ARIMA modeling of the performance of different photovoltaic technologies

Alexander Phinikarides[a], George Makrides[a], Nitsa Kindyni[a], Andreas Kyprianou[b], George E. Georghiou[a]

[a]Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, 1678, Cyprus
[b]Department of Mechanical and Manufacturing Engineering, University of Cyprus, Nicosia, 1678, Cyprus

Abstract — **In this paper, the performance of different technology photovoltaic (PV) systems was modeled using autoregressive integrated moving average (ARIMA) processes. Measurements from mono-crystalline (mono-c-Si), multi-crystalline (multi-c-Si) and amorphous (a-Si) silicon, cadmium telluride (CdTe) and copper indium gallium diselenide (CIGS) systems were used to construct monthly dc performance ratio (PR) time-series, from outdoor measurements. Each PR time-series was modeled a) with multiplicative ARIMA, b) with linear regression and c) with Seasonal-Trend Decomposition by Loess (STL) using the first 4 years of each time-series in order to compare the accuracy of the different methods. The models were used to forecast the PR of the 5th year of the different PV technologies and the results from the aforementioned statistical methods were compared based on the root-mean-square error (RMSE). The results showed that ARIMA produced the lowest RMSE for crystalline silicon (c-Si) technologies, whereas for thin-film technologies, STL was more accurate. The results from ARIMA also showed that thin-film technologies were optimally modeled with identical model orders, whereas for c-Si, each technology required a different optimal model order.**

Index Terms — **photovoltaic systems, modeling, autoregressive processes, performance evaluation, forecasting, performance ratio.**

I. Introduction

Modeling the performance of photovoltaic (PV) systems operating in the field is a tedious task due to the multitude of environmental factors affecting their performance and the differences between the seasonal behaviors of different technologies. Studies have proposed performance metrics such as performance ratio (PR) [1] and PVUSA ratings [2] in order to assess the PV power output in relation to the prevailing meteorological conditions and to construct performance time-series on a daily or monthly basis.

In its simplest form, a time-series with a constant trend can be modeled using linear regression, which can be very effective for well-behaved time-series but is sensitive to seasonal variations, non-constant trends, random errors and outliers. In particular, for PV system performance time-series, which exhibit pronounced seasonal variations [3], the need for seasonal adjustment is high. A simple method of seasonal adjustment uses a moving average (MA) or exponential smoothing to average the seasonal variations. A more advanced method is Seasonal-Trend decomposition based on Loess (STL) [4], which uses local polynomial regression to determine the trend and remove it from the time-series in order to compute seasonal components for each time unit by averaging over all periods. Finally, the residuals are calculated by removing the trend and the seasonal components from the original series. An autocorrelation function (ACF) plot of the residuals provides evidence whether they have Gaussian white noise (GWN) properties, indicating good model fit [5].

In this paper, and expanding on the subject of seasonal decomposition, a multiplicative autoregressive integrated moving average (ARIMA) model [6], [7] was fitted to the performance time-series of different technology PV systems. The ARIMA method is much more capable and flexible than other classical methods since it can effectively deal with seasonal variations, random errors and outliers and can therefore be used to remove all autocorrelations in the model residuals. Previous investigations showed that multiplicative ARIMA could be used to model monthly PVUSA metrics [8], to calculate degradation on even a relatively short time-series and reduce the impact of outliers and other errors on the degradation rate [9]. In this investigation, it was shown that the optimal use of multiplicative ARIMA offers a large improvement in forecasting the PR of crystalline silicon (c-Si) technologies over traditional methods, whereas for thin-film technologies, other forecasting methods such as STL, produced better results, due to the weaker seasonal component of their PR time-series.

II. Methodology

A. Outdoor PV testing facility

The outdoor testing facility at the University of Cyprus (UCY) in Nicosia, Cyprus was commissioned in May 2006 and includes, amongst others, 12 different technology grid-connected PV systems of nominal power 1 kW_p each. The PV technologies range from fixed-plane mono- and multi-crystalline silicon (mono-c-Si, multi-c-Si) to amorphous-silicon (a-Si), cadmium telluride (CdTe), copper indium gallium diselenide (CIGS) and others [10]. Table I lists the PV systems analyzed in this paper and their technology.

978-1-4799-3299-3/13/$31.00 ©2013 IEEE

TABLE I
PV SYSTEMS DESCRIPTION

Identifier	Manufacturer	Technology
ucy0-05	Sanyo	mono-c-Si
ucy0-07	Suntechnics	mono-c-Si
ucy0-10	SolarWorld	multi-c-Si
ucy0-11	Schott Solar	multi-c-Si
ucy0-14	Mitsubishi	a-Si
ucy0-13	First Solar	CdTe
ucy0-12	Würth	CIGS

The monitoring of the PV systems started in June 2006. Both meteorological and PV operational measurements are being recorded and stored through the advanced measurement platform. The platform comprises of meteorological and electrical sensors connected to a central data logging system that stores measurement data every second. The meteorological measurements include the total plane-of-array irradiance, G_I, wind speed, S_W, as well as ambient, T_{am}, and module, T_m, temperatures. The operational measurements include maximum power point (MPP) dc current, I_{MPP}, dc voltage, V_{MPP}, dc power, P_{MPP}, as obtained at each PV system output. Measurements from dc P_{mpp} and G_I were used to create time-series of monthly PR ratings.

B. ARIMA model selection and validation

General seasonal time-series can be modeled with multiplicative ARIMA models [11], [12]. Due to the non-linear parametric nature of ARIMA modeling, statistical software is essential in order to calculate model parameters [13], [14]. The general model for multiplicative ARIMA is given in (1) and is abbreviated as ARIMA(p,d,q)(P,D,Q), where p is the AR (auto-regressive) order, d is the differencing order, q is the MA (moving average) order, P is the seasonal AR order, D is the seasonal differencing order and Q is the seasonal MA order:

$$\Phi(T)\Phi_s(T^s)\nabla^d\nabla_s^D y_t = \theta(T)\theta_s(T^s)e_t \qquad (1)$$

where T is the delay operator, $\Phi(T) = (1 - \varphi_1 T - ... - \varphi_p T^p)$ is an auto-regressive polynomial in T of degree p, $\Phi(T^s)$ is an auto-regressive polynomial in T^s of degree P_s, $\theta(T)$ a moving average polynomial in T of degree q, $\theta_s(T^s)$ a moving average polynomial of degree Q_s in T^s. The operator $\nabla^d = (1-T)^d$ is a non-seasonal differencing operator and $\nabla^D = (1-T^s)^D$ is a seasonal differencing operator and captures non-stationarity in the corresponding locations in successive periods.

The observed time-series was initially checked for stationarity and then transformed using differencing in order to achieve stationarity and enable model fitting. A strictly stationary Gaussian time-series is characterized by a constant mean and variance, therefore, no trend. Non-stationarity in time-series is detected through the augmented Dickey-Fuller (ADF) test [15] and can be achieved by differencing, i.e.

$$\Delta y_t = y_t - y_{t-1} \qquad (2)$$

where Δy_t is the series of 1st differences. The ADF tests for unit roots in the time-series and returns a value that the more negative it is, the less likely it is for the time-series to have a unit root and thus more likely to be stationary.

The sample autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the transformed time-series can be used to determine the lags p, q, P, Q of the ARIMA model, as detailed in [16] and are determined by the numbers of the first significant peaks in the ACF and PACF.

The seasonal period identified in the original time-series of monthly dc PR is 12 months. This requires a multiplicative (seasonal) ARIMA model, with additional AR, MA terms and differencing for the seasonal component. In the case of PV performance metrics, the time-series is highly seasonal and thus the test detailed in [17] is used to detect unit-roots (i.e. non-stationarity) in the seasonal component and set the order of seasonal differencing.

The model selection procedure as summarized in this section yields multiple models that fit the data well. In system identification, the optimum model is the one with the lowest order (i.e. parsimonious), with the lowest mean square error (MSE) and the minimum value of the corrected Akaike information criterion (AICc). In order to validate the goodness of fit for the data, the residuals are checked for GWN properties (i.e. un-correlated, normally distributed). The model is also validated through forecasting [11] by comparing the root-mean-square errors (RMSE) of the forecasted values with the observations (i.e. out-of-sample validation).

C. Comparison of forecasting methods

Using the first 4 years of the PR time-series, a comparison of forecasting methods was performed by calculating the RMSE between the observed and forecasted PR values of the 5th year of operation of the PV systems, for ARIMA, STL and linear regression. A small RMSE value indicates a good model fit to the data, whereas a large RMSE value indicates that the chosen model does not fit the data well.

III. RESULTS

In this section, the optimal multiplicative ARIMA model orders for the different technologies are presented. The models in Table II were chosen using the procedure outlined in the previous section and more specifically by checking the residuals for GWN properties within a 95% confidence interval. From Table II it is clear that different technology systems resulted in different optimal ARIMA model orders and that the modeling of thin-film technologies, such as a-Si, CdTe and CIGS, resulted in the same optimal model order. Furthermore, it can be seen that the seasonal component model orders were identical for all systems under study.

TABLE II
OPTIMAL ARIMA MODELS FOR 4 YEARS

Identifier	Technology	ARIMA(p,d,q)(P,D,Q) model
ucy0-05	mono-c-Si	ARIMA(3,0,0)(1,1,0)
ucy0-07	mono-c-Si	ARIMA(1,0,1)(1,1,0)
ucy0-10	multi-c-Si	ARIMA(3,0,1)(1,1,0)
ucy0-11	multi-c-Si	ARIMA(1,0,2)(1,1,0)
ucy0-14	a-Si	ARIMA(1,0,1)(1,1,0)
ucy0-13	CdTe	ARIMA(1,0,1)(1,1,0)
ucy0-12	CIGS	ARIMA(1,0,1)(1,1,0)

The forecasting results for almost all c-Si systems from ARIMA were shown to be more accurate than other methods, based on the lower RMSE. The results for a-Si, CdTe and CIGS showed that ARIMA may not be the ideal modeling method and that STL produced lower RMSE for the next year of forecasting, as can be seen in Table III.

TABLE III
ROOT-MEAN-SQUARE ERRORS FOR THE 5TH YEAR PREDICTIONS

Identifier	Technology	RMSE Linear reg.	RMSE STL	RMSE ARIMA
ucy0-05	mono-c-Si	4.32	2.05	1.94
ucy0-07	mono-c-Si	4.74	1	1.49
ucy0-10	multi-c-Si	4.24	0.95	0.85
ucy0-11	multi-c-Si	3.8	1.3	0.99
ucy0-14	a-Si	2.88	1.45	1.49
ucy0-13	CdTe	2.36	0.6	1.3
ucy0-12	CIGS	3.12	0.96	1.45

Fig. 1 shows the actual monthly DC PR of the c-Si PV systems for all 5 years and the forecast of the 5th year using the 4-year multiplicative ARIMA model, where the dashed lines represent the upper and lower limits of the 95% confidence interval and the solid blue line represents the mean. It is evident that ARIMA was able to successfully capture the temporal characteristics of the monthly PV performance time-series.

In the case of the thin-film systems (ucy0-14, ucy0-13 and ucy0-12), linear regression produced lower RMSE, in comparison to the c-Si technologies. This was due to the weaker seasonality of the thin-film PR time-series, as is evident in Fig. 2, which shows the actual monthly DC PR of the thin-film PV systems for all 5 years and the forecast of the 5th year using the 4-year multiplicative ARIMA model. Fig. 2 also shows that the forecasts of the optimal ARIMA model underestimated the true performance of the systems. A comparison of linear regression, STL and ARIMA showed that STL produced the lowest RMSE and was therefore more suitable for modeling their behavior. The optimal ARIMA model orders for these systems were identical, with differences only in the coefficients, φ, and θ (Eq. 1).

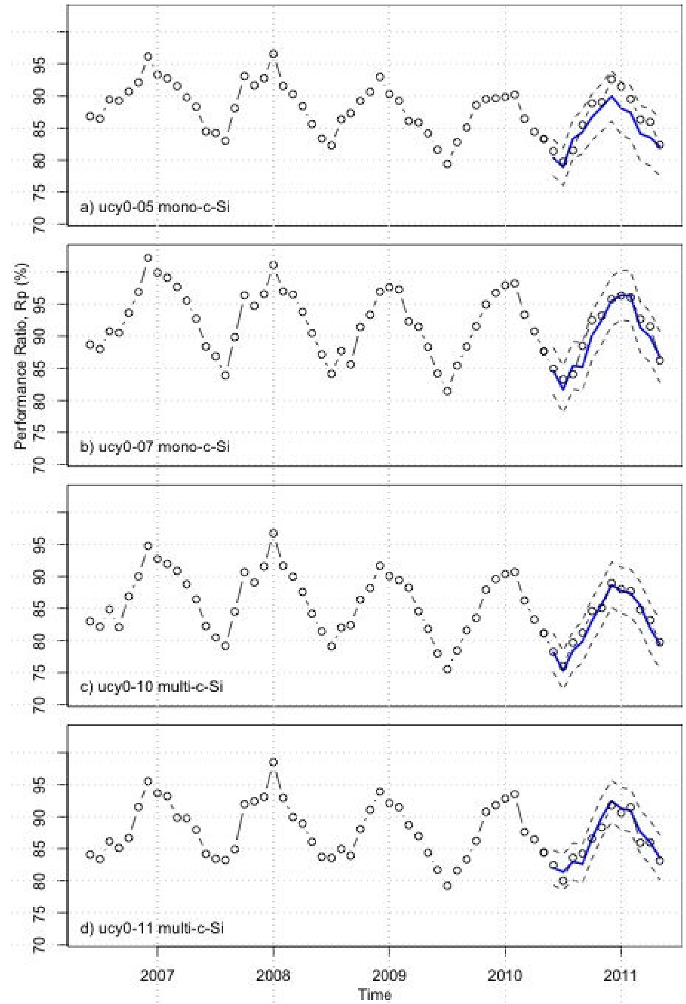

Fig. 1. Monthly DC PR time-series (black circles) and 5th year forecasts (blue line) for the c-Si technologies: (a) ucy0-05 mono-c-Si, (b) ucy0-07 mono-c-Si, (c) ucy0-10 multi-c-Si, (d) ucy0-11 multi-c-Si. The dashed lines represent the upper and lower limits of the 95% confidence interval and the solid blue line represents the mean.

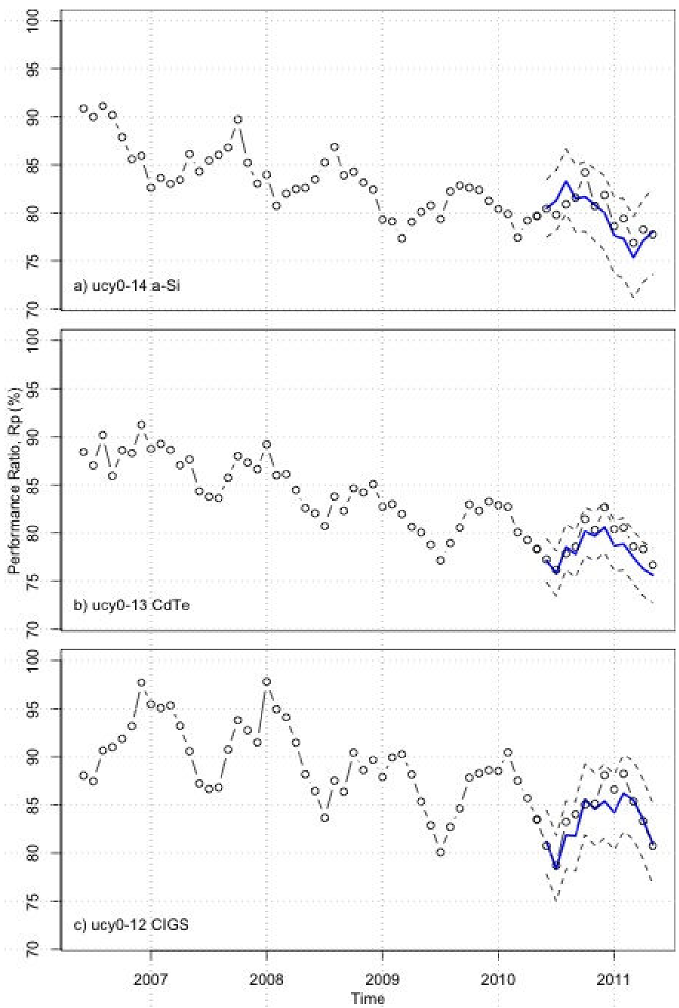

Fig. 2. Monthly DC PR time-series (black circles) and 5th year forecasts (blue line) for the thin-film technologies: (a) ucy0-14 a-Si, (b) ucy0-13 CdTe, (c) ucy0-12 CIGS. The dashed lines represent the upper and lower limits of the 95% confidence interval and the solid blue line represents the mean.

IV. Conclusions

The modeling of performance time-series of different technology PV systems using multiplicative ARIMA was investigated in this paper. The results showed that ARIMA was able to produce accurate modeling estimates with residuals exhibiting GWN properties within a 95% confidence interval. The results have proven that forecasting using these models produced a good estimate of the performance for the next year for c-Si technologies.

Furthermore, it was evident that different technologies required different model orders. This was due to slight differences in their monthly behavior and seasonality. From the definition of a stationary series it can be inferred that the systems requiring non-seasonal and seasonal differencing in ARIMA exhibit a change in their monthly and yearly mean respectively, which may be attributed to steady performance loss. All systems in this study required one order of seasonal differencing, and furthermore, the seasonal model orders were identical for all systems. From this it can be inferred that these PV systems, under the meteorological conditions in Cyprus, were successfully modeled with the same seasonal (annual) profile, but with differences in the coefficients, φ, and θ.

Finally, a comparison of the different forecasting methods used in this paper (linear regression, STL and ARIMA) showed that ARIMA provided better forecasts for c-Si PV systems, whereas for thin-film systems, STL produced better forecasts.

References

[1] IEC 61724:1998, *Photovoltaic system performance monitoring - Guidelines for measurement, data exchange and analysis*, 1st ed. Geneva, Switzerland: IEC, 1998.

[2] A. Kimber, T. Dierauf, L. Mitchell, C. Whitaker, T. Townsend, J. NewMiller, D. King, J. Granata, K. Emery, C. Osterwald, D. Myers, B. Marion, A. Pligavko, A. Panchula, T. Levitsky, J. Forbess, and F. Talmud, "Improved test method to verify the power rating of a photovoltaic (PV) project," in *34th Annual Conference of IEEE Industrial Electronics, IECON*, 2009, pp. 316–321.

[3] G. Makrides, B. Zinsser, A. Phinikarides, M. Schubert, and G. E. Georghiou, "Temperature and thermal annealing effects on different photovoltaic technologies," *Renewable Energy*, vol. 43, pp. 407–417, Jul. 2012.

[4] R. Cleveland, W. Cleveland, J. McRae, and I. Terpenning, "STL: A seasonal-trend decomposition procedure based on Loess," *Journal of Official Statistics*, vol. 6, no. 1, pp. 3–73, 1990.

[5] A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2006.

[6] G. E. P. Box and G. M. Jenkins, "Some recent advances in forecasting and control," *Applied Statistics*, vol. 23, no. 2, pp. 158–179, 1968.

[7] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: forecasting and control*. Prentice Hall, 1994.

[8] A. N. Dunea, D. N. Dunea, V. I. Moise, and M. F. Olariu, "Forecasting methods used for performance's simulation and optimization of photovoltaic grids," in *IEEE Porto Power Tech Proceedings*, 2001, p. 5.

[9] D. C. Jordan and S. R. Kurtz, "Analytical improvements in PV degradation rate determination," in *35th IEEE PVSC*, 2010, pp. 2688–2693.

[10] G. Makrides, B. Zinsser, M. Norton, G. E. Georghiou, M. Schubert, and J. H. Werner, "Outdoor Performance Evaluation of Grid-Connected PV technologies in Cyprus," *Journal of Energy & Power Engineering*, vol. 4, no. 2, pp. 52–57, 2010.

[11] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. 2012.

[12] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to Time Series Analysis and Forecasting*. Wiley, 2008, p. 472.

[13] R. J. Hyndman and Y. Khandakar, "Automatic Time Series Forecasting: The forecast Package for R," *Journal Of Statistical Software*, vol. 27, no. 3, pp. 1–22, 2008.

[14] J. Maindonald and W. J. Braun, *Data Analysis and Graphics Using R: An Example-Based Approach*. Cambridge University Press, 2010.

[15] D. Dickey and W. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, vol. 74, no. 366, pp. 427–431, 1979.
[16] NIST/SEMATECH, *e-Handbook of Statistical Methods*. http://www.itl.nist.gov/div898/handbook/, 2012.
[17] D. R. Osborn, A. P. L. Chui, J. P. Smith, and C. R. Birchenhall, "Seasonality and the order of integration for consumption," *Oxford Bulletin of Economics and Statistics*, vol. 50, no. 4, pp. 361–377, May 2009.