

# Optimization of a solar irradiation forecasting tool based on artificial intelligence

F. Rodríguez<sup>1,2</sup>, A. Galarza<sup>1,2</sup> and L. Fontán<sup>1,2</sup>

<sup>1</sup> Ceit, Manuel Lardizabal 15, 20018 Donostia/San Sebastián, Spain.  
 Phone number:+0034 943 212800, e-mail: [frlalanne@ceit.es](mailto:frlalanne@ceit.es)

<sup>2</sup> Universidad de Navarra, Tecnun, Manuel Lardizabal 13, 20018 Donostia/San Sebastián, Spain

**Abstract.** In current electric markets, where many stakeholders can take part, power network operators need accurate predictions of the energy generated by intermittent renewable sources in order to control the whole system. Therefore, the capacity to accurately forecast solar irradiance is key when it comes to the large-scale integration of solar energy generators in the traditional network. One of the challenges, however, consists of providing accurate very short-term predictions (minutes ahead) due to the variability of solar irradiance caused by different meteorological phenomena.

This study addresses this need for very short-term forecasts through the development of an irradiance prediction scheme for 10 minutes ahead. The irradiance prediction algorithm is based on a parallel combination of two different layer recurrent networks and has been trained with a two-year historical database of solar irradiance. The accuracy of the proposed tool has been validated through forecasting a whole year using data that is not in the database used in the training step. This tool was then used to forecast the irradiance in two Spanish locations with different weather conditions to analyse whether the accuracy changes. The accuracy between predicted and actual values demonstrates that this tool outperforms similar forecasters.

## Key words

Solar irradiance, Forecasting, Artificial intelligence, Renewable sources control.

## 1. Introduction

Electrical energy consumption has increased in recent decades due to the development of technology. Energy has been traditionally provided by non-renewable sources as thermal power, but requirements for decreasing pollution and the greenhouse effect has allowed new cleaner energy technologies to appear.

Solar photovoltaic and wind energy are considered clean, cheap and limitless power generators and have arisen as a solution to the global energy crisis, which explains the increased number of photovoltaic and wind power plants in recent years [1].

However, if the goal is to connect renewable energies to the traditional network on a large scale, some challenges need to be overcome. In terms of solar energy generation, there are two main issues that must be addressed, namely variability and location dependence [2].

Due to the fact that these generators depend on solar irradiance to produce energy, sudden changes in this meteorological parameter will affect overall generation. This problem can be solved in two different ways: by improving forecasting algorithms or increasing the number of energy storage devices in order to provide the difference between the forecasted and actual energy produced by solar generators. In addition, solar resources are geography-dependant and unlike fossil fuels such as coal or biomass resources, it cannot be transported to a power generator; instead new transmission lines are needed to connect renewable generators to the main grid [2]. This study addresses the variability issue through a forecaster in order to improve solar irradiance prediction.

A review of the available literature shows that the prediction horizon of the forecaster will change depending on the information that is needed to make a decision. Authors currently differentiate between four different horizons: very short-term, which is related to a few minutes' horizon and real-time dispatch; short-term, which involves a horizon of a few hours to days ahead and economic dispatch; medium-term, which entails from a few days to weeks ahead and maintenance scheduling; and long-term, where the associated horizon is from a few months to years and is used in solar power plant planning [3]. Table I summarizes the information related to the different forecast horizons.

Table I. – Summary of prediction horizons

FORECASTER	HORIZON	APPLICATION
Very short-term	Few minutes	Real-time dispatch
Short-term	Few hours	Economic dispatch
Medium-term	Few days	Maintenance plan
Long-term	Few months	Solar plant planning

To ensure the integration of solar power plants into the traditional network and reduce the variability associated with solar technology, this study proposes a simple and accurate method for reducing variability. To that end, a very short-term forecasting tool for solar irradiation, with a 10-minutes ahead horizon, was developed.

## 2. Related works on irradiation forecasting

Methods for predicting solar irradiation can be classified into the following categories: persistence method, physical techniques, statistical approaches and artificial intelligence (AI).

### A. Persistence method

The persistence method is based on the extrapolation of current conditions to future horizons. This is the simplest forecasting method and it is usually used as a reference model for new short-term forecasters. For solar irradiation predictions, persistence supposes clear sky conditions and no variation in solar irradiance between time steps. Thus, this method is not able to predict sudden changes and it is used only as a comparative method [4]-[6].

### B. Physical techniques

Physical forecasting techniques are divided in two groups: techniques related to numerical weather prediction (NWP) and those which use satellite imagery. NWP methods are based on a number of equations that characterize the movement of the fluids that are involved in weather. If these equations are computerized, numerical methods are used and boundary conditions are applied, and an NWP is obtained. References [7] and [8] show examples where this method is used for forecasting. These techniques, however, are usually avoided due to their complex programming.

Regarding satellite and sky imagery, cloud cover and cloud depth are the key aspects that affect the solar irradiance that is received on the earth's surface. This method is related to the study of cloud structures in their initial time steps and extrapolating movement in order to predict their position in the future [9]. This technique can be used only if there is an available satellite, something that it is not always possible [10], [11].

### C. Statistical approaches

These methods have been traditionally implemented as an alternative to physical techniques. These methods are based on time-series forecasting models. Therefore, a huge historical database is needed to establish the relationship between the variables that characterize the process and to predict future values. These approaches include: autoregressive (AR), moving average (MA) and a combination of these (ARMA).

An example of this type of method is found in [12] is, which can forecast irradiance values with reasonable accuracy. However, these methods will make better predictions when there is a linear relationship between the variables involved.

### D. Artificial intelligence

Artificial intelligence is based on the development of machines that are able to behave like humans. AI has become a key challenge for the industry of the future, known as Industry 4.0. Much effort has gone into introducing this technology in current products. Some of the issues that are addressed through AI are learning, planning and perception.

Solar irradiation forecasting has been addressed through different AI techniques: artificial neural networks (ANNs), which simulate human brain abilities [2]; wavelet neural networks (WNN), which consist of a combination of wavelet theory and neural networks [5]; support vector machines (SVM) [13] or fuzzy logic [14], which are associated with the generalization of classical logic. Moreover, there are techniques that are based on the combination of these methods [15].

Although all these techniques could be used for solar irradiation forecasting, each method has its strengths and weakness. For instance, fuzzy logic approaches are not able to learn directly from a historical database, and wavelet theory-based methods need a NN to overcome the challenge of modelling complex structures and reduce the required computational cost [2]. The strength of ANNs and SVMs stem from their capacity to model non-linear processes and their ease of implementation, in addition to providing accurate very-short term forecasts.

Of these two methods, an ANN has been chosen to develop this forecaster. ANNs outperform SVMs in the vast majority of cases if just a historical database is used and no optimized decision making step is added to an SVM forecaster. ANNs can be divided in two main categories: feed-forward neural networks (FFNNs) and recurrent neural networks (RNNs). Since including the characteristic feedback loops of RNNs influences an ANN's learning capability and can produce better results in time series prediction, the developed forecaster is based on an RNN network [16].

Finally, when developing a solar irradiation forecasting tool, it is necessary to define certain parameters of the RNN as the number of hidden layers, the number of neurons at each layer or the delay related to the feedback loop. Figure 1 shows the layout of an RNN.

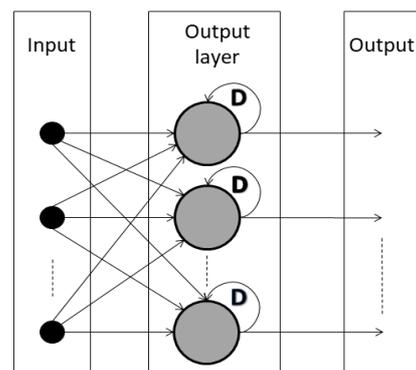


Fig. 1. A recurrent neural network (RNN) layout

### 3. The proposed prediction technique

This section summarizes the steps of the technique proposed for solar irradiation forecasting. More details can be found in [17].

#### A. Input parameter selection and database length

Before the forecasting tool can be developed, it is necessary to analyse which parameters characterize the process as they will constitute the inputs. Through this process, variables with little or no contribution to the network output are removed [18].

The analysis was performed using the Pearson Correlation analysis tool in Excel®. Correlation tests are typically used for analysing whether or not a true relationship exists between two parameters. After running different Pearson Correlation tests, we decided to use the following parameters as inputs: season, time and irradiation values for the 24 hours preceding the prediction.

Regarding historical database length, the available literature on solar irradiation forecasting disagrees on what the optimal size is. In spite of this, the vast majority of the references [8], [19], [20] use at least a one-year historical database for their prediction methods. Therefore, to ensure proper learning in the forecaster, a two-year database covering 2015 and 2016 was used for the training step.

#### B. Fixing ANN parameters

To fix the ANN parameters and obtain the optimal architecture, data from the first four months of 2015 were used to examine different designs. The root mean squared error value of the predictions made by those designs were analysed to select the optimal one. To determine the delay, the number of hidden neurons was held constant and the value of the delay was changed incrementally in each test. The method for fixing the number of hidden neurons was the same, but in this case, the delay was held constant [17].

After running several tests, it was concluded that an ANN with the same architecture and input values will make similar predictions with slight differences. To ensure that the optimal design was selected, the test for each architecture was repeated five times.

#### C. Learning algorithm and validation step

Forecasters have to be able to obtain reasonable predictions from previously unseen values. This goal is achieved by having correctly trained the ANN; ensuring proper train will allow the network to produce accurate results in areas like prediction and pattern recognition [21].

An ANN is trained with learning algorithms, a process that is based on modifying random initial connections of the neurons in an orderly way until the minimal error is obtained [16]. A review of the literature shows there is agreement on using the Levenberg-Marquardt (LM) as the learning algorithm due to its robustness [5], [22].

There are two possible approaches to the validation step. While some authors use a single database divided into three different pieces for training, testing and validation [2], [4], other authors use two different databases, one for training and one for validating [6], [22]. For this study's validation step, the second option was chosen.

#### D. Accuracy metrics

Several error metrics are used in the literature [3], [5], [19] to evaluate the performance of new prediction models. In order to analyse the accuracy of the developed tool and be able to compare it with previous related works, root mean square error (RMSE) was chosen.

The RMSE calculates the square root of the sum of the squared difference between the actual and predicted value. The RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{N} \cdot \sum_{n=1}^N (X_i - X_i')^2} \quad (1)$$

where  $X_i$  represents actual values,  $X_i'$  is forecasted points and  $N$  is the number of predictions for equations (1).

#### E. Results obtained in previous works

This solar irradiation forecasting technique was initially developed to predict irradiation values in the city of Vitoria-Gasteiz (Álava), Spain [17]. This technique was then applied in additional locations in Spain: Almike (Bizkaia) and Tudela (Navarra). Two architectures were repeatedly found to be the optimal: (1:4 delay and 20 neurons) and (1:2 delay and 20 neurons). The present work presents the result of an ensemble network that is based on the parallelization of both architectures in order to obtain better solar irradiation forecasts. The aim of this technique is to ensure that the error of the ensemble is lower than the error of each individual network.

## 4. Results analysis and discussion

The accuracy of predictions through the new forecaster, which was based on the combination of two architectures, was compared with the accuracy of the original forecasters, which were satisfactorily developed with the methodology explained in [17]. Two locations in Spain, Almike (Bizkaia) and Tudela (Navarra), were analysed in order to support the obtained conclusions.

In order to properly compare both forecasters, for each place 20 sunny days, 20 partially cloudy days and 20 cloudy days were selected and the errors averaged. As an example of the differences between the types of days, Figures 2, 3 and 4 give information for a sunny day, whereas Figures 5, 6 and 7 are for a cloudy day, obtained in all cases through the ensemble forecaster. In addition, the results obtained through the ensemble forecaster are also shown in order to compare the results of both forecasters.

Figure 2 shows the evolution of solar irradiation in Tudela on 13 July, 2017. From the smooth evolution of the irradiation, it can be concluded that it is a sunny day. While actual values are represented through a continuous line, forecasted values are represented by a discontinuous one.

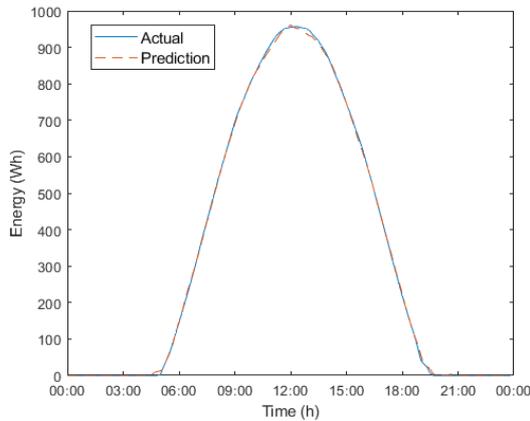


Fig. 2. Actual vs predicted irradiation values with ensemble forecaster on 13 July, 2017

For the whole day of 13 July, 2017, the RMSE obtained by the non-combined forecaster was  $5.00 \text{ W/m}^2$  while that of the ensemble forecaster was  $4.59 \text{ W/m}^2$ . Figure 3 shows the distribution of the errors made by the tool throughout the day. Although the vast majority of the errors are located next to zero values, the normal distribution bell curve is slightly off-centred to the negative side because there are more negative than positive errors.

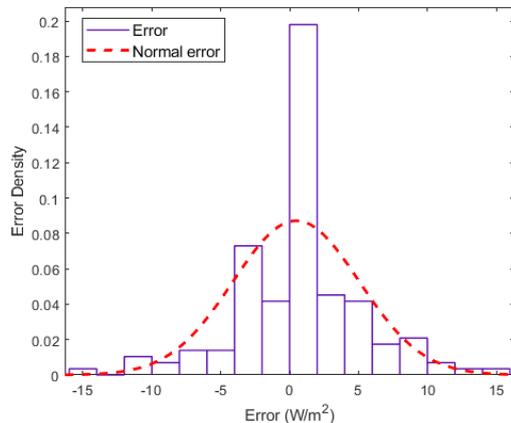


Fig. 3. Distribution of error on a sunny day with ensemble forecaster, 13 July, 2017

Moreover, Figure 4 represents the actual and predicted energy produced by a photovoltaic panel. The difference between real values and values predicted throughout the whole day is 0.43 % by the non-combined forecaster and 0.13 % by the ensemble forecaster.

The values for the power output have been calculated using (2).

$$P_s = \eta SI(1 - 0.005(t_0 - 25)) \quad (2)$$

where  $P_s$  is the generated electrical power (W),  $\eta$  is the conversion efficiency coefficient,  $S$  is the area of the module ( $\text{m}^2$ ),  $I$  is the solar irradiation ( $\text{W/m}^2$ ), and  $t_0$  is the

measured temperature ( $^{\circ}\text{C}$ ). To calculate the electrical power generated, a PV solar panel whose parameters are  $\eta = 17.59 \%$  and  $S = 1.6767 \text{ m}^2$  [17] was chosen.

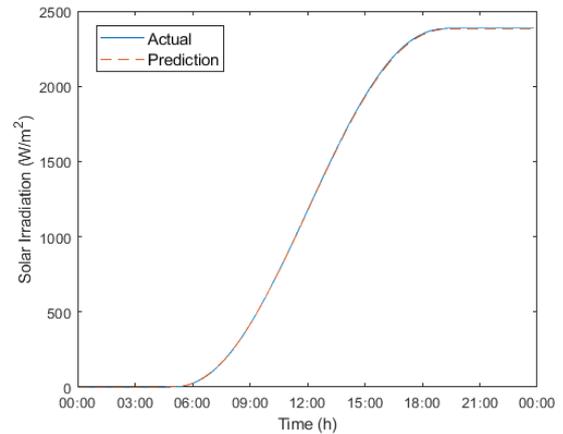


Fig. 4. Comparison of actual and predicted energy generation with ensemble forecaster on 13 July, 2017

Figure 5 represents the evolution of solar irradiation on a cloudy day in Tudela on 12 March, 2017. It is considered a cloudy day due to the sudden changes that the clouds produced in the solar irradiation, as is demonstrated in Figure 5.

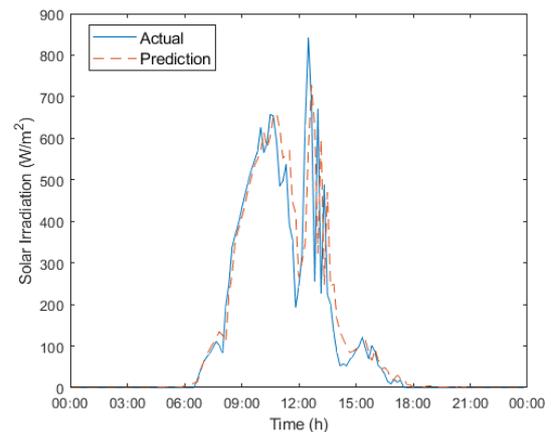


Fig. 5. Actual vs predicted irradiation values with ensemble forecaster on 12 March, 2017

The RMSE for the whole day, 12 March, 2017, was  $77.06 \text{ W/m}^2$  for the non-combined forecaster and  $77.15 \text{ W/m}^2$  for the ensemble forecaster. Figure 6 shows the distribution of the errors that the tool made throughout the day.

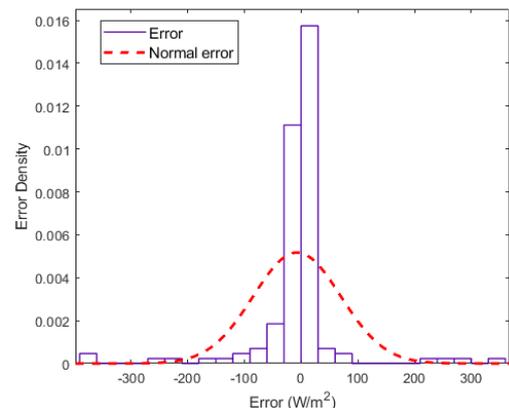


Fig. 6. Comparison between actual and predicted energy generation with ensemble forecaster on 12 March, 2017

In this second case, the RMSE is higher than in the previous one due to the sudden changes produced by the clouds. Therefore, it can be concluded that the more periods with clouds in a day, the higher the RMSE will be [17]. If Figure 6 is analysed closely, it can be seen that the vast majority of errors are next to zero and inside the distribution bell, similar to the sunny day. However, there are some values at the normal distribution tails which are responsible for the increase in the RMSE value.

Moreover, Figure 7 shows the difference between the actual and predicted energy produced by a photovoltaic panel. The difference in this case is 6.62% when using the non-combined forecaster and is 5.86% when using the ensemble forecaster. The increase in the values is related to the sudden changes in solar irradiation due to the cloud movement.

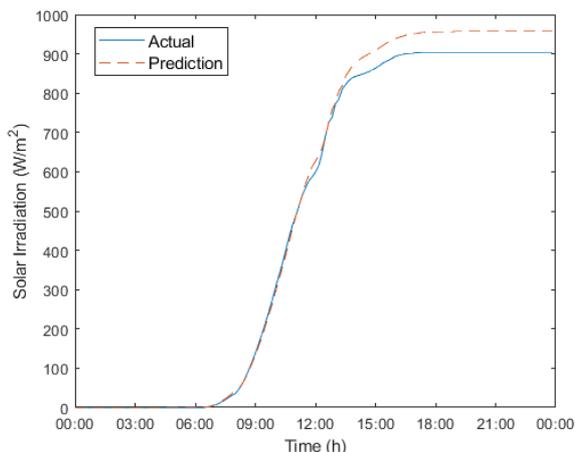


Fig. 7. Comparison between actual and predicted energy generation with ensemble forecaster on 12 March, 2017

Figures 8 and 9 show the difference between the original and the new forecaster used in Almike to predict the solar irradiation throughout 2017. Moreover, Tables II and III demonstrate that no matter whether the day was sunny, partially cloudy or cloudy, the new tool surpassed the original one. Concerning the error distribution throughout 2017, it remains similar in both cases, with slight changes.

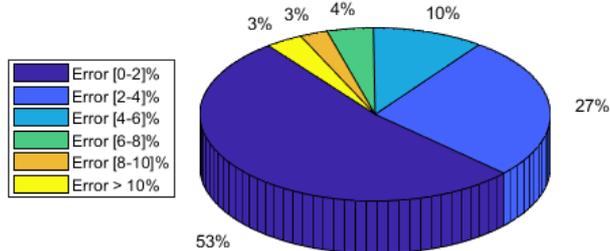


Fig. 8. Distribution of the error throughout 2017 for the Almike architecture (1:2 delay, 20 neurons)

Table II. –Almike architecture (1:2 delay, 20 neurons)

PARAMETER	SUNNY	PARTIALLY CLOUDY	CLOUDY
RMSE	8.45	35.70	85.06
Error (%)	1.30	1.71	2.75

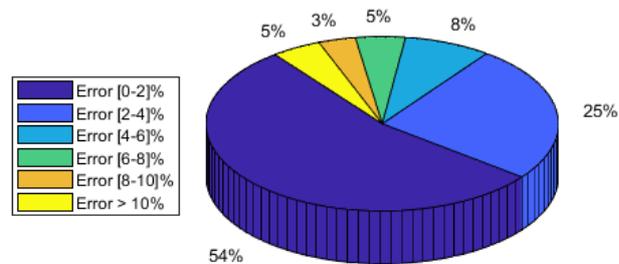


Fig. 9. Distribution of the error throughout 2017 for the new Almike solar irradiation forecaster

Table III. – Newly developed Almike irradiation forecaster

PARAMETER	SUNNY	PARTIALLY CLOUDY	CLOUDY
RMSE	7.52	34.68	81.76
Error (%)	0.95	1.71	2.36

Figures 10 and 11 show the difference between the original and new forecaster in Tudela in predicting the solar irradiation throughout 2017. Furthermore, Tables IV and V demonstrate once more that, for the RMSE, no matter the conditions—sunny, partially cloudy or cloudy—the new tool outperforms the original one. If the error (%) is analysed, it is clear that on sunny days the original forecaster makes better predictions than the new one. However, the difference between both values is not too large.

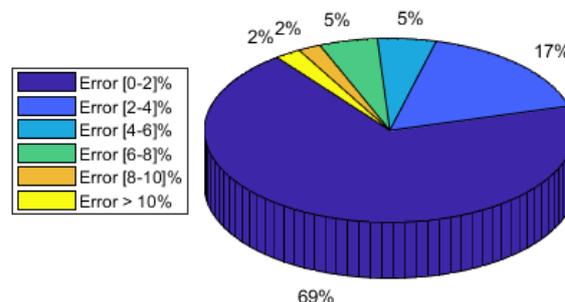


Fig. 10. Distribution of the error throughout 2017 in Tudela architecture (1:4 delay, 20 neurons)

Table IV. –Tudela architecture (1:4 delay, 20 neurons)

PARAMETER	SUNNY	PARTIALLY CLOUDY	CLOUDY
RMSE	8.24	31.99	87.02
Error (%)	0.49	1.37	2.92

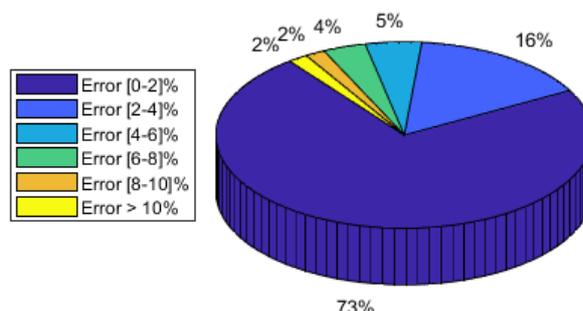


Fig. 11. Distribution of the error throughout 2017 in Tudela for the new solar irradiation forecaster

Table V. – Newly developed Tudela irradiation forecaster

PARAMETER	SUNNY	PARTIALLY CLOUDY	CLOUDY
RMSE	7.90	31.67	86.83
Error (%)	0.55	1.29	2.54

In addition, if the error distribution of both forecasters in Tudela throughout 2017 is compared, the results obtained with the new forecaster are better than with the original one. The accuracy for days with an error lower than 2% improves from 69% to 73% with the ensemble forecaster.

Finally, if the RMSE in Tables III and V, which are related to the ensemble forecaster, are compared with the results in the literature [4], [19], [23], it can be considered that the results obtained through this forecaster are similar to or better than other results obtained through different models. This improvement relies on the fact that the average of the forecasts made by different structures reduces the probability of less accurate predictions.

## 5. Conclusion

This paper presents a forecaster that is able to predict solar irradiation with a high standard of accuracy in a very short-term horizon, next 10 minutes. The results demonstrated that the predictions made with the ensemble forecaster (Figures 9 and 11, Tables II and V) equalled or surpassed the predictions of non-ensemble forecasters, even when they were found to be the optimal ones for each location, as seen in [17].

This improvement the accuracy relies on the fact that the ensemble forecaster is formed by two RNNs with different structures, and whose predictions are averaged to obtain a single forecast. The architecture of these RNNs are the most common structures obtained from applying the methodology summarized in this paper to two different Spanish locations. Finally, this tool can be integrated into microgrids with solar generators to reduce uncertainty.

## References

[1] H.T. Yang, C.H. Ming, Y.C. Huang, and Y.S. Pai, "A Weather-Based Hybrid Method for 1-Day ahead Hourly Forecasting of PV Power Output", in *IEEE Trans. Sustainable energy*, Vol5, pp. 917-926, July 2014.

[2] A.K. Sahoo, and S.K. Sahoo, "Energy Forecasting For Grid Connected MW Range Solar PV System", in *7th India International Conference on Power Electronics (IICPE)*, 2016.

[3] I. Majumder, M.K. Behera and N. Nayak, "Solar Power Forecasting Using a Hybrid EMD-ELM Method", in *International Conference on circuits Power and Computing Technologies (ICCPCT)*, 2017.

[4] A. Kaur, L. Nonnenmacher, H.T.P. Pedro and C.F.M. Coimbra, "Benefits of solar forecasting for energy imbalance markets", in *Renewable Energy*, Vol. 86, pp. 819-830, 2016.

[5] V. Sharma, D. Yang, W. Walsh and T. Reindl, "Short term solar irradiance forecasting using a mixed wavelet neural network", in *Renewable Energy*, Vol. 90, pp. 481-492, 2016.

[6] X. Qing and Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM", in *Energy*, Vol. 148, pp. 461-468, 2018.

[7] M. Paulescu, E. Paulescu, P. Gravila and V. Badescu, "Weather Modeling and Forecasting of PV Systems Operation," Springer, pp. 17-42, 2013.

[8] F.M. Lopes, H.G. Silva, R. Salgado, A. Cavaco, P. Canhoto and M. Collares-Pereira, "Short-term forecast of GHI and DNI for solar energy systems operation: assessment of the ECMWF integrated forecasting system in southern Portugal" in *Solar Energy*, Vol. 170, pp. 14-30, 2018.

[9] E.B. Ssekulima, M.B. Anwar, A.A. Hinai and M.S. El Morursi, "Wind speed and solar irradiance forecasting techniques for enhances renewable energy integration with the grid: a review", in *IET Renewable Power Generation*, Vol. 10, pp. 885- 898, 2016.

[10] E. Lorenz, J. Kühnert and D. Heinemann, "Overview of irradiance and photovoltaic power prediction", in A. Troccoli, L. Dubus, S.E. Haupt (Eds.): "Weather matters for energy", Springer, pp. 429-454, 2014.

[11] S. Pelland, J. Remund, J. Kleissl T. oozkeiand K. De Barbandere, "Photovoltaic and solar forecasting: State of the art", in *IEA PVPS, Task 14*, pp. 1-36, 2013.

[12] J. Boland, M. Korolkiewicz, M. Agrawal and J. Huang, "Forecasting solar irradiation on short time scales using a coupled autoregressive and dynamical system (cards) model", *Porc. Of the Australian Solar Energy Conf., Melbourne*, pp. 6-7, 2012.

[13] H. Jiang and Y. Dong, "A nonlinear support vector machine model with hard penalty function based on glowworm swarm optimization for forecasting daily global solar irradiation", in *Energy Conversion and Management*, Col. 126, pp. 991-1002, 2016.

[14] S.X. Chen, H.B. Gooi, and M.Q. Wang, "Solar irradiation forecast based on fuzzy logic and neural networks", in *Renewable Energy*, Vol. 52, pp. 118-127, 2013.

[15] P.S. Loh, J.V. Chua, A.C. Tan and C.I. Khaw, "Data-driven short-term forecasting of solar irradiance profile", in *World Engineers Submit – Applied energy Symposium & Forum: Low Carbon Cities & Urban Energy Joint Conference, WES-CUE 2017, 19-21 July 2017, Singapore*.

[16] S. Haykin, *Neural Networks and Learning Machines - 3rd ed.*, Upper Saddle River: Pearson Education, Inc., 2009.

[17] F. Rodríguez, A. Fleetwood, A. Galarza and F. Fontán, "Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control", in *Renewable Energy*, Vol. 126, pp. 855-864, 2018.

[18] S.A. Kalogirou, "Artificial neural networks in renewable energy systems applications: a review", in *Renewable & Sustainable Energy Reviews*, Vol. 5, pp. 373-401, 2001.

[19] A. Shakya, S. Michael, C. Saunders, D. Armstrong, P. Pandey, S. Chalise and R. Tonkoski, "Solar Irradiance Forecasting in Remote Microgrids Using Markove Switching Model", in *IEEE Transaction on Sustainable Energy*, Vol. 8, pp.895-905,2017

[20] H. Zhou, W. Xu, C. Xue, H.B. Cao, X. Gu and J. Wang, "A Short-Term Forecasting Model for Photovoltaic Plants Based on Data Mining", in *3<sup>rd</sup> IEEE International Conference on Computer and Communications*, 2017.

[21] J. E. Dayhoff and J. M. De Leo, "Artificial neural networks", in *Conference on Prognostic Factors and Staging in Cancer Management: Contributions of Artificial Neural Networks and Other Statistical Models*.

[22] B. Sivaneasan, C.Y. Yu and K.P. Goh, "Solar Forecasting using ANN with Fuzzy Logic Pre-processing", in *World Engineers Submit – Applied energy Symposium & Forum: Low Carbon Cities & Urban Energy Joint Conference, WES-CUE 2017, 19-21 July 2017, Singapore*.

[23] H. Sheng, J. Xiao, Y. Cheng, Q. Ni and S. Wang, "Short Term Solar Power Forecasting Based on Weighted Gaussian Process Regression", in *IEEE Transactions on Industrial Electronics*, Vol. 65, NO1, 2018.