20th International Conference on Renewable Energies and Power Quality (ICREPQ'22)
Vigo (Spain), 27th to 29th July 2022
Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, Volume No.20, September 2022

ANN-Based Large-Scale Cooperative Solar Generation Forecasting

Lejla Pašić[1,3], Azra Pašić[2], Alija Pašić[2] and István Vokony[1]

[1] Department of Electric Power Engineering
Faculty of Electrical Engineering and Informatics
Budapest University of Technology and Economics
Műegyetem rkp. 3., H-1111 Budapest, Hungary
e-mail: **pasic.lejla@edu.bme.hu**, **vokony.istvan@vet.bme.hu**

[2] Department of Telecommunications and Media Informatics
Faculty of Electrical Engineering and Informatics
Budapest University of Technology and Economics
Műegyetem rkp. 3., H-1111 Budapest, Hungary
e-mail: **pasic.azra@edu.bme.hu, pasic@tmit.bme.hu,**

[3] ELKH-BME Information Systems Research Group,
1117, Magyar Tudósok krt. 2, Budapest, Hungary

Abstract. In this work we introduce the concept and method of so-called cooperative solar generation forecasting, where geographically close data sources are utilized in order to improve forecasting accuracy. We devised and examined various large-scale one-hour-ahead artificial neural networks based solar generation forecasting scenarios to prove the benefits of cooperation. The introduced cooperative solar generation forecasting method showed significant improvement in forecasting accuracy, especially when combined with previous generation data, where a root mean square error reduction of at least 50% could be achieved in the majority of cases. We believe these results point to a scientific and economical benefit of international cooperation in solar generation forecasting.

Keywords. Renewable energy, solar generation, forecasting, artificial neural networks, input optimization

1. Interest and Objectives

At present, the energy production system still mostly relies on non-renewable sources, such as fossil fuel power plants. These types of power plants are known to be harmful to the environment, since they are responsible for the emission of greenhouse gases, which lead to global warming [1]. Renewable energy does not only offer an alternative source for clean energy – it could also help to reduce a possible energy crisis (which is looking likelier day by day) by playing a key role in meeting future electricity demands. Solar energy is considered to be one of the most promising forms of renewable energy sources [1].

The intermittent nature of renewable energy generation poses a significant operational challenge to power systems, which traditionally operate under deterministic rules. Generation forecasting is considered to be a good strategy for the mitigation of these effects [2]. In addition to the operational benefits, generation forecasting can have significant economic advantages, especially in cases where short term electricity trading is possible [3]. Overall, a reliable forecasting method for renewable energy production would have a very positive influence on the reduction of the integration costs, the decrease of the average annual operating costs and the minimizing of the reserve shortfalls [4]. This study examines the possibilities of artificial neural networks based, less explored [5] large-scale generation forecasting. With an optimisation of inputs, we aim to improve the one-hour-ahead short term solar generation forecast accuracy. For this purpose, a cooperative forecasting approach was developed and analysed in several test scenarios, with very promising results.

This paper is organized the following way. Section 2 discusses the necessity of solar power generation and the typically used forecasting horizons, while Section 3 introduces the data set and the methodology. The experimental results are presented in Section 4, and a short summary is offered in Section 5.

2. Solar Power Generation Forecasting

Solar power forecasting plays an important role in the operation, scheduling and balancing of electricity production by standalone photovoltaic (PV) plants, but also in large grid-interconnected solar PV plants [6].

Solar generation forecasting can help the grid operators to manage the system more efficiently, for example when deciding whether to commit or decommit generators to accommodate changes in generation and react to extreme events [4]. This increases system reliability, since reliability is dependent on the system's ability to accommodate expected and unexpected changes and disturbances while maintaining quality and continuity of service for the customers [4]. Solar generation forecasting is not only important for the day-to-day grid operation, but it is also vital to estimate the reserves, schedule the power system, manage congestion and storage optimally, and also to trade in the electricity market. Besides helping to overcome operational challenges resulting from the volatility and uncertainty of the solar energy sources, forecasting has a significant monetary impact too.

Forecasts grant the possibility to reduce the amount of operating reserves needed for the power system, hence reducing the system balancing costs. With forecasts, grid operators can schedule and operate other generating capacity efficiently, reducing fuel consumption, operation and maintenance costs and gas emissions as compared to simply dealing with the energy generated by variable sources without any beforehand knowledge [4]. In literature, we can also find specific examples of how forecasting can affect costs and profits. For instance, [4] shows how a six-million-dollar saving could be achieved in one year as a result of forecasting [7]. Moreover, it is shown how the use of production forecasts reduced operating costs over the course of one year by up to 14%, or 5 billion US dollars per year, which was achieved through a reduction of operating costs by 12–20 $/MWh (in [8]).

Solar generation forecasting models are usually classified into three categories: statistical, physical and hybrid [6]. Statistical methods have the advantage that they do not require the internal state information of the system to be modeled. Artificial intelligence (AI) is often used for complex or non-linear data, particularly for data arranging, pattern recognition, simulation, and optimization. It is often utilized for solar generation forecasting, mainly based on Artificial Neural Networks (ANNs) [9]. For example, [10] introduces an ANN model based on Extreme Learning Machine (ELM) trained to forecast solar photovoltaic power.

Forecasting methods can be classified according to the forecast horizon into four categories: (i) very short-term forecasting (a time horizon of a few seconds to a few minutes); (ii) short-term forecasting (up to three days ahead); (iii) medium-term forecasting (from a few days to one week ahead); and (iv) long-term forecasting from a few months to even several years ahead [11]. The forecasting horizon categories usually have specific applications. Short-term forecasting is mainly used for the control of power system operations, economic dispatch, unit commitment, etc. Conversely, medium and long-term horizons are usually used for the maintenance and the planning of PV plants.

Machine learning methods are believed to be best suited for short term forecasting [11]. The forecasting time horizon used in this study is the one-hour-ahead forecasting, which also falls into the short term category. Table I presents the start-up time of some electricity production plant types. It shows that these have a relatively long start-up time, while quicker responding technologies have a start-up time within an hour. Hence a one-hour-ahead forecast of the solar energy generation could be helpful in the management of the energy resources.

Table I - Characteristics of Electricity Production Plants [4]

Type	Size (MW)	Start-up Time (h)
Nuclear power plant	400–1300 per reactor	40 (cold)–18 h (hot)
Steam thermal plant	200–800 per turbine	11–20 h (cold)– 5 h (hot)
Fossil-fired power plants	1–200	10 min–1 h
Combined-cycle plant	100–400	1–4 h
Hydropower plant	50–1300	5 min
Combustion turbine (light fuel)	25	15–20 min
Internal combustion engine	20	45–60 min

3. Data Set and Methodology

This study implemented and examined an Artificial Neural Network (ANN) based solar generation forecasting algorithm in different forecasting scenarios. For the ANN algorithm, a Multi-layer Perceptron Regressor (MLPRegressor) architecture was applied from Python's scikit-learn library, which optimizes the squared-loss using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) or stochastic gradient descent approach [12]. According to the analysis in [9], artificial intelligence approaches are widely used and often outperform the traditional methods for short term solar generation forecasting. Most of the research is focused on forecasting at a single location, while little work has been done on regional models [5]. This paper examines cooperative one-hour-ahead generation forecasting on large-scale models (encompassing entire countries), and analyses their fluctuation in performance depending on their input data. It reveals how the one-hour-ahead solar generation forecasting accuracy can be improved with cooperation between different data sources.

Six large-scale open-source data sets were used from six different countries, namely Austria, the Czech Republic, Switzerland, Spain, France and Italy. The examined time period for all data sets was between 2017 and 2019 (where 2017 and 2018 made up the training data set, and 2019 was the test data set). Finding the optimal length of the training data for a solar generation forecasting model is mainly an experimental process, the data of two to three years is considered to be the most suitable for model training to get the best results [4]. We obtained the data from the Open Power System Data platform [13]. After pre-processing and cleaning, the data was converted into pandas dataframes, which were split into training and testing sets. The training set contained the weather and timeseries generation data for two years (66.66% of the data whole data set), and the test set contained the weather and timeseries generation data for one year (33.33% of the data set).

For the validation of the results of forecasting, the root mean square error (RMSE) of each algorithm on the test data set was examined, while taking algorithm runtime into consideration as well. The RMSE (see Equation 1) is a popular regression verification measure, which is defined as the square root of the mean of the squared differences between corresponding elements of the forecast (z_{fi}) and the observation (z_{oi}) – in this case, the difference between predicted generation and real generation [14]. Runtime is the time needed for the training of the algorithm.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(z_{fi} - z_{oi}\right)^2}{N}} \qquad (1)$$

4. Experimental Results

The six data sets were assigned to two groups – one group consists of the Austrian (AT), Czech (CZ) and Swiss (CH) data, while the other group includes the Spanish (SE), French (FR) and Italian (IT) data. The data sets (countries) in each group can share data between each other in the cooperative scenarios (see Figure 1). The dotted line connects the data sets which can cooperate with each other, but only in the cooperative scenarios.

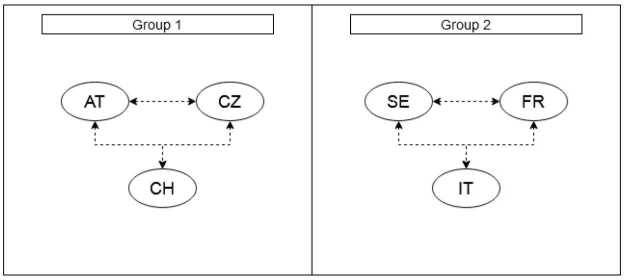

Figure 1: Data set groups

The output of our ANN algorithm was always the one-hour ahead prediction of the solar generation (in MW) for the examined data set, while we varied the input according to several test case scenarios. In our base case, we used the temperature ($^{\circ}C$), the direct horizontal radiation (W/m^2) and the diffuse horizontal radiation (W/m^2) – henceforth referred to collectively as *weather data* – of the examined data set as the input. In the cooperative forecasting scenarios, the weather data of each country in the respective group was applied. In the 24 h cooperative forecasting scenario, the input was the one-hour-before weather data of the examined country, and the 24 h before weather data from the cooperating countries (countries in the same group). It was also examined how adding the one-hour-before generation data to the weather as input in the next scenario impacts forecasting results. Accordingly, in the cooperative scenario with added generation data, the input was the one-hour-before weather and generation data for all countries in the group. In the 24 h one-hour-ahead cooperative generation forecasting with generation data, the input was the one-hour-before weather and generation data of the examined country and the 24-hour-before weather and generation data of the cooperating countries. An example is shown in Table II, where one time step (2019. 05. 18. 13:00) of a forecast on one data set (Spanish data set) is presented for all examined scenarios. The different input parameters can be observed here, while naturally, the output is the solar generation forecast for the mentioned time step.

Table II – Examples of input and output parameters in different scenarios (Group 2, Spanish data set)

Example input parameters	Corresponding output parameter (forecasted value)
One hour ahead forecasting	
Spanish weather data from 2019. 05. 18. 12:00	Solar energy generation in Spain for 2019. 05. 18. 13:00
One-hour-ahead cooperative forecasting	
Spanish, French and Italian weather data from 2019. 05. 18. 12:00	Solar energy generation in Spain for 2019. 05. 18. 13:00
One-hour-ahead cooperative forecasting (24 h)	
Spanish weather data from 2019. 05. 18. 12:00 French and Italian weather data from 2019. 05. 17. 12:00	Solar energy generation in Spain for 2019. 05. 18. 13:00
One-hour-ahead forecasting with generation data	
Spanish weather and generation data from 2019. 05. 18. 12:00	Solar energy generation in Spain for 2019. 05. 18. 13:00
One-hour-ahead cooperative forecasting with generation data	
Spanish, French and Italian weather and generation data from 2019. 05. 18. 12:00	Solar energy generation in Spain for 2019. 05. 18. 13:00
One-hour-ahead cooperative forecasting with generation data (24 h)	
Spanish weather and generation data from 2019. 05. 18. 12:00 French and Italian weather and generation data from 2019. 05. 17. 12:00	Solar energy generation in Spain for 2019. 05. 18. 13:00

The experimental results of the two groups, Group 1 and Group 2, are presented in Table III and Table IV, respectively. Within the tables, the achieved test values – root mean square errors (RMSE) and algorithm training time – are given for each data set in each scenario separately.

Table III – Solar generation forecasting (Group 1)

	AT	CZ	CH
One hour ahead forecasting			
RMSE	61.4708	140.8735	29.1212
Runtime [s]	2.7989	6.5168	5.7042
One hour ahead cooperative forecasting			
RMSE	53.8516	105.4581	22.0411
Runtime [s]	8.5347	12.3986	15.9387
One hour ahead cooperative forecasting (24 h)			
RMSE	50.8506	95.9488	21.1447

Runtime [s]	9.2895	13.4802	16.2596
One hour ahead forecasting with generation data			
RMSE	52.6571	114.8401	22.2997
Runtime [s]	5.9555	5.8936	5.1496
One hour ahead cooperative forecasting with generation data			
RMSE	29.4456	55.1854	10.0242
Runtime [s]	9.9245	10.0438	11.3595
One hour ahead cooperative forecasting with generation data (24 h)			
RMSE	24.5071	48.4023	9.5708
Runtime [s]	6.6001	8.1927	9.2013

Table IV – Solar generation forecasting (Group 2)

	ES	FR	IT
One hour ahead forecasting			
RMSE	626.6078	644.5594	948.8579
Runtime [s]	4.9257	2.9964	3.3367
One hour ahead cooperative forecasting			
RMSE	535.5095	529.4183	474.3324
Runtime [s]	8.4420	9.5291	8.1970
One hour ahead cooperative forecasting (24 h)			
RMSE	540.1418	520.0801	561.1038
Runtime [s]	7.9667	6.5511	12.9118
One hour ahead forecasting with generation data			
RMSE	538.4999	627.4544	885.8472
Runtime [s]	8.2998	5.2325	3.9643
One-hour-ahead cooperative forecasting with generation data			
RMSE	340.4299	451.4564	285.0834
Runtime [s]	4.6663	4.1397	5.3836
One-hour-ahead cooperative forecasting with generation data (24 h)			
RMSE	330.9977	461.3612	377.8792
Runtime [s]	5.2398	3.0809	7.6705

The decrease seen in the RMSE (in percent) compared to the base case for different scenarios is illustrated in Figure 2 and Figure 3 for Group 1 and Group 2, respectively. These figures (and the tables) show that a small cooperation between three countries can already significantly improve the solar generation forecasting for each data set, as in all cases, the cooperation decreased the error at least by 12%. The smallest percentual decrease was seen on the Austrian data set (12%), which in this case means that with cooperation only, we could decrease the error of the forecast by approximately 8 MW/hour on average. The largest percentual RMSE decrease was 50%, which was observed on the Italian data set.

The data closest in time to the forecast does not necessarily bring the best results – when using the 24-hour-before data from the cooperating countries, the accuracy was improved in all cases for Group 1. However, the data in Group 2 proved that the one-hour-before cooperative data yielded better results than the 24 h before data in the case of the Italian data set. So using the example from Table II, we can say that the forecast would improve on most of the data sets if the "One hour ahead cooperative forecasting (24 h)" scenario is used – i.e., if, for example, the Spanish weather data from 2019. 05. 18. 12:00 and French and Italian weather data from 2019. 05. 17. 12:00 were used instead of the Spanish, French and Italian weather data from 2019. 05. 18. 12:00 for the forecasting of the solar energy generation in Spain for 2019. 05. 18. 13:00. However, this in not necessarily the best approach for all data sets and cooperation. This is not surprising, since the geographical area can strongly influence the optimal input parameters for the forecasting model [6]. However, it can be concluded that with the individual optimisation of the utilized cooperative data sample-time the cooperative forecasting accuracy can be further improved.

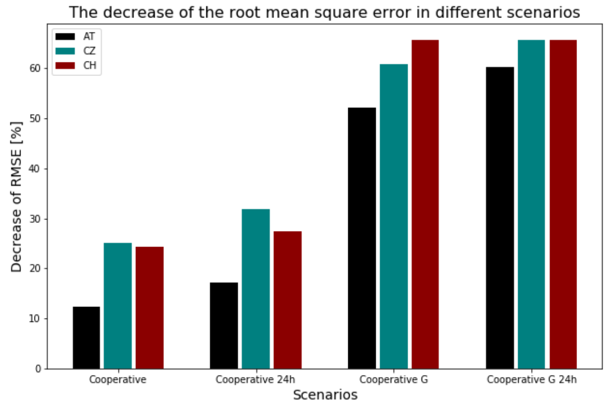

Figure 2: The decrease of RMSE in Group 1

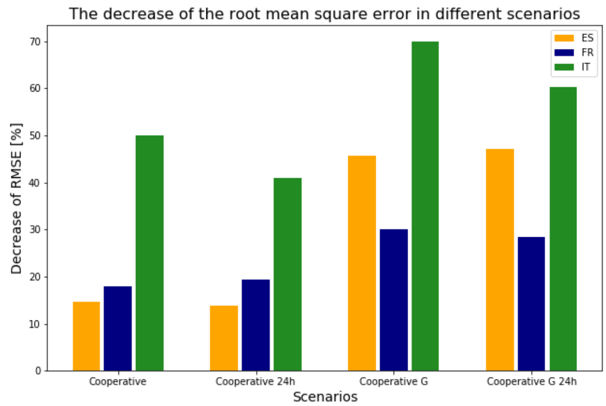

Figure 3: The decrease of RMSE in Group 2

In our further input optimization experiments, it was discovered that adding the previous generation data to the input of the cooperative solar generation forecasting can greatly improve the forecasting accuracy. With this added step, an even higher improvement in forecasting accuracy can be achieved for all examined data sets. In Group 1, the accuracy improved at least by 52% compared to the baseline case (for all test cases). In Group 2, the smallest improvement was on the French set, around 29%, but the highest improvement decreased the error by 69%.

It can also be observed that while all the data sets benefit from the cooperation, the extent of this benefit is not equal. As Figure 2 and Figure 3 show, in Group 1 the amount of the error decrease (in percent) for the separate data sets is much closer than in Group 2.

Naturally, in most cases there was a runtime trade-off when using the cooperative forecasting approach due to the increased data volume, but interestingly, in some cases (e.g. Spain one-hour-ahead forecasting with generation data) the cooperation also decreased the training time, most likely due to quicker convergence of the ANN.

5. Conclusion

This paper discussed how solar photovoltaic power forecasting can play a crucial role in the planning and modelling of the solar photovoltaic plants, as well as in the managing of the power demand and supply [9]. For this purpose, we implemented an artificial neural network based one-hour-ahead solar generation forecasting model, which was tested on different large-scale data sets in different scenarios. A custom-made ANN-Based Large-Scale Cooperative Solar Generation Forecasting technique was introduced, which was optimized using different input parameters.

With cooperative data sharing introduced between previously separately functioning (geographically close) data sets, a great improvement in forecasting accuracy can be achieved. By cooperative data sharing only, the RMSE of the forecasting decreased at least by 12%, but an error reduction of even 50% could be achieved. With an optimization of the utilized cooperative data sample-time and through the inclusion of previous generation data as input parameters, the error of cooperative forecasting can be further decreased – in our test cases, the RMSE reduction was as high as 29 to 69% compared to the baseline case.

Further directions of improvement for our work include finding the optimal number of data sets for a cooperative group, as well as establishing a set of criteria for the optimal selection of specific data sets for respective cooperative groups.

Acknowledgement

The research leading to these results was supported by Ericsson and the High Speed Networks Laboratory (HSNLab). Project no. 137698 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the PD_21 funding scheme. This paper was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. In addition we would like to thank Monodon Solutions Ltd. for their support.

References

[1] Atika Qazi, Fayaz Hussain, Nasrudin ABD Rahim, Glenn Hardaker, Daniyal Alghazzawi, Khaled Shaban, and Khalid Haruna. Towards sustainable energy: a systematic review of renewable energy sources, technologies, and public opinions. IEEE Access, 7:63837–63851, 2019.

[2] Orwig, K. D., Ahlstrom, M. L., Banunarayanan, V., Sharp, J., Wilczak, J. M., Freedman, J., ... & Marquis, M. (2014). Recent trends in variable generation forecasting and its value to the power system. IEEE Transactions on Sustainable Energy, 6(3), 924-933.

[3] Adel Mellit, Alessandro Massi Pavan, Emanuele Ogliari, Sonia Leva, and Vanni Lughi. Advanced methods for photovoltaic output power forecasting: A review. Applied Sciences, 10(2):487, 2020.

[4] Notton, G., & Voyant, C. (2018). Forecasting of intermittent solar energy resource. In Advances in Renewable Energies and Power Technologies (77–114). Elsevier

[5] Adel Mellit, Alessandro Massi Pavan, Emanuele Ogliari, Sonia Leva, and Vanni Lughi. Advanced methods for photovoltaic output power forecasting: A review. Applied Sciences, 10(2):487, 2020.

[6] Singla, P., Duhan, M., & Saroha, S. (2021). A comprehensive review and analysis of solar forecasting techniques. Frontiers in Energy, 1–37.

[7] Parks, K. (2011, February). Xcel energy/NCAR wind energy forecasting system. In Talk on the UWIG Forecasting Workshop, Albany (US) (23–24).

[8] Energy, G. E. (2010). Western wind and solar integration study (pp. 1-536). NREL/SR-550-47434. Golden, Colorado: National Renewable Energy Laboratory.

[9] Sobrina Sobri, Sam Koohi-Kamali, and Nasrudin Abd Rahim. Solar photovoltaic generation forecasting methods: A review. Energy Conversion and Management, 156:459–497, 2018.

[10] Teo, T. T., Logenthiran, T., & Woo, W. L. (2015, November). Forecasting of photovoltaic power using extreme learning machine. In 2015 IEEE Innovative Smart Grid Technologies-Asia (ISGT ASIA) (1–6). IEEE.

[11] Mellit, A., Massi Pavan, A., Ogliari, E., Leva, S., & Lughi, V. (2020). Advanced methods for photovoltaic output power forecasting: A review. Applied Sciences, 10(2), 487.

[12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12: 2825–2830, 2011

[13] Open power system data. https://open-power-system-data.org/.

[14] Sameer Al-Dahidi, Osama Ayadi, Mohammed Alrbai, and Jihad Adeeb. Ensemble approach of optimized artificial neural networks for solar photovoltaic power prediction. IEEE Access, 7: 81741–81758, 2019.