed various challenges for developing accurate PV power output predictions, including inaccuracies in meteorological forecasts, seasonal differences, PV system differences, influences from the sun position, snow, shadow, and dust, as well as limited availability of actual PV power output data as Ground Truth in reasonable quality and quantity. Besides the more traditional physical and statistical methods, most recent approaches focused on applying AI/ML techniques [e.g., Li2019, Giorgi2015], for both, the intra-day and intra-hour predictions. An overview of the most promising recent approaches can be found in the following table:

Table 1: Overview of previous research related to PV power output prediction, split into intra-day and intra-hour

Category	Techniques Used	References
Intra-day	Hybrid Systems	Ciapala2018
	Neural Network (*RNN, **BPNN)	AlDahidi2018**, Theocharidies2018*, Kou2013**
	Regression	Clack2017, Wolff2014, Sunday2015, DeBock2014, Mouatasim2018, Ramenah2018
	Other	Mellit2005,
Intra-hour	Hybrid systems	Lee2019-1, Li2019
	LSTM	Huang2019, Lee2019-1, Li2019
	Neural Network (*RNN)	Lee2019-1*, Li2019*, Sfetsos2000
	Regression (* with BP)	Li2019, Zhong2018*, Reindl2017, Bouzerdoum2013, Shi2015, DaSilva2012
	Deep Reinforcement Learning (DRL)	Dorokhov2022, O'Grady2022, Lee2019-2,
	Other	Pelland2013, Crisosto2018, Zhong2017

In summary, the wide variety of ML approaches, the very diverging results and qualities of the resulting models, and the stark differences in relevant features impacting model quality, show that this area of research is still in its infancy and only very few best practices for operational use are established as of today.

3 Approach

The explorative nature of our task resulted in an initial focus set on developing **intra-day** predictions of PV power output on the intra-day timeframe of 12-24 hours. To our knowledge, other than a few commercial approaches with limited prediction quality, there are only very few (published) systematic approaches studying PV power output predictions in Switzerland. In what follows, we describe our approach, which starts with identifying, aggregating, and cleaning the required data, developing features as input for our ML pipeline, scoring the resulting models and optimizing them based on residual analysis.

3.1 Data Collection & Feature Engineering

To develop PV power output predictions for the intra-day case, we collected data for three locations: PV system information (such as location, tilt, number of modules, types of modules and inverters), measured PV power output, and meteorological data (such as temperature, global irradiance, clouds).

3.1.1 Measured PV Power Data (Ground Truth)

The independent variable (aka Ground Truth) is the PV power output (in watts, sometimes known as power yield). It was measured and logged by each systems' sensors. We neglected the sensors' measurement errors, since they are generally very small compared to the prediction errors [Li2019].

For the three PV systems we had access to, the following table summarizes PV system producer information for the modules and inverters, their location, installation with respect to the sun (azimuth and tilt) and the amount of data we aggregated per location. Note that the locations are all Switzerland, geographically located close to each other (1.8 - 5kms apart), and are in the same time zone (GMT+1). They are situated in pre-alpine (rugged) terrain, between 0.5 and 1 kilometers south of the Lake of Zurich.

Table 2: Description of PV Systems used for Ground Truth (and amount of available data)

PV System	Location Latitude Longitude Elevation	Azimuth	Tilt	Modules	Modules per String	Strings per Inverter	Inverter	Number of days (hours) of data	Comments
Lab32	47.19763 8.72126 510m	90° 270°	30° 30°	21x SunPower SPR-327NE-WHT	3 6	3 2	1x SolarEdge Technologies Ltd SE10K-480V-CEC	814 (11'379)	Lab MIT Coaching PV power output measured as one system, managed by Solar-Log 1200.
EWH1	47.18902 8.70555 650m	145°	15°	217x Jiangyin Hareon Power_HR_260P-18-Bb	11	5	4x Solarmax 15MT 15MT3A-480V-CEC *	768 (10'728)	Data via our partner EW Höfe AG
EWH2	47.19632 8.77167 683m	163°	10°	1031x Seraphim Solar System SRP-260-6PB	17	5	12x Huawei Technologies SUN2000-22KTL-US-CEC *	768 (10'742)	

** The EWH1 and EWH2 PV systems use several inverters models and not just the one listed in the 'Inverter' column. To simplify the models, the most often used inverter was used. The other inverters' performance characteristics were compared using the NREL SAM database and they were found to behave very similarly.*

3.1.2 Meteorological Data

Our partner, Kachelmann GmbH, provided us with industry-leading, high resolution meteorological data for the three PV system locations. The forecast granularity is 1 hour, which in turn impacts the PV power output prediction granularity (i.e., also 1 hour). The forecast time window is 13-18 hours (forecast calculation every 6 hours). The resolution is 1 km^2. The meteorological data is available for the entire Ground Truth dataset.

3.2 Pre-Processing: Data Understanding and -Cleaning

A core part of our work was an extensive pre-processing of the aggregated data, which includes plausibility checks (inter- and intra-comparisons of the 3 datasets and seasonal corrections), removing missing data, removing outliers (using IQR and rolling median filtering), normal distribution checks (Shapiro-Wilk, Lilliefors), removing data before sunrise and after sunset, and correlation tests between input factors (Pearson and pairwise correlation heatmaps).

3.3 Feature Engineering

Based on the aggregated and carefully cleaned data, we developed and extracted 14 features, categorized as meteorological, PV system specific, and seasonal.

Table 3: Extracted features (with units) and determination of whether they were selected by the final General Model (using SelectKBest)

Categories	Feature	Units	Features selected
Meteorological	Temperature	Celsius	*
Meteorological	Effective irradiance (total plane of array (PoA) irradiance, adjusted by sun position, array orientation (tilt, azimuth), meteorological characteristics (diffuse light), based on global predicted irradiance, applied using Python Pvlib [William2018])	Watts	*
Meteorological	Sunshine duration	0-1	*
Meteorological	Clouds low	%	*

Meteorological	Clouds medium	%	*
Meteorological	Clouds high	%	*
Meteorological	Precipitation	mm	*
Meteorological	Wind speed	m/s	*
Meteorological	Relative Humidity	%	*
Meteorological	Snow	mm	-
PV system specific	Model-based power output calculation (using panel-specific constants from NREL SAM, assumption: no shadow, calculations based on [King2004], applied using Python Pvlib [William2018])	Watts	*
Seasonal	Time of Day	Hour	*
Seasonal	Day of Month	Day	-
Seasonal	Month of Year	Month	-

3.4 Training & Scoring ML models and Residual Analysis

The feature dataset was split into training and test sets; the test set accounting for 11% of the data (every 1st, 10th, 20th of each month). We then developed a regression pipeline to train, validate and test predictive models using a broad variety of regression classifiers (R.): Linear R., Ridge R., Lasso R., ElasticNet R., Logistic R., Random Forest R., K-nearest Neighbor R., Support Vector R, Boosted Decision Tree R., Gradient Boosting R., Extreme Gradient Boosting R., and Multi-Layer Perceptron (Feedforward Artificial Neural Network). As described above, the pipeline performs scaling (StandardScaler, RobustScaler, QuantileTransformer) and dimensionality reduction (feature selection using Principal Component Analysis (PCA), SelectKBest, recursive feature elimination (RFE)) as a pre-processing step, and then runs a 10-fold cross-validation with each regression classifier. Hyperparameter tuning is performed to determine the optimal parameters for each regression classifier, by optimizing for both *rsquare (r2)* and the *negative root mean squared error (nRMSE)* in separate runs. We adapted the cross-validation approach to account for the temporal dependency of samples, using TimeSeriesSplit. The cross-validation splits the training data set into training and validation sets, and trains the predictive models on the first, and validates them with the latter. Using the test set, the resulting predictive models were then scored and ranked using standardized scores for regressions: R2, adj. R2, MAE, relative MAE (MAE / max yield), MSE, RMSE, nRMSE, CV/bias, ME, MPE, SMAPE, Pearson Correlation, Kolmogorov-Smirnov, and others. Visual inspection and residual analysis were performed on the PV output predictions, which, for example, led to the introduction of seasonal features to better model temporal dependencies.

4 Prediction Performance

Our results show that training a general model on the entire dataset (i.e., LAB32, EWH1, EWH2) results in overall good performance. Individual models perform only slightly better. Of the 12 trained and hyperparameter tuned models, the three best performing classifiers (in increasing order of performance) are Random Forest, Extreme Gradient Boost, and the Multi-Layer Perceptron (ANN).

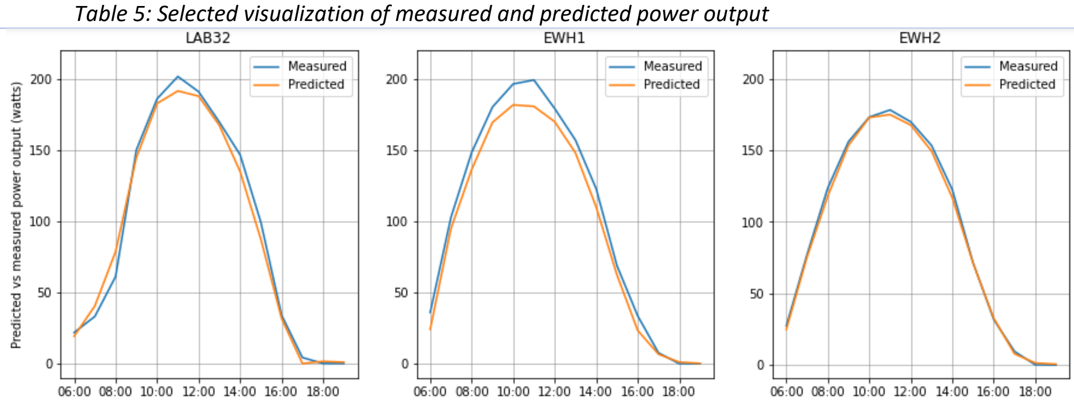
Table 4: Standard scores for the Individual Models and General Model (Combined) trained with the Multi-Layer Perceptron (ANN)

	R2 (%)	adj. R2 (%)	MAE (w)	rel. MAE (%)	MSE	RMSE (w)	NRMSE (%)
LAB32	0.874026	0.872673	14.051059	0.051121	585.551378	24.198169	0.088039
EWH1	0.850271	0.848562	14.404942	0.065842	563.367171	23.735357	0.111123
EWH2	0.844395	0.842618	12.446203	0.058011	460.995214	21.470799	0.108088
Combined	**0.860481**	**0.859965**	**13.397499**	**0.048743**	**536.527724**	**23.163068**	**0.084273**

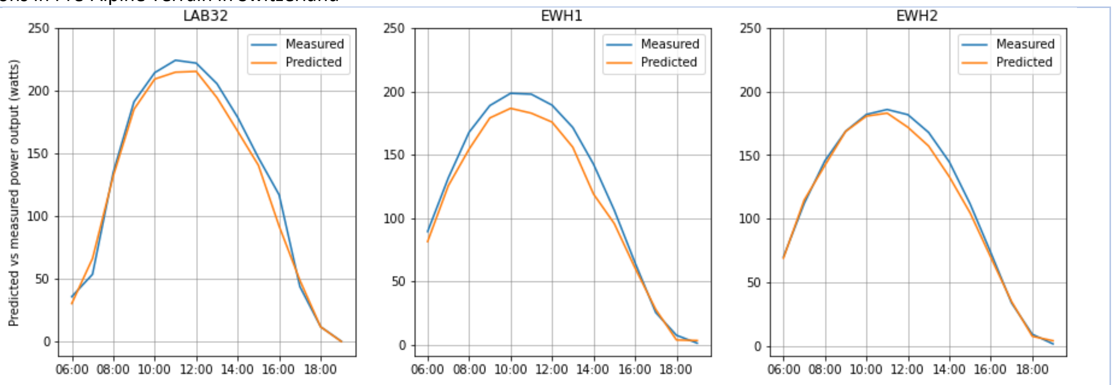
Table 5 shows **excerpts of good predictions** for all 3 sites on a selected day per season (note the different y-Axes).

Table 5: Selected visualization of measured and predicted power output

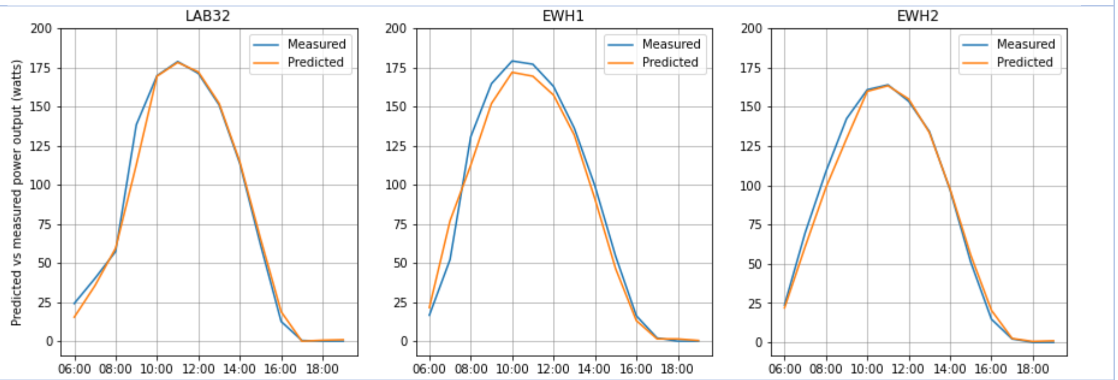
Spring



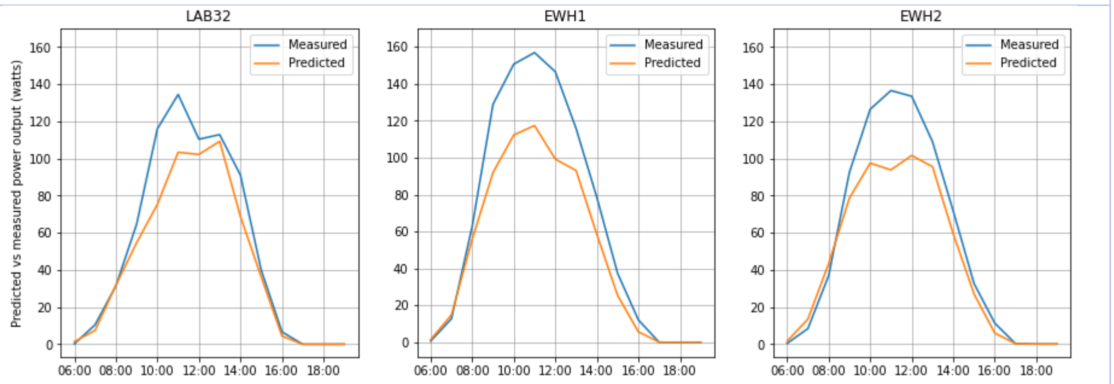
Summer



Autumn



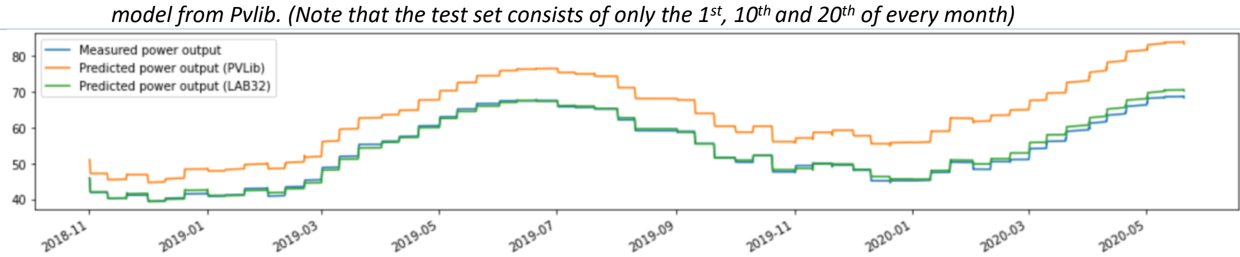
Winter



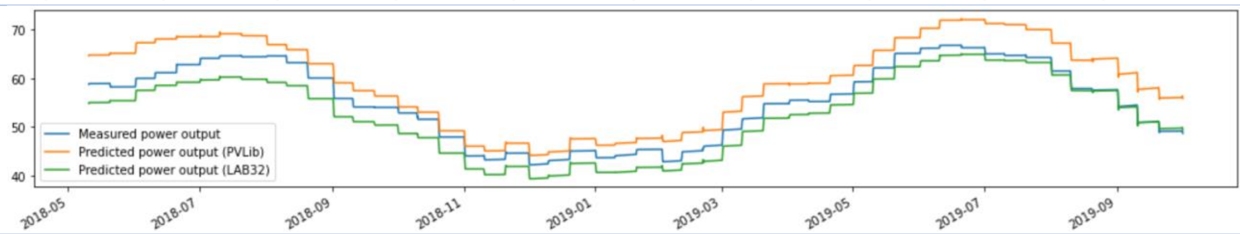
In a separate analysis, we investigated the ability to train **seasonal prediction models**. Overall, they performed worse than the entire dataset, but better smoothed out large residuals. With a larger dataset consisting of more than just two seasons each, we expect that seasonal models to perform better. A seasonal residual analysis shows that since the power output is generally higher during the spring and summer seasons, the absolute residual values are higher as well. The following table visualizes the seasonal analysis exemplary, as the seasonal trend for the measured and predicted power output (including a comparison to [Python Pvlib]) is visualized.

Table 6: Visualization of the trend when comparing measured with predicted power of our model (Lab32), as well as a comparison with a model from Pvlib. (Note that the test set consists of only the 1st, 10th and 20th of every month)

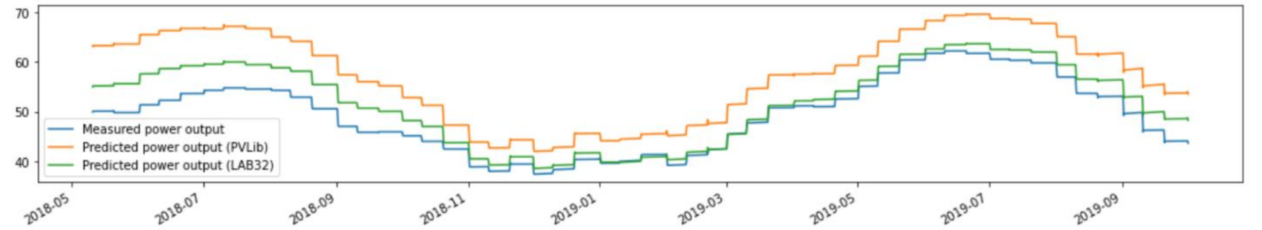
Lab32



EWH1



EWH2



While the high-quality and high-resolution meteorological data and inclusion of PV system characteristics allowed for near-perfect predictions on sunny (i.e., clearsky) days (e.g., the summer and spring example days in Table 5), there are also many days which are much harder to predict accurately on a intra-day timeframe of 13-18h. The large variability and volatility of weather conditions and mountainous, pre-alpine location of the three PV systems, aggravate accurate predictions on this longer timeframe. Overall, the prediction performance is comparable to results from previous work, which in most cases were optimized for locations with easier meteorological setups, such as no pre-alpine terrains.

5 Contributions & next steps

Our resulting model is trained with 13 features that were extracted from three large datasets in Switzerland after extensive data cleaning and understanding, and hyperparameter tuning on a broad set of 12 regression algorithms. In this report, we described the **intermediate** results of our predictions and show that it is feasible to predict PV power output by extracting not just meteorological features, but also extracting two features that consider the PV system characteristics.

In 2021, we will continue to **improve our intra-day models**, by extracting additional features (e.g., humidity), by continuing improving the model based on residual analysis, by testing our hyperparameter tuning with additional scoring functions, and by training the models with additional data from longer time periods and new locations. Larger and separate datasets will further allow us to train and optimize several models, e.g., by splitting them up by season or weather classification types (sunny/clearsky, cloudy, rainy/foggy). We also intend to optimize model efficiency, by selecting and developing models for reduced training times and resource consumption, to improve economic and ecological aspects, which will be relevant for applying our models in practice.

Another core focus will be on developing accurate PV power output predictions for our **intra-hour (15-60min) use case** for utility companies and power grid operators. We expect to be able to develop significantly better models than in the intra-day approach, based on our access of live data with minimal delay, including measured PV power data (access via PV systems' web API) and additional meteorological live data from local weather stations (e.g., Netatmo). This will allow us to not only include adjacent days, but also intra-data data into the models, which will be based on Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) techniques.

Finally, we are working on several practical **applications for consuming our intra-day and intra-hour PV power data prediction models**. This includes a private consumer facing interface to optimize power usage for the next day, based on expected PV power output and estimated future power use. Finally, we also hope to establish a partnership with a utility company or power grid operator to co-develop models for smart grid management and power trading, by leveraging intra-hour PV power data predictions and forecasts of (private) customers' power consumption.

Needs/Asks for Future Work

To achieve our so far self-financed visions and goals, we are looking to establish **industry and research partnerships**:

- New industry partnerships, with energy management or monitoring platform providers, power grid operators and energy provider companies; to gain (real-time) access to additional measured PV power output data, integrate power output predictions into power/grid management platforms, and co-fund the projects.
- Extending research partnerships for knowledge and expertise exchange.
- API access to consumers' power usage to develop models for smart grid and load power management.

Please contact us via meyer@mit-coaching.com in case you have any questions, inputs, and suggestions to our analysis, or want to discuss potential collaborations. Thank you!

Acknowledgements

The authors thank the Kachelmann GmbH for providing us with meteorological forecast data for our 3 sites as well as valuable discussion on impacts of different meteorological factors, the EW Höfe AG for providing us with PV power output data for their two sites and valuable inputs on our work, and Girsberger Informatik AG for valuable inputs on energy grids and applied energy predictions.

References

[**AlDahidi2018**] Al-Dahidi, Sameer & Ayadi, Osama & Adeeb, Jehad & Alrbai, Mohammad & Qawasmeh, Bashar. (2018). Extreme Learning Machines for Solar Photovoltaic Power Predictions. Energies. 11. 2725. 10.3390/en11102725.

[**Ayon2017**] Ayón, X.; Moreno, M.Á.; Usaola, J. (2017) Aggregators' Optimal Bidding Strategy in Sequential Day-Ahead and Intraday Electricity Spot Markets. Energies 2017, 10, 450, doi:10.3390/en10040450.

[**Bouzerdoum2013**] Bouzerdoum, M. & Mellit, Adel & Massi Pavan, Alessandro. (2013). A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. Solar Energy. 98. 226–235. 10.1016/j.solener.2013.10.002.

[**Ciapala2018**] Jurasz, J.; Ciapała, B. Solar–hydro Hybrid Power Station as a Way to Smooth Power Output and Increase Water Retention. Sol. Energy 2018, 173, 675–690, doi:10.1016/j.solener.2018.07.087.

[**Clack2017**] Clack, C. T. M. (2017). Modeling Solar Irradiance and Solar PV Power Output to Create a Resource Assessment Using Linear Multiple Multivariate Regression. Journal of Applied Meteorology and Climatology 56, 1, 109-125.

[**Crisosto2018**] Crisosto, Cristian & Hofmann, Martin & Mubarak, Riyad & Seckmeyer, Gunther. (2018). One-Hour Prediction of the Global Solar Irradiance from All-Sky Images Using Artificial Neural Networks. Energies. 11. 2906. 10.3390/en11112906.

[**DaSilva2012**] Da Silva Fonseca, J.G.; Oozeki, T.; Takashima, T.; Koshimizu, G.; Uchida, Y.; Ogimoto, K. (2012) Use of Support Vector Regression and Numerically Predicted Cloudiness to Forecast Power Output of a Photovoltaic Power Plant in Kitakyushu, Japan. Proc. Photovolt. Res. Appl.

[**DeBock2014**] Bock, Veerle & De Backer, Hugo & Van Malderen, Roeland & Mangold, Alexander & Delcloo, Andy. (2014). Relations between erythemal UV dose, global solar radiation, total ozone column and aerosol optical depth at Uccle, Belgium. Atmospheric Chemistry and Physics Discussions.

[**Dorokhov2022**] Dorokhova, M., Martinson, Y., Ballif, C., & Wyrsch, N. (2021). Deep reinforcement learning control of electric vehicle charging in the presence of photovoltaic generation. Applied Energy, 301, 117504.

[**Giorgi2015**] Giorgi, M.G.D.; Congedo, P.M.; Malvoni, M.; Laforgia, D. Error analysis of hybrid photovoltaic power forecasting models: A case study of mediterranean climate. Energy Convers. Manag. 2015, 100, 117–130.

[**King2004**] King, David & Galbraith, Gary & Boyson, William. (2007). Performance Model for Grid-Connected Photovoltaic Inverters. Sandia Natl. Lab. 38.

[**Lee2019-1**] Lee, Donghun & Kim, Kwanho. (2019). Recurrent Neural Network-Based Hourly Prediction of Photovoltaic Power Output Using Meteorological Information. Energies. 12. 215. 10.3390/en12020215.

[**Lee2019-2**] Lee, S., & Choi, D. H. (2019). Reinforcement learning-based energy management of smart home with rooftop solar photovoltaic system, energy storage system, and home appliances. Sensors, 19(18), 3937.

[**Li2019**] Li, & Wang, & Zhang, Min & Xin, & Zhaodong, Liu. (2019). Recurrent Neural Networks Based Photovoltaic Power Forecasting Approach. Energies. 12. 2538. 10.3390/en12132538.

[**Huang2019**] W. Huang, C. Zhang, X. Zhang, J. Meng, X. Liu and B. Yuan, "Photovoltaic Power Prediction Model Based on Weather Forecast," 2019 IEEE Sustainable Power and Energy Conference (iSPEC), Beijing, China, 2019, pp. 1596-1600.

[**Kou2013**] Kou, Jiahao & Liu, Jun & Li, Qifan & Fang, Wanliang & Chen, Zhenhuan & Liu, Linlin & Guan, Tieying. (2013). Photovoltaic power forecasting based on artificial neural network and meteorological data. IEEE Region 10 Annual International Conference, Proceedings/TENCON. 1-4. 10.1109/TENCON.2013.6718512.

[**Mellit2005**] Mellit, Adel & Benghanem, Mohamed & Bendekhis, M.. (2005). Artificial neural network model for prediction solar radiation data: Application for sizing stand-alone photovoltaic power system. 2005 IEEE Power Engineering Society General Meeting. 1. 40 - 44 Vol. 1.

[**Mouatasim2018**] El Mouatasim, Abdelkrim & Darmane, Yassin. (2018). Regression Analysis of a Photovoltaic (PV) System in FPO. AIP Conference Proceedings. 2056. 0200081-02000811. 10.1063/1.5084981.

[**O'Grady2022**] O'Grady, J. (2022). Optimizing Solar Energy Production with Reinforcement Learning. https://www.jackogrady.me/reinforcement-learning-solar/research-summary. Last accessed: 19.07.2022.

[**Pelland2013**] Pelland, Sophie & Remund, Jan & Kleissl, Jan & Oozeki, Takashi & De Brabandere, Karel. (2013). Photovoltaic and Solar Forecasting: State of the Art.

[**Ramenah2018**] Ramenah, H. & Casin, Philippe & Ba, Mouhamadou Moustapha & Benne, Michel & Tanougast, Camel. (2018). Accurate determination of parameters relationship for photovoltaic power output by Augmented Dickey Fuller test and engle granger method. AIMS Energy. 6. 19-48. 10.3934/energy.2018.1.19.

[**Reindl2017**] Reindl, Thomas & Walsh, Wilfred & Yanqin, Zhan & Bieri, Monika. (2017). Energy meteorology for accurate forecasting of PV power output on different time horizons. Energy Procedia. 130. 130-138. 10.1016/j.egypro.2017.09.415.

[**Sfetsos2000**] Sfetsos, Thanasis & Coonick, A.H.. (2000). Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. Solar Energy. 68. 169-178. 10.1016/S0038-092X(99)00064-X.

[**Shi2015**] Jie Shi, Wei-Jen Lee, Yongqian Liu, Yongping Yang and Peng Wang, "Forecasting power output of photovoltaic system based on weather classification and support vector machine," 2011 IEEE Industry Applications Society Annual Meeting, Orlando, FL, 2011, pp. 1-6.

[**Sunday2015**] Sunday, V. E., Simeon, O., & Anthony, U. M. (2017). Multiple linear regression photovoltaic cell temperature model for PVSyst simulation software. International Journal of Theoretical and Applied Mathematics, 2(2), 140.

[**Theocharidies2018**] Theocharides, S., Makrides, G., Theristis, M., Almonacid, F., Fernández, E. F., & Georghiou, G. E. Short-term photovoltaic power forecasting based on artificial neural networks: a numerical weather prediction-free approach.

[**William2018**] William F. Holmgren, Clifford W. Hansen, and Mark A. Mikofski. "Pvlib Python: a python package for modeling solar energy systems." Journal of Open Source Software, 3(29), 884, (2018). https://doi.org/10.21105/joss.00884

[**Wolff2014**] Wolff, B., Kühnert, J., Lorenz, E., Kramer, O., & Heinemann, D. (2016). Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data. Solar Energy, 135, 197-208.

[**Zhong2017**] Zhong, Z., Yang, C., Cao, W., & Yan, C. (2017). Short-term photovoltaic power generation forecasting based on multivariable grey theory model with parameter optimization. Mathematical Problems in Engineering, 2017.

[**Zhong2018**] Zhong, J., Liu, L., Sun, Q., & Wang, X. (2018). Prediction of photovoltaic power generation based on general regression and back propagation neural network. Energy Procedia, 152, 1224-1229.

[**Zhou2017**] Zhou, Y.; Wang, C.; Wu, J.; Wang, J.; Cheng, M.; Li, G. (2017). Optimal Scheduling of Aggregated Thermostatically Controlled Loads with Renewable Generation in the Intraday Electricity Market. Appl. Energy 2017, 188, 456–465