Application of Time Series and Artificial Neural Network Models in Short-term Forecasting of PV Power Generation

E. G. Kardakos*, M. C. Alexiadis, S. I. Vagropoulos, C. K. Simoglou, P. N. Biskas, and A. G. Bakirtzis

Power Systems Laboratory, Department of Electrical and Computer Engineering
Aristotle University of Thessaloniki
54124, Thessaloniki, GREECE

*Corresponding author: ekardako@hotmail.com

Abstract- **This paper addresses two practical methods for electricity generation forecasting of grid-connected PV plants. The first model is based on seasonal ARIMA time-series analysis and is further improved by incorporating short-term solar radiation forecasts derived from NWP models. The second model adopts artificial neural networks with multiple inputs. Day-ahead and rolling intra-day forecast updates are implemented to evaluate the forecasting errors. All models are compared in terms of the Normalized (with respect to the PV installed capacity) Root Mean Square Error (NRMSE). Simulation results from the application of the forecasting models in different PV plants of the Greek power system are presented.**

Index Terms-- **Artificial neural networks, autoregressive integrated moving average (ARIMA) models, day-ahead forecasting, photovoltaic plants, solar radiation.**

I. Introduction

Political directions for a decarbonized future [1] have already led to various support schemes in many countries towards the large-scale RES penetration. Apart from the obvious environmental benefits, the large share of intermittent RES electricity production presents a big challenge in the efficient management of the power systems, since it can seriously affect the short-term scheduling of the conventional generating units, the maintenance scheduling of generating plants and transmission lines, the required levels of system reserves, etc.

Besides wind parks, photovoltaic (PV) units are continuously gaining acceptance worldwide, since they can be easily installed and operate in various sites. However, although PV generation follows a relative expected daily cycle, it still remains an intermittent and non-dispatchable power source, due to the inherent difficulty in accurately predicting meteorological conditions, such as cloud cover. Therefore, the research community has focused lately on the development of advanced PV forecasting models and tools, which are generally divided into two main groups. The first group comprises the pure physical models, which transform the output of numerical weather prediction (NWP) models into PV power output by performing appropriate post-process calculations and often implementing Model Output Statistics (MOS) to correct the forecasts. Such models are described in [2], where a forecasting approach based on numerical weather forecasts is developed and in [3], where the forecasting approach is based on inputs from sky-cover predictions. Global horizontal irradiance forecasts from three operational NWP models are analyzed and evaluated in [4]. The second group comprises the statistical models (i.e. time series, artificial neural networks, etc.), which focus on the emulation of the relation between historical data in order to predict the PV power output, usually taking into account correlations with exogenous meteorological data (e.g. solar radiation, temperature, etc.). Statistical models can be found in [5]-[7]. A comparison of six different statistical models for very short-term forecasting (0-4 hours ahead) is presented in [8], whereas indicative hybrid forecasting models, which combine different statistical approaches to improve the forecasting efficiency, are described in [9]-[10]. An extensive review of state-of-the art methods for solar irradiation forecasting can be found in [11].

In this paper, two practical methods for electricity generation forecasting of grid-connected PV plants are presented. The first model is based on seasonal ARIMA time-series analysis and is further improved by incorporating short-term solar radiation forecasts derived from NWP models. The second model adopts artificial neural networks with multiple inputs. Day-ahead and rolling intra-day forecast updates are implemented to evaluate the forecasting errors. Both models are compared in terms of the Normalized (with respect to the PV installed capacity) Root Mean Square Error (NRMSE). Simulation results from the application of all models in different PV plants of the Greek power system are presented.

II. PV Forecasting Models

In this section, the main features of both forecasting models are described in detail.

This work was supported by the General Secretariat of Research and Technology (GSRT), Hellenic Ministry of Education and Religious Affairs, Culture and Sports, in the context of the Action "ARISTEIA" (Project Code: 1522).

978-1-4799-3254-2/13/$31.00 ©IEEE

A. Time Series Models

A class of time series techniques, namely ARIMA, can be employed for the short-term forecasting of PV power generation. ARIMA is a method first introduced by Box and Jenkins [12] and has now become one of the most popular methods for time series forecasting.

In this paper, a variation of the classical ARIMA model, namely the seasonal ARIMA model (i.e. SARIMA) is used, in order to account for the inherent seasonal effect of the PV power output. The seasonal ARIMA model is generally referred to as $SARIMA\left(p,d,q\right)\, x\, \left(P,D,Q\right)_s$, where *p, d, q* and *P, D, Q* are non-negative integers that refer to the polynomial order of the autoregressive (AR), integrated (I), and moving average (MA) parts of the non-seasonal and seasonal components of the model, respectively.

The SARIMA model is described mathematically as follows:

$$\varphi_p(B)\Phi_P(B)\nabla^d\nabla_s^D y_t = \theta_q(B)\Theta_Q(B)\varepsilon_t$$

where:

y_t is the forecast variable (i.e. PV production)

$\varphi_p(B)$ is the regular AR polynomial of order *p*

$\theta_q(B)$ is the regular MA polynomial of order *q*

$\Phi_P(B)$ is the seasonal AR polynomial of order *P*

$\Theta_Q(B)$ is the seasonal MA polynomial of order *Q*

The differentiating operator ∇^d and the seasonal differentiating operator ∇_s^D eliminate the non-seasonal and seasonal non-stationarity, respectively. *B* is the backshift operator, which operates on the observation y_t by shifting it one point in time (i.e. $B^k\left(y_t\right) = y_{t-k}$). The term ε_t follows a white noise process and *s* defines the seasonal period. The polynomials and all operators are defined mathematically as follows:

$$\varphi_p(B) = 1 - \sum_{i=1}^{p} \varphi_i B^i \qquad \Phi_P(B) = 1 - \sum_{i=1}^{P} \Phi_i B^{s,i}$$

$$\theta_q(B) = 1 - \sum_{i=1}^{q} \theta_i B^i \qquad \Theta_Q(B) = 1 - \sum_{i=1}^{Q} \Theta_i B^{s,i}$$

$$\nabla^d = (1-B)^d \qquad \nabla_s^D = (1-B^s)^D$$

The model development was based on the Box-Jenkins methodology, which consists of four iterative steps: a) Identification, b) Estimation, c) Diagnostic Checking and d) Forecasting. Further details on the description of the adopted methodology can be found in [12].

The selection of the most appropriate model was based on the calculation of the day-ahead forecast error using the NRMSE for different model orders and taking into account only the daylight hours. The model with the smallest average yearly NRMSE was selected as the most suitable.

The selected SARIMA model was further improved by incorporating short-term solar radiation forecasts derived from NWP models. For this purpose, the initial hourly forecasts (output of the pure SARIMA model) were multiplied by the following factor:

$$K = \frac{R'_d}{f_1 \cdot R_{d-1} + f_2 \cdot R_{d-2} + .. + f_P \cdot R_{d-P}}$$

where

R'_d denotes the total radiation forecast of day *d* (forecast day),

R_{d-i}, $i = 1, 2, ..., P$, denotes the total real radiation of each of the previous *P* days to day *d*, where *P* is the order of the seasonal AR polynomial of the SARIMA model, and

f_i, $i = 1, 2, ..., P$, are weighting factors calculated on the basis of the respective coefficients of the seasonal AR polynomial of the SARIMA model, as follows:

$$f_i = \Phi_i \Big/ \sum_{i=1}^{P} \Phi_i$$

Finally, the modified hourly forecasts are calculated as follows:

$$F'_{SARIMA}(t) = K \cdot F_{SARIMA}(t)$$

B. Artificial Neural Networks

Artificial Neural Networks (ANNs) have been successfully used in various forecasting applications. They are based on the operation of biological neural networks and supposedly possess the ability of a human-like learning process. A typical ANN structure consists of an input layer, a hidden layer and an output layer. Each layer is comprised of neurons that process the input signals and produce an output, while connections between the layers have a weight factor. An ANN easily adjusts to any set of input-output patterns and through a robust training process forms a model function with the minimum possible error.

Several ANN models were investigated for day-ahead forecasting of power output in single PV parks. The proposed models have one output, namely the forecasted PV power generation *P(D,H),* where *D=forecast day, H=forecast hour.*

Thus, the same formed model is used consecutively for all hours of the forecast day. Many inputs were tested based on the autocorrelation function of the power time series and the strong diurnal periodicity. Two of the ANN models are presented in Table I. Both of them have 7 inputs, 1 output and one hidden layer with 4 neurons.

TABLE I
ANN MODELS

ANN Model	Inputs	Output
A	P(D-1,H), P(D-1,H-1), P(D-2,H), P(D-3,H), P_{AVE}(H), R_{MAX}(H), R'(D,H)	P(D,H)
B	ΔP(D-1,H), ΔP(D-1,H-1), ΔP(D-2,H), ΔP(D-3,H), ΔR'(D,H), ΔR'(D,H-1), ΔR'(D,H+1)	ΔP(D,H)

where

P(D,H) is the power output of the PV plant

$P_{AVE}(H)$ is the average *P(d, H)* for d=D-4,…, D-10

R'(D,H) is the available Radiation Forecast

$R_{MAX}(D,H)$ is the maximum (clear sky) radiation for day *D* and hour *H*

$\Delta F(D,H)=F(D,H) - F(D-1,H)$ is the daily deviation referring either to values of power output (*P*) or radiation forecasts (*R'*).

III. TEST RESULTS

The developed forecasting models were applied for the day-ahead and intraday hourly PV power generation forecast of four PV plants located at different sites throughout the Greek territory.

The data set includes: a) the hourly real PV power generation of the four PV plants and b) the hourly radiation measurements and the respective forecasts collected from four weather stations located adjacent to the PV plants sites. The available data range covers the time period from 1 Jan 2011 to 31 Dec 2012.

For each of the four PV time series, a variety of SARIMA models in terms of the polynomial order of the non-seasonal and seasonal components was examined. The SARIMA $(3,1,2)x(3,1,2)_{24}$ model was found to exhibit the best performance for all sites in terms of the average yearly NRMSE.

In the following paragraphs, simulation results from the application of all forecasting models in a single PV plant (Plant A) for a single day of 2012 as well as total statistics for the entire 2012 are presented. In addition, test results from the application of the forecasting models in four PV Plants (Plants A, B, C, D) located in different sites for a single day of 2012 are provided.

In order to evaluate the performance of the proposed forecasting models, they are all compared to the persistence model. In this paper, persistence is defined as if PV generation in each hour of the forecast day equals the real PV generation of the respective hour of the previous day.

A. Simulation Results - PV Plant A

PV plant A is located in the prefecture of Attica, outside Athens. Its nominal capacity is 0.15 MW, whereas the solar radiation data (real measurements and forecasts) are provided by a weather station located at Spata, very close to the PV plant.

Figures 1-4 illustrate the performance of the examined forecasting models for a spring day (i.e. 26 May 2012). Both ANN models examined, namely A and B, exhibit similar performance throughout the year (see also Table II). For the sake of clarity, only the simulation results obtained with the ANN-Model B are presented in those figures.

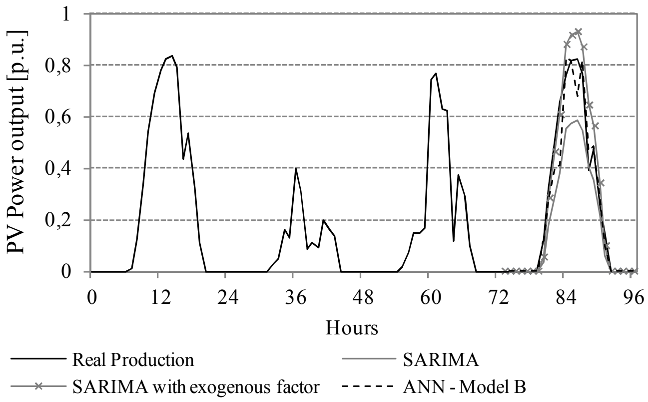

Fig. 1 Day-ahead forecast for Plant A - 26/05/2012

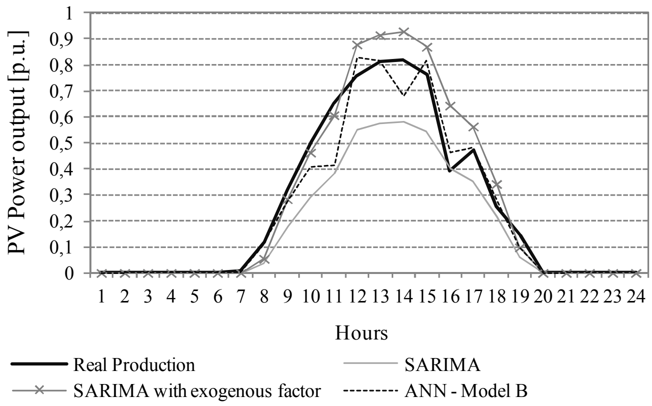

Fig. 2 Day-ahead forecast for Plant A - 26/05/2012 (high resolution)

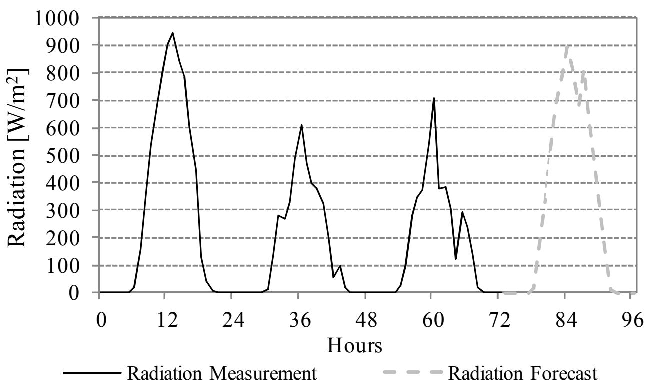

Fig. 3 Solar radiation data - 26/05/2012

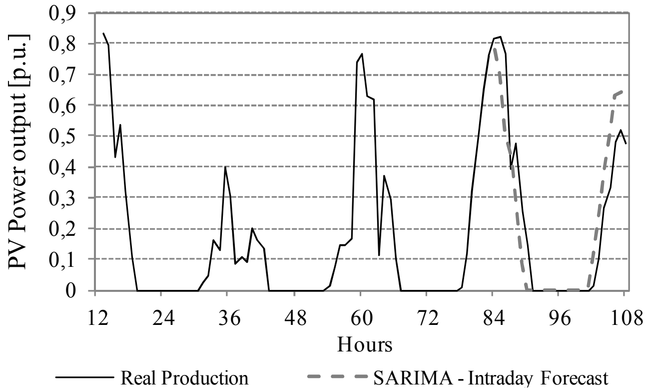

Fig. 4 Intraday forecast with SARIMA model - 26/05/2012

TABLE II
FORECAST ERRORS - PV PLANT A

Forecast Model		NRMSE [%] Winter	Spring	Summer	Autumn	Average Yearly
Day-Ahead Forecasting	Persistence	20,35	18,00	**3,17**	13,33	13,71
	SARIMA $(3,1,2) \times (3,1,2)_{24}$	18,89	16,66	3,55	12,45	12,89
	SARIMA $(3,1,2) \times (3,1,2)_{24}$ with exogenous factor	15,06	14,02	3,61	**11,82**	**11,12**
	ANN - Model A	**14,50**	**12,64**	5,70	12,85	11,42
	ANN - Model B	14,76	13,25	4,06	12,98	11,26

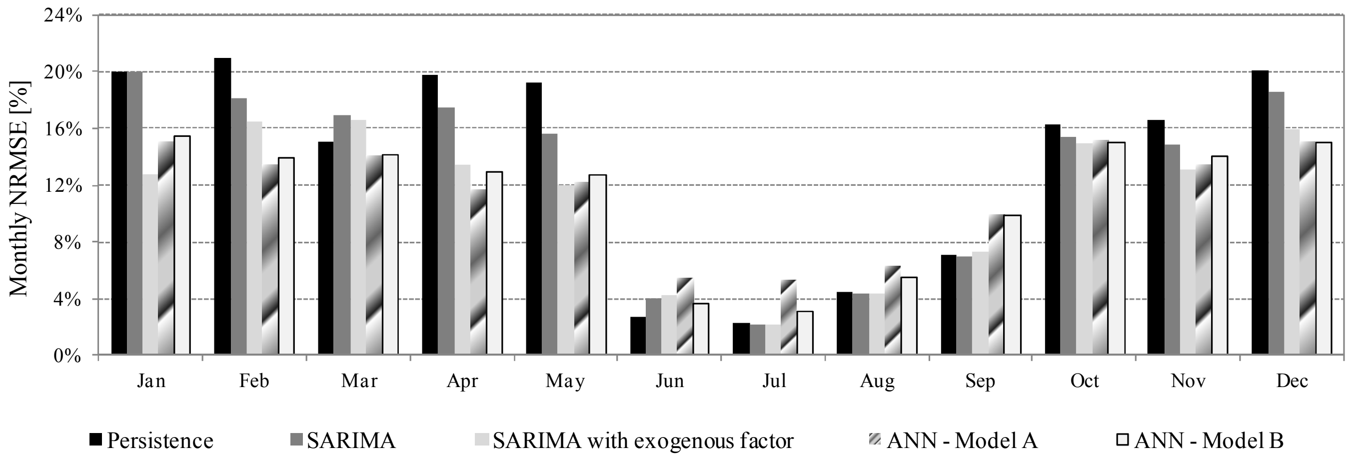

Fig. 5 Average monthly NRMSE for all forecasting models

Figures 1-2 show that the pure SARIMA model cannot predict accurately the day-ahead PV power generation, when the previous days exhibit irregular production patterns with respect to the forecast day. However, the proposed modification of the SARIMA model (i.e. SARIMA model with exogenous factor) improves significantly the day-ahead forecast and leads to comparable forecasts with respect to the ANN-Model B, since both proposed models make use of the available solar radiation data shown in Fig. 3, as already described in Section II.

Regarding the intraday forecasting (e.g. with a 13h time delay), the pure SARIMA model leads to improved forecasts that are very close to the real PV generation (see Fig. 4). This is due to the fact that the pure SARIMA model takes into account the most recent real PV measurements observed during the beginning of the forecast day through its non-seasonal component. In this sense, it is adjusted to the evolving PV generation profile accordingly, thus improving significantly the forecast of the remaining hours of the forecast day.

Table II shows the average seasonal NRMSE as well as the average yearly NRMSE for each forecasting model. The model obtaining the lowest forecast error in each season is shown in bold. It is shown that, in general, the ANN models and the modified SARIMA model are superior in terms of the forecast error as compared to the persistence model and the pure SARIMA model, demonstrating that the use of solar radiation data (real measurements and forecasts) improves significantly the day-ahead forecasting of the PV generation. The only exception lies in summer, where the persistence model and the pure SARIMA model exhibit the lowest errors with all other models obtaining higher errors. This indicates that both these models, which do not make use of solar radiation data, can be used equivalently, due to the clear sky conditions that prevail during summer in Greece. However, the persistence model and the pure SARIMA model lead to higher average yearly NRMSE as compared to the modified SARIMA and the ANN models. This is also made clear in Fig. 5, where it is shown that the modified SARIMA model and the ANN models perform better during the majority of the months of 2012. On the contrary, the persistence model and the pure SARIMA model lead to superior or comparable forecasts only during the summer months and September, where the favorable weather conditions in Greece eliminate the need to use a more sophisticated short-term forecasting model of PV power generation.

B. Simulation Results - All PV plants

In this section, indicative simulation results from the application of the examined forecasting models in four PV Plants (Plants A, B, C, D) located in different zones of the Greek territory for a single autumn day of 2012 (i.e. 13 November 2012) are shown. As already explained in the previous section regarding ANNs, only the simulation results obtained with the ANN-Model B are presented.

Table III shows the main data of the PV plants and the associated weather stations.

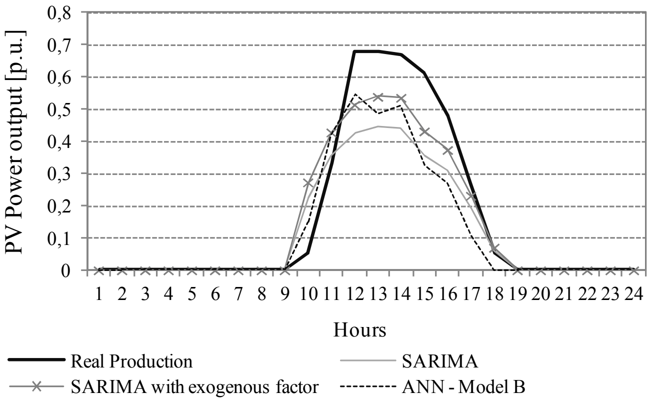

Fig. 6 Day-ahead forecast for Plant A - 13/11/2012

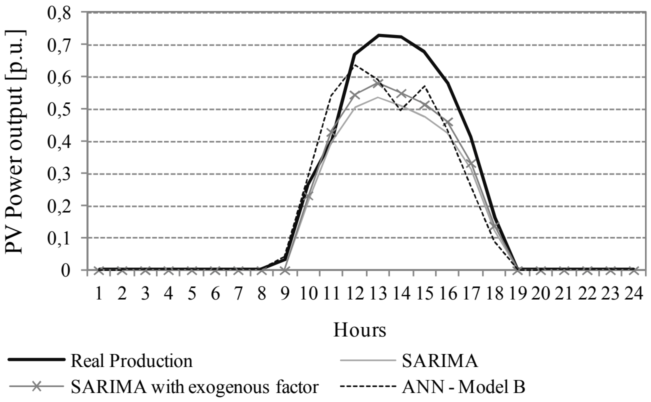

Fig. 7 Day-ahead forecast for Plant B - 13/11/2012

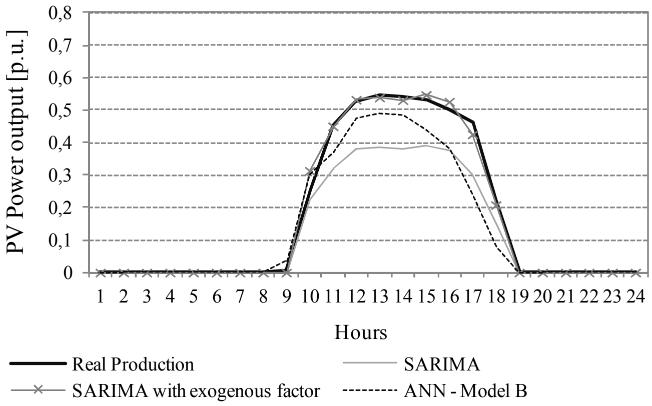

Fig. 8 Day-ahead forecast for Plant C - 13/11/2012

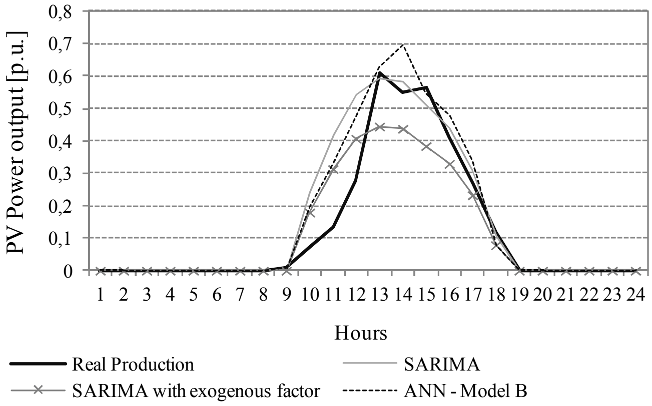

Fig. 9 Day-ahead forecast for Plant D - 13/11/2012

TABLE III
PV PLANTS & WEATHER STATIONS DATA

PV Plant	Location	Nominal Capacity [MW]	Weather Station
A	Attica	0.15	Spata
B	Viotia	1.0	Amfikleia
C	Arta	0.15	Arta
D	Drama	5.0	Drama

Figures 6-9 illustrate the performance of the examined forecasting models for the selected autumn day for the four PV plants. In accordance with the discussion made in the previous section, in all cases the forecasting models that make use of the solar radiation data, namely the SARIMA model with exogenous factor and the ANN model, lead to improved forecasts and lower NRMSEs as compared to the pure SARIMA model. Extensive simulations performed for different days of 2012 for the four PV plants resulted in similar findings, proving the robustness of the proposed forecasting models.

IV. CONCLUSIONS

In this paper a seasonal ARIMA (SARIMA) model that may incorporate short-term solar radiation forecasts derived from NWP models and two artificial neural network models with multiple inputs are presented for the day-ahead and intra-day forecasting of grid-connected PV power plants. The performance of all forecasting models is evaluated in terms of the NRMSE. Test results from the application of all models in different PV plants of the Greek power system show that, in general, those models that make use of available external solar radiation data, namely the modified SARIMA model and the ANN models are preferable, since they lead to considerably improved day-ahead forecasts and lower NRMSEs as compared to the persistence model or the pure SARIMA model. However, both the persistence model and the pure SARIMA model that do not take into account solar radiation data can be equally applied during summer, where the favorable weather conditions in Greece eliminate the need to use a more sophisticated short-term forecasting model of PV power generation.

ACKNOWLEDGEMENTS

The authors are thankful to the Independent Power Transmission Operator of Greece (IPTO or ADMIE) S.A. and the National Observatory of Athens for providing the PV production data and solar radiation data, respectively.

REFERENCES

[1] Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC. [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:140:0016:0062:en:PDF

[2] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 2, no. 1, pp. 2-10, Mar. 2009.

[3] R. Perez, K. Moore, S. Wilcox, D. Renne, and A. Zelenka, "Forecasting solar radiation – Preliminary evaluation of an approach based upon the national forecast database," *Solar Energy*, vol. 81, issue 6, pp. 809-812, Jun. 2007.

[4] P. Mathiesen and J. Kleissl, "Evaluation of numerical weather prediction for intra-day solar forecasting in the continental united states," *Solar Energy*, vol. 85, issue 5, pp. 967-977, May 2011.

[5] 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," *Solar Energy*, vol. 84, issue. 5, pp. 807-821, Jun. 2010.

[6] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," *Solar Energy*, vol. 83, issue 10, pp. 1772-1783, Oct. 2009.

[7] C. Chen, S. Duan, T. Cai, and B. Liu, "Online 24-h solar power forecasting based on the weather type classification using artificial neural network," *Solar Energy*, vol. 85, issue. 11, pp. 2856-2870, Nov. 2011.

[8] G. Reikard, "Predicting solar radiation at high resolutions: A comparison of time series forecasts," *Solar Energy*, vol. 83, issue. 3, pp. 342-349, Mar. 2009.

[9] A. Mellit, M. Benghanem, A. H. Arab, and A. Guessoum, "A simplified model for generating sequences of global solar radiation data for isolated sites: Using artificial neural network and a library of Markov transition matrices approach," *Solar Energy*, vol. 79, no. 5, pp. 469-482, Nov. 2005.

[10] M. Cococcioni, E. D'Andrea, and B. Lazzerini, "24-hour-ahead forecasting of energy production in solar PV systems," in *Proc. of the International Conference on Intelligent Systems Design and Applications 2011*, Cordoba, Spain, Nov. 22–24, 2011.

[11] H.M. Diagne, M. David, P. Lauret, and J. Bolan, "Solar irradiation forecasting: State-of-the-art and proposition for future developments for small-scale insular grids," in *Proc. of WREF 2012*, Denver, Colorado, May 2012.

[12] G.E. Box and G. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden Day, 1976.