Intra-hour Forecasting for a 50 MW Photovoltaic System in Uruguay: a Baseline Approach

Spyros Theocharides[1,*], Rodrigo Alonso-Suarez[2], Gianina Giacosa[2], George Makrides[1], Marios Theristis[1] and George E. Georghiou[1]

[1]PV Technology Laboratory, FOSS Research Centre for Sustainable Energy,
Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, 1678, Cyprus

[2]Laboratorio de Energía Solar, Universidad de la República (Udelar), Uruguay

***ABSTRACT* — The increased penetration of photovoltaic (PV) generation introduces new challenges for the stability of electricity grids. In this work, machine learning (ML) techniques were implemented to forecast PV power production up to 1-hour ahead with a 10-minute granularity. Three different input combinations were utilised: Model 1 (M1) using the AC power only, Model 2 (M2) using the elevation angle (α), azimuth angle (φ) and AC power and Model 3 (M3) using the AC power, α, φ and satellite observations (SAT) aiming to improve the forecasting performance. Historical PV operational data are used for the training and validation stages of intra-hour PV forecasting models for time t + 10 to 60 minutes ahead. The results obtained over the test set period (15% of the data, i.e. ≈ 110 days) have shown that M2 exhibits the best-performance with a normalised root mean square error (nRMSE) varying between 7.6% to 14.2%, whereas the skill score (SS) ranged between 6.5% and 30.9% for the 10- to 60-minute ahead respectively.**

I. Introduction

Accurate PV production forecasting can mitigate power quality effects posed by large shares of distributed PV systems through active grid management. Therefore, it is an important feature that can assist utilities and plant operators for energy management and dispatch. More specifically, short-term PV production forecasting (intra-hour) is necessary for power ramp and voltage flicker prediction as well as control operations and real-time energy dispatch. On the other hand, mid-term PV production forecasting (intra-day and day-ahead) is used for matching the demand unit commitment and production monitoring in order to control voltage and frequency levels and reduce the back-up reserves.

Parametric models for PV production forecasting have already been developed [1], [2], but their ability to forecast the power output of PV systems is not a straightforward process since information of the system characteristics and behaviour should be provided. Therefore, an important share of the research is devoted to the development of flexible prediction techniques using non-parametric models based on machine learning algorithms [3]–[6]. In order to train a PV power forecasting model, weather classification and machine learning techniques may be performed [7], [8]. Moreover, models that combine a physical model coupled to artificial neural networks (ANNs) have started emerging [5], [8]. Although a significant number of PV power forecasting tools have been developed, the challenge to provide a global and validated (against large scale data-sets) model for different PV plants remains unsolved. Additionally, to improve the accuracy of the PV power prediction, adaptive methods that can capture system information and behaviour without the need of datasheet and installation information must be employed. This is crucial because a large proportion of PV systems includes decentralized rooftop installations where knowledge of system information is not always available.

Furthermore, system behaviour can be estimated by processing recent PV operational data-sets using the classical approach of the feedforward neural network (FFNN). This algorithm is widely used in other fields for prediction, modelling and classification purposes. The classical approach of the FFNN with given inputs, a hidden layer and an output layer of linear and non-linear activation functions can be viewed as a convenient way to predict PV power output. FFNN can be trained to develop relational weighted chains between internal nodes in order to overcome the limitations of traditional methods in solving complex problems, which can be modelled through a supervised learning technique based on historical data. Because of this chain of relationships, theoretically, multi-layered neural networks can be universal approximators and have a tremendous potential to perform any nonlinear mapping based on historical time-series [9]. In addition, ANNs are efficient for online (i.e. real-time) training due to their capability of reflecting the information of new instances on a model by changing the weight values only.

In this work, a baseline approach is investigated for intra-hour (i.e. up to 1-hour ahead) PV power forecasting by utilising ANNs. Two-year data from a 50 MW PV power plant located in Uruguay (Salto) were used for training and testing the forecasting model. A forecasting performance assessment was also conducted in order to obtain a baseline performance level for PV power forecasting in north-western Uruguay. It is important to mention that the night hours are filtered out from the data.

II. Methodology

A. Neural Networks

The strength of ANN models relies on their ability to approximate non-linear functions through a supervised learning process. The training step is formulated in respect to minimising a loss function. The error term accounts for the discrepancy between the produced output of the network and the desired output. This is approximated using common statistical metrics such as the mean square error (MSE). In a multi-layer network, minimising the error in the training phase is achieved with the back-propagation (BP) algorithm, which is used to calculate the error contribution of each neuron after a batch of data is distributed back from the output through the network layers [10]. In addition, a regularisation term is used in the cost function to prevent overfitting, by controlling the effective complexity of the neural network. The regularisation of the designed networks in this study was performed by adding a penalty equal to the L2-norm of the weights, in order to reduce the value of the weights by the same factor [11].

B. Model selection and training

This model was selected through a series of validation steps performed by varying the topology (hyperparameters, combinations of input parameters and sizes) and architectural design (optimal hidden layers, neurons, iterations and learning function) of the ANN model. Here, the ANN was trained and validated using the historical PV production data-set.

Specifically, three models were tested with different input combinations, namely, Model 1 (M1) with the AC power as a sole input, Model 2 (M2) with the elevation angle (α), azimuth angle (φ) and AC power and Model 3 (M3) with the AC power, α, φ and satellite information (SAT) as inputs to the intra-hour PV power forecasting models. Different ANN models (weights) were trained to forecast the PV generation (output parameter) up to t + 10 to 60 minutes ahead while the input parameters were utilised at time t. Figure 1 demonstrates the training phase of the ANN model.

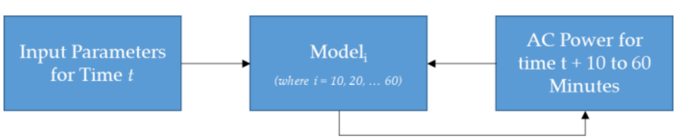

Fig. 1. Training procedure of the intra-hour PV power forecasting model.

The annual data-set was divided into three subsets: 70% for training, 15% for validation and 15% for testing. The training and validation sets were evaluated in a way to allow the implementation of the best-performing model. The weights of the neural network were obtained by varying the number of iterations in order to prevent overfitting. The validation set was used to evaluate the performance of the models.

C. Performance Metrics

The forecasting performance accuracy is assessed by utilising statistical metrics such as the the root mean square error (RMSE) that describes the standard deviation of the prediction errors, the normalised RMSE (nRMSE). Additionally, the mean bias error (MBE) was utilised to measure the bias between the forecasted and observed data and the skill score (SS) that measures the superiority over a reference model (for this study the persistence model (PM) was used as the reference model). The metrics used are demonstrated as follows:

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{i=1}^{n} (e_i)^2} \tag{2}$$

$$nRMSE = \frac{100}{\text{Pnominal}} \times \sqrt{\frac{1}{n} \times \sum_{i=1}^{n} (e_i)^2} \tag{3}$$

$$MBE = \frac{100}{n} \times \sum_{i}^{n} \frac{e_i}{y_{i.observed}} \tag{4}$$

$$e_i = y_{i,forecast} - y_{i,observed} \tag{5}$$

$$SS = 100 \cdot \left(1 - \frac{RMSE_{forecast}}{RMSE_{reference}}\right) \tag{6}$$

where $y_{\text{i,observed}}$ and $y_{\text{i,forecasted}}$, is the observed and forecasted AC power respectively, P_{nominal} is the nominal capacity of the PV system (i.e. 50 MW) and $RMSE_{\text{forecast}}$ and $RMSE_{\text{reference}}$ are the root mean square errors of the predicted and reference models (predictions of the PM).

IV. Results

The development of the PV generation forecasting was done using a five-neuron for M1 and M2 and six-neurons for M3 ANN topology, which is a good trade-off between accuracy and simplicity, as demonstrated in previous studies [11]. The network interface diagrams (NID) for the three models are demonstrated in Fig. 2.

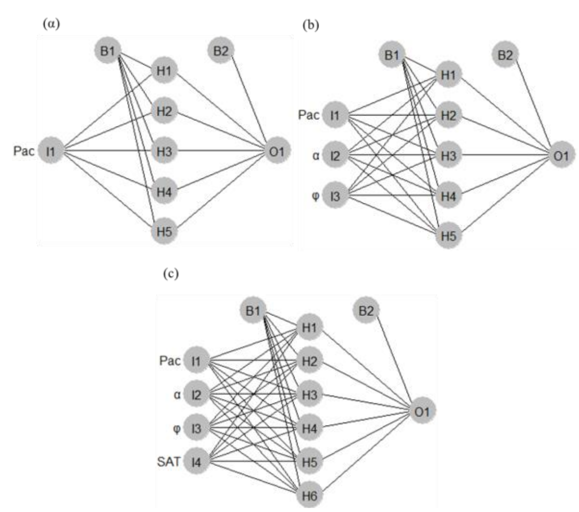

Fig. 2. Network interface diagrams (NID) of: (a) M1, (b) M2 and (c) M3.

Table I and II summarise the forecasting accuracy of the models at different intra-hour forecasting horizons over the period of the test set. The best-performing input combination was exhibited by M2 with an nRMSE varying between 7.6% to 14.2% and an SS of 6.5% to 30.9%. Additionally, the SS results of M2 demonstrated interesting improvements from 10 to 60 minutes when compared with the persistence model (PM).

TABLE I
STATISTICAL ANALYSIS OF THE RESULTS (NRMSE).

Forecasts (Minutes)	nRMSE (%)		
	M1	M2	M3
10	8.1	7.6	7.5
20	11.4	10.2	10.9
30	13.7	11.6	13.1
40	15.8	12.6	13.6
50	17.7	13.4	13.9
60	19.6	14.2	15.4

TABLE II
STATISTICAL ANALYSIS OF THE RESULTS (SKILL SCORE).

Forecasts (Minutes)	SS (%)		
	M1	M2	M3
10	1.0	6.5	7.6
20	1.6	12.1	5.2
30	2.2	17.4	6.1
40	3.0	22.5	16.2
50	3.9	27.3	24.3
60	5.0	30.9	25.1

Furthermore, Table III summarises the results of the MBE, demonstrating an MBE ranging from -0.1 to -0.4 for M2, which was the model with the lowest average MBE. However, all three models exhibited low MBE for 10 to 60 minutes forecasting, indicating no biases between the forecasted and the observed data.

TABLE III
STATISTICAL ANALYSIS OF THE RESULTS (MBE).

Forecasts (Minutes)	MBE (%)		
	M1	M2	M3
10	-0.1	-0.1	-0.4
20	-0.2	-0.1	-0.9
30	-0.3	-0.2	-1.2
40	-0.4	-0.3	-1.0
50	-0.5	-0.4	-1.2
60	-0.6	-0.4	-1.1

In addition, based on the results of Table I, II and III, the best-performing model was M2. Fig. 3 demonstrates a histogram of the error distribution of 10, 30 and 60 minutes forecasts. As could be observed, the major amount of error is at the range of zero. However, moving from 10 to 60 minutes forecasts the error distribution is scattered through largest numbers.

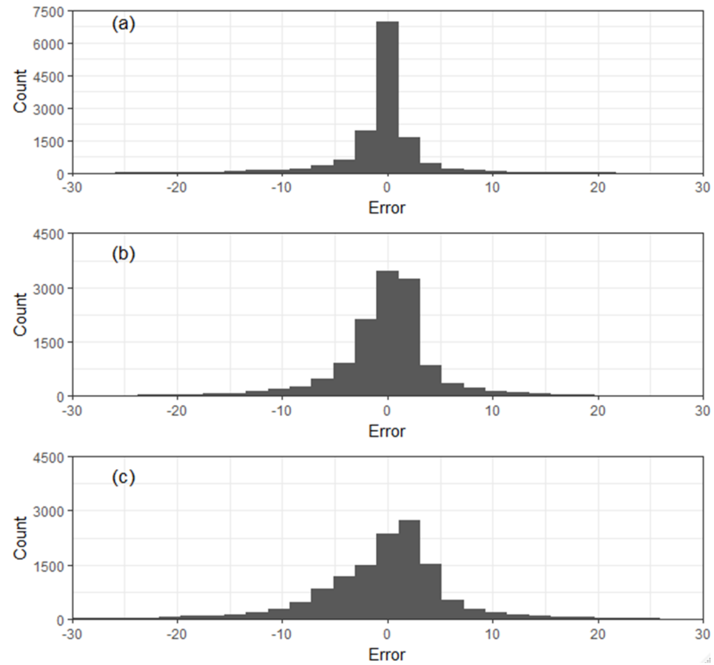

Fig. 3. Error histogram of: (a) 10 minutes, (b) 30 minutes and (c) 60 minutes forecasts. The histogram demonstrated a scattered variation moving from 10 to 60 minute forecasting.

Additionally, Fig. 4 demonstrates the actual against the forecasted AC power for the 10 to 60 minutes forecasts for the M2 model (as the best-performing model) (Fig. 4a to Fig 4c respectively). All three forecasts captured the systems behaviour. For days with high irradiance, the M2 model could accurately forecast AC power. However, when there are fluctuations, the model still requires some improvements.

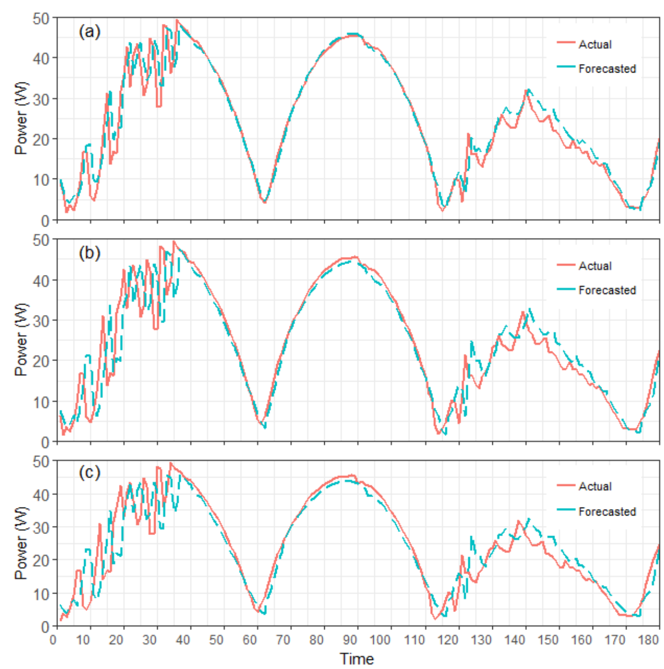

Fig. 4. Observed against forecasted AC power for: (a) 10 minutes, (b) 30 minutes and (c) 60 minutes forecasts. The night hours are filtered out.

V. Conclusions

In this study, an ANN was developed in order to implement a non-parametric intra-hour PV power generation forecasting model. Two input combinations were tested and trained with a random sample 70:15:15% of train, validation and test set approach. The models were adjusted and validated against real AC power measurements at a 50 MW PV power plant in Salto, Uruguay.

The results showed that the best-performing input combinations were exhibited by M2 (taking into account the a, φ and AC power) with an nRMSE varying between 7.6% to 14.2% and an SS of 6.5% to 30.9% compared to M1 and M3. Additionally, the SS results of M2 demonstrated improvements from 10 to 60 minutes when compared with the persistence model (PM). In addition, the MBE results for all three models demonstrated that no biases exists between the forecasted and the observed data.

The plot of error histogram of M2 (best-performing model) for 10, 30 and 60 minute forecasts demonstrated that the major amount of error is in the range of 0. However moving from 10 to 60 minutes ahead forecasts the error distribution is scattered to largest numbers.

Finally, when the AC power plot of the observed against the forecasted data is demonstrated for M2 (for 10, 30 and 60 minutes forecasting), the model was accurately recorded the system's behaviour for clear sky days. However, when fluctuations appeared to the irradiance, the model still requires some improvements.

Acknowledgement

This work was co-financed by the European Regional Development Fund and the Republic of Cyprus in the framework of the project "INFORPV" with grant agreement number SOLAR-ERA.NET/1215/02. Suarez and Giacosa would like to thank the ANII, Uruguay for the financial support (grant number: FSE-ANII 131799). All authors would also like to acknowledge the DIEGO project, which is funded through the ERASMUS+ programme for education (project number: 561795-EPP-1-2015-1-IT-EPPKA2-CBHE-JP).

References

[1] E. Lorenz, T. Scheidsteger, J. Hurka, D. Heinemann, and C. Kurz, "Regional PV power prediction for improved grid integration," *Prog. Photovoltaics Res. Appl.*, vol. 19, no. 7, pp. 757–771, 2011.

[2] S. Pelland, G. Galanis, and G. Kallos, "Solar and photovoltaic forecasting through post-processing of the Global Environmental Multiscale numerical weather prediction model," *Prog. Photovoltaics Res. Appl.*, vol. 21, no. 3, pp. 284–296, 2013.

[3] I. Ashraf and A. Chandra, "Artificial neural network based models for forecasting electricity generation of grid connected solar PV power plant," *Int. J. Glob. Energy Issues*, vol. 21, no. 1, pp. 119–130, 2004.

[4] D. Caputo, F. Grimaccia, M. Mussetta, and R. E. Zich, "Photovoltaic plants predictive model by means of ANN trained by a hybrid evolutionary algorithm," in *Proceedings of the International Joint Conference on Neural Networks*, 2010.

[5] A. Gandelli, F. Grimaccia, S. Leva, M. Mussetta, and E. Ogliari, "Hybrid model analysis and validation for PV energy production forecasting," in *Proceedings of the International Joint Conference on Neural Networks*, 2014, pp. 1957–1962.

[6] F. Almonacid, P. J. Pérez-Higueras, E. F. Fernández, and L. Hontoria, "A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator," *Energy Convers. Manag.*, vol. 85, pp. 389–398, 2014.

[7] C. Chen, S. Duan, T. Cai, and B. Liu, "Online 24-h solar power forecasting based on weather type classification using artificial neural network," *Sol. Energy*, vol. 85, no. 11, pp. 2856–2870, 2011.

[8] A. Dolara, F. Grimaccia, S. Leva, M. Mussetta, and E. Ogliari, "A physical hybrid artificial neural network for short term forecasting of PV plant power output," *Energies*, vol. 8, no. 2, pp. 1138–1153, 2015.

[9] M. P. Almeida, M. Muñoz, I. de la Parra, O. P. Lamigueiro, and L. N. Fernández, "Comparative study of nonparametric and parametric PV models to forecast AC power output of PV plants," in *EU PVSEC Proceedings*, 2015, pp. 2230–2234.

[10] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[11] S. Theocharides, G. Makrides, V. Venizelou, P. Kaimakis, and G. E. Georghiou, "PV Production Forecasting Model Based on Artificial Neural Networks (ANN)," *33rd Eur. Photovolt. Sol. Energy Conf.*, pp. 1830–1894, 2017.