SAND2020-9426C

HYBRID MODELLING OF PV POWER GENERATION FOR ENHANCED FORECASTING

Spyros Theocharides[1, *], George Makrides[1], Marios Theristis[2], Marios Kynigos[1], 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]Sandia National Laboratories Albuquerque, NM 87185, USA
*corresponding author email: theocharidis.spyros@ucy.ac.cy

ABSTRACT: Accurate photovoltaic (PV) production forecasting is an important feature that can assist utilities and plant operators in the direction of energy management and dispatchability planning. Although numerous forecasting models have been reported in literature, the challenge of improved accuracy remains unsolved. In this work, a day-ahead PV power model utilising a hybrid approach is derived to feed into an Artificial Neural Network (ANN) and a linear regression model trained for PV power forecasting. The study focuses on improving the forecasting accuracy by employing machine learning and linear regression models that could record the behaviour of the PV system. The performance of the hybrid model was assessed against a single ANN model using a historical test set. The results showed that the hybrid model outperformed the single ANN model exhibiting a normalised mean square error (nRMSE) of 7.05%.
Keywords: artificial neural networks, forecasting, performance, photovoltaic, power.

1 INTRODUCTION

PV production forecasting can mitigate the power quality effects posed by large shares of distributed systems through active grid management and is, therefore, an important feature that can assist utilities and plant operators in the direction of energy management and dispatchability planning. More specifically, short-term PV production forecasts (intra-hour) are necessary for power ramp and voltage flicker prediction as well as control operations and dispatch management. On the other hand, mid-term PV production forecasting (intra-day and day-ahead) is used for load consumption and production monitoring to control voltage and frequency levels and reduce secondary reserve.

During the last decades, electricity system operation has been upgraded involving PV production forecasts and most commonly adopted PV production forecasting approaches are based on time series analysis techniques [1]. In addition, parametric models for PV production forecasting have already been developed [2], [3] but their ability to forecast the power output of PV systems is not a straightforward process since accurate knowledge of system characteristics and behaviour is required. Therefore, a huge share of research is devoted to the development of more sophisticated and flexible prediction techniques using non-parametric models based on machine learning algorithms [4]–[9]. In order to train a PV power forecasting model with identical data, weather classification and machine learning techniques may be performed [10], [11]. Moreover, joint models comprised of a combination of features of a physical model and artificial neural networks (ANNs) were presented elsewhere [7], [12]. Although a significant number of PV power forecasting tools have been developed, the challenge to provide a location-independent and validated (against large scale data-sets) model for different PV module types remains unsolved. Additionally, to improve the accuracy of 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 de-centralized rooftop installations where knowledge of system information is not always available.

Furthermore, the system's behaviour can be estimated by processing recent PV operational data-sets using the classical approach of the feedforward neural network (FFNN). The classical approach of the FFNN with an input, a hidden and an output layer of linear and non-linear activation functions can be viewed as a convenient way to predict the power output of PV systems. FFNN can be trained to develop relational weighted chains between internal nodes to overcome the limitations of traditional methods to solve 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 tremendous potential to perform any nonlinear mapping through a learning process based on historical time-series [13]. In addition, ANNs are efficient for online training due to their capability of reflecting the information of new instances on a model by changing the weight values only.

Another approach is to utilise Numerical Weather Prediction (NWP) models to forecast weather variables [3], [14]–[16]. NWP models deliberate the atmosphere as a fluid thus the concept of weather forecasts is to parametrise the state of the fluid at a certain time (t) by utilising the fluid- and thermo-dynamic equations to forecast the state of the fluid at time (t + n) [17]–[19].

In this work, a day-ahead PV power model utilising a hybrid approach is derived to feed into an ANN and a linear regression model trained for PV power forecasting. The study focuses on improving the forecasting accuracy by employing machine learning and linear regression models that could record the behaviour of the PV system.

2 EXPERIMENTAL SETUP

2.1. Outdoor test facility description

The outdoor test facility (OTF) of the University of Cyprus includes a fixed plane infrastructure for outdoor performance assessments at both the module and system level. The installed poly-c-Si system was mounted in a

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

portrait arrangement on aluminium mountings, at the optimum annual energy yield plane-of-array (POA) angle for Cyprus of 27.5˚.

The PV system was connected to a data-acquisition platform, used for the monitoring and storage of meteorological and PV operational data. The performance of the system and the prevailing meteorological conditions were recorded according to the requirements set by the IEC 61724 [20]. Specifically, the irradiance and meteorological measurements include the global horizontal irradiance (*GHI*), in-plane global irradiance (G_I), relative humidity (*RH*), wind direction (W_a), wind speed (W_s) and ambient temperature (T_{amb}). The PV system operational measurements include the maximum power current (I_{mp}), voltage (V_{mp}) and power (P_{mp}), as measured at the output of the PV array (DC side) [21], [22]. Additionally, the elevation angle of the sun (α) and the azimuth angle of the sun (φ_s) were calculated [23] for the location of the OTF.

2.2 PV system description

The PV system comprises of five poly-c-Si PV modules. The modules of the system are connected in series to form a PV string at the input of a string inverter. The main technical specifications of the test PV system are summarised in Table I.

Table I: Installed PV system technical characteristics.

Technical characteristic	**Parameter**
Modules	5 × poly-c-Si
System power (datasheet)	1365 W_p
Installation date	01/06/2015
Efficiency	14.40%

3 METHODOLOGY

The hybrid model comprised of two separate models. The ANN model was deployed to forecast the PV power during irradiance levels ≥ 300 W/m^2, whereas the linear regression model was used to forecast the PV power during irradiance levels lower than 300 W/m^2. This approach was selected after a series of stress tests on both models, where their corresponding behaviours on high and low irradiance levels was recorded and analysed. Both models were fed with historical meteorological and PV operational data see Figure 1). Furthermore, the models were optimised according to their input parameters, training/testing dataset range and their architectural (hyper) parameters.

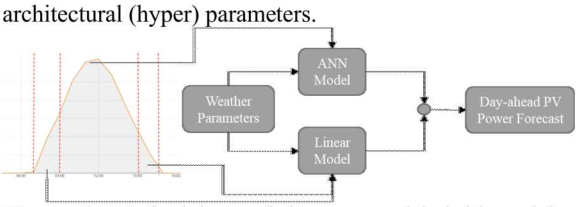

Figure 1: Methodology of the proposed hybrid model. Each model was trained with data within the predefined boundaries, which are described by the red dotted lines.

The best-performing model was selected through a series of validation tests including the optimisation of input combinations and sizes and hyper parameters of the ANN model. More specifically, the annual dataset was divided into three subsets including a training, validation and testing set. Training and validation sets were used to identify the best-performing model, by varying the input parameters and training period from 30% to 70% of the actual dataset. Additionally, the hyper parameters of the ANN model were tested by employing the validation set while the test set was used as the final assessment of the best-performing model (when the NWP data were employed).

3.1 Artificial Neural Networks

A Bayesian Regularization Neural Network (BRNN) is essentially a simple Multi-Layer Perceptron (MLP) in which a Bayesian regularization has been applied to its training function. The proposed model is given by Eq. (1) [24], [25]:

$$y_i = g(x_i) + e_i = \sum_{k=1}^{9} w_k g_k \Big(b_k + \sum_{j=1}^{p} x_{ij}\beta_j^{[k]} + e_i\Big), i = 1, ..., n \quad (1)$$

where $e_i \sim N(0, \sigma^2 e)$, *s* is the number of neurons, w_k is the weight of the *k*-th neuron, b_k is a bias for the *k*-th neuron, $\beta_j^{[k]}$ is the weight of the *j*-th input to the net, *and* $gk(\cdot)$ is the activation function:

$$g_k(x) = \frac{exp\left(2x\right) - 1}{exp\left(2x\right) + 1} \quad (2)$$

The model will minimise according to Eq. (3) [24], [25]

$$F = \beta E_D + aE_W \quad (3)$$

where E_D is the error of sum squares, E_W is the sum of squares of network parameters (weights and biases). E_W is the sum of squares of network parameters (weights and biases), β and α are the dispersion parameters for weights and biases.

The regularisation term applied was the squared sum of the weights of the neural network [26]:

$$\varepsilon_T = \beta\varepsilon_D + \alpha\varepsilon_R = \frac{\beta}{2}\sum_{k=1}^{N}(y_k - t_k) + \frac{\alpha}{2}\sum_{i=2}^{M} w_i \quad (4)$$

where α and β are coefficients assigned to each term. The second term in Eq. (4) is called weight decay and it ensures that the weights of the network do not exceed the total error of the network.

Once the data were fed into the network, the density function for the weights can be updated according to the Bayes' rule [25]:

$$P(w|\, D, \alpha, \beta, M) = \frac{P(D|\, w; \beta, M)P(w|\, a, M)}{P(D|\, a, \beta, M)} \quad (5)$$

where *D* represents the data-set, *M* the model used for the Neural Network and *w* is the vector of neural network's weights. *P(w|α, M)* represents the values of weights prior to the data-set input. *P(D|w,β, M)* is the probability of the data occurring based on the weights. *P(D|a,β, M)* is a normalisation factor, which ensures that the total summation of the probability is one.

In addition, the regularisation term is used 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.

3.2. Linear regression model

A linear regression model is a linear method to model the association among a scalar response of one or more independent variables [27]:

$$y=\beta_0+\beta_1\cdot X_1+\beta_2\cdot X_2+\ldots \quad (6)$$

where β_0 is the intercept, β_1, β_2... are the coefficients for each parameter and X_1, X_2 … are the input parameters.

3.3 Numerical Weather Predictions

The utilised NWP data were derived from the Weather Research and Forecasting (WRF) Model, which is a mesoscale NWP model designed for atmospheric research and operational forecasting applications [28].

3.4 Model performance assessment

The forecasting performance accuracy was assessed based on several predefined metrics when the test set was applied to the developed algorithms. The metrics commonly used in PV production forecasting applications include the mean absolute percentage error (MAPE), root mean square error (RMSE) and normalised RMSE (nRMSE) to the nominal PV system peak power:

$$MAPE=\frac{100}{n}\times\sum_{i=1}^{n}\left|\frac{y_{\mathrm{actual},i}-y_{\mathrm{predicted},i}}{y_{\mathrm{actual},i}}\right| \quad (7)$$

$$RMSE=\sqrt{\frac{1}{n}\times\sum_{i=1}^{n}(y_{\mathrm{actual},i}-y_{\mathrm{predicted},i})^2} \quad (8)$$

$$nRMSE=\frac{100}{P_{\mathrm{nominal}}}\times\sqrt{\frac{1}{n}\times\sum_{i=1}^{n}(y_{\mathrm{actual},i}-y_{\mathrm{predicted},i})^2} \quad (9)$$

where $y_{\mathrm{actual},i}$ and $y_{\mathrm{forecasted},}$ is the actual and predicted irradiance and power respectively, Y_{nominal} is the nominal peak power of the PV system (1365 W).

4 RESULTS

The hybrid model comprised of 4 input parameters [29] (forecasted *GHI,* forecasted T_{amb}, α and φ_s).

Table II demonstrates the error performance results of the single ANN model against the hybrid model over 110 days. The hybrid model exhibits improved accuracy compared to the performance of the single ANN model. Specifically, the hybrid model demonstrates an nRMSE and MAPE (7.05%, 5.86%) with a ~2% decrease compared to the single ANN model (9.35%, 8.72%).

Table II: Performance metrics of the single ANN model against the hybrid model.

Forecasts	Metrics		
	MAPE (%)	**RMSE (W)**	**nRMSE (%)**
ANN	8.72%	127.62	9.35%
Hybrid Model	5.86%	96.23	7.05%

In addition, Figure 1 demonstrates the daily nRMSE results of the hybrid model. As can be observed, the majority of days are below the 6% error, while all the tested days are exhibiting nRMSE below the 9%.

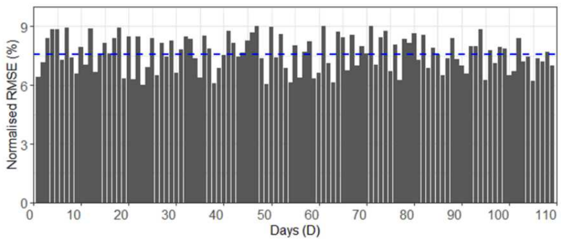

Figure 2: Daily nRMSE results of the hybrid model. The blue dashed line demonstrates the average nRMSE which is 7.05%.

In addition, Figure 3 demonstrates the histogram of the error distribution of the hybrid model (Fig. 3a) and single ANN model (Fig. 3b). It can be seen that the hybrid model exhibits the minimum error distribution compared to the single ANN model.

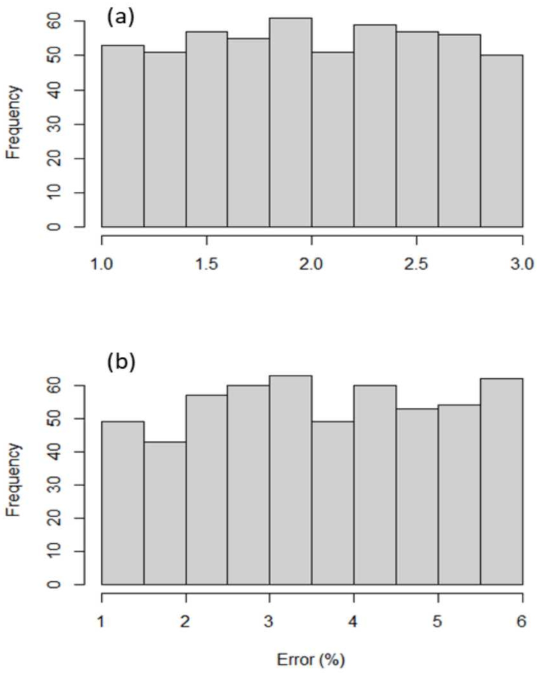

Figure 1 Histogram of the error distribution of : (a) Hybrid model and (b) single ANN model.

Finally, Figure 4a demonstrates the diurnal comparison of a typical day exhibiting higher errors during the low irradiance hours whereas Figure 4b demonstrated the improvements by applying the hybrid model.

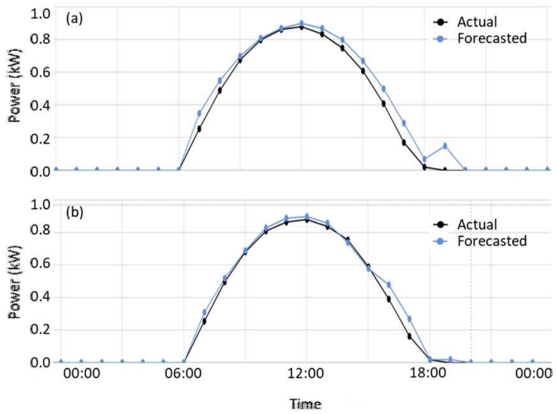

Figure **2:** Typical day of the observed against the forecasted power of: (a) single ANN model and (b) Hybrid model. The hybrid model corrected the high errors during the low irradiance hours.

5 CONCLUSIONS

A robust hybrid model was developed comprised of an ANN model and a linear regression. The ANN model was deployed to forecast the PV power during irradiance levels ≥ 300 W/m^2, whereas the linear regression model was used to forecast the PV power during irradiance levels lower than 300 W/m^2.

The results showed that the hybrid model exhibits the best performance compared to the single ANN model. Specifically, the hybrid model demonstrated an nRMSE and MAPE with a ~2% absolute decrease compared to the single ANN model, while all the test days of the hybrid models were exhibiting errors below 9%. Furthermore, the error distribution of the hybrid model has a significant reduction during the low irradiance hours compared to the single ANN model.

6 ACKNOWLEDGEMENT

It has received funding from the European Union's Horizon 2020 research and innovation programme under the SOLAR ERANET COFUND network (Project Number: P2P/SOLAR/1216/0003).Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

7 REFERENCES

[1] S. Pelland, J. Remund, J. Kleissl, T. Oozeki, and K. De Brabandere, "Photovoltaic and Solar Forecasting: State of the Art," *Int. Energy Agency Photovolt. Power Syst. Program. Rep. IEA PVPS T14*, pp. 1–40, 2013, [Online]. Available: https://www.researchgate.net/publication/259717376_Photovoltaic_and_Solar_Forecasting_State_of_the_Art.

[2] 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, doi: 10.1002/pip.1033.

[3] 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, doi: 10.1002/pip.1180.

[4] 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, [Online]. Available: http://inderscience.metapress.com/content/N0R76HDTKC19GWVH.

[5] D. Caputo, F. Grimaccia, M. Mussetta, and R. E. Zich, "Photovoltaic plants predictive model by means of ANN trained by a hybrid evolutionary algorithm," 2010, doi: 10.1109/IJCNN.2010.5596782.

[6] S. A. Kalogirou, "Applications of artificial neural-networks for energy systems," *Appl. Energy*, vol. 67, no. 1–2, pp. 17–35, 2000, doi: 10.1016/S0306-2619(00)00005-2.

[7] 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, doi: 10.1109/IJCNN.2014.6889786.

[8] A. Mellit and A. M. Pavan, "A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy," *Sol. Energy*, vol. 84, no. 5, pp. 807–821, 2010, doi: 10.1016/j.solener.2010.02.006.

[9] 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, doi: 10.1016/j.enconman.2014.05.090.

[10] 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, doi: 10.1016/j.solener.2011.08.027.

[11] J. Shi, W. J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic system based on weather classification and support vector machine," 2011, doi: 10.1109/IAS.2011.6074294.

[12] 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, doi: 10.3390/en8021138.

[13] 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, [Online]. Available: http://oa.upm.es/42462/.

[14] S. E. Haupt *et al.*, "Building the Sun4Cast System: Improvements in Solar Power Forecasting," *Bull. Am. Meteorol. Soc.*, vol. 99, no. 1, pp. 121–136, 2018, doi: 10.1175/BAMS-D-16-0221.1.

[15] D. P. Larson, L. Nonnenmacher, and C. F. M. Coimbra, "Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest," *Renew. Energy*, vol. 91, pp. 11–20, 2016, doi: 10.1016/j.renene.2016.01.039.

[16] X. Sun and T. Zhang, "Solar Power Prediction in Smart Grid Based on NWP Data and an Improved Boosting Method," in *Proceedings - 1st IEEE International Conference on Energy Internet, ICEI 2017*, 2017, pp. 89–94, doi: 10.1109/ICEI.2017.23.

[17] P. Lynch, "The origins of computer weather prediction and climate modeling," *J. Comput. Phys.*, vol. 227, no. 7, pp. 3431–3444, 2008, doi: 10.1016/j.jcp.2007.02.034.

[18] L. F. Richardson, *Weather prediction by numerical process, second edition*. 2007.

[19] R. Kimura, "Numerical weather prediction," *J. Wind Eng. Ind. Aerodyn.*, vol. 90, no. 12–15, pp. 1403–1414, 2002, doi: 10.1016/S0167-6105(02)00261-1.

[20] International Electrotechnical Commission, "IEC 61724:1998. Photovoltaic system performance monitoring — Guidelines for measurement , data exchange and analysis," p. 20, 1998, doi: 10.1016/j.solener.2008.04.009.

[21] G. Makrides, B. Zinsser, M. Norton, G. E. Georghiou, M. Schubert, and J. H. Werner, "Potential of photovoltaic systems in countries with high solar irradiation," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 2. pp. 754–762, 2010, doi: 10.1016/j.rser.2009.07.021.

[22] G. Makrides, B. Zinsser, M. Schubert, and G. E. Georghiou, "Energy yield prediction errors and uncertainties of different photovoltaic models," *Prog. Photovoltaics Res. Appl.*, vol. 21, no. 4, pp. 500–516, 2013, doi: 10.1002/pip.1218.

[23] I. Reda and A. Andreas, "Solar position algorithm for solar radiation applications," *Sol. Energy*, vol. 76, no. 5, pp. 577–589, 2004, doi: 10.1016/j.solener.2003.12.003.

[24] 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, doi: 10.1038/323533a0.

[25] D. J. C. MacKay, "Bayesian Interpolation," *Neural Comput.*, vol. 4, no. 3, pp. 415–447, 1992, doi: 10.1162/neco.1992.4.3.415.

[26] F. Dan Foresee and M. T. Hagan, "Gauss-Newton approximation to bayesian learning," in *IEEE International Conference on Neural Networks - Conference Proceedings*, 1997, vol. 3, pp. 1930–1935, doi: 10.1109/ICNN.1997.614194.

[27] A. C. Rencher and G. B. Schaalje, *Linear Models in Statistics*. 2007.

[28] N. C. for A. R. NCAR, "ARW Modelling System UserGuide V.3," *Book*, no. January, p. 408, 2016, doi: 10.5065/D68S4MVH.

[29] S. Theocharides, G. Makrides, A. Livera, M. Theristis, P. Kaimakis, and G. E. Georghiou, "Day-ahead photovoltaic power production forecasting methodology based on machine learning and statistical post-processing," *Appl. Energy*, vol. 268, no. December 2019, p. 115023, 2020, doi: 10.1016/j.apenergy.2020.115023.