Comparison of trend extraction methods for calculating performance loss rates of different photovoltaic technologies

Alexander Phinikarides, George Makrides, Nitsa Kindyni, George E. Georghiou

FOSS Research Centre for Sustainable Energy, Photovoltaic Technology Laboratory, Department of Electrical and Computer Engineering, University of Cyprus
75 Kallipoleos Avenue, P.O. Box 20537, Nicosia, 1678, Cyprus

Abstract — In this work, the performance loss rates of eleven grid-connected photovoltaic (PV) systems of different technologies were evaluated by applying linear regression (LR) and trend extraction methods to Performance Ratio, \( R_p \), time series. In particular, model-based methods such as Classical Seasonal Decomposition (CSD), Holt-Winters (HW) exponential smoothing and Autoregressive Integrated Moving Average (ARIMA), as well as non-parametric filtering methods such as LOcally wEighted Scatterplot Smoothing (LOESS) were used to extract the trend from monthly \( R_p \) time series of the first five years of operation of each PV system. The results showed that applying LR on the time series produced the lowest performance loss rates for most systems, but with significant autocorrelations in the residuals, signifying statistical inaccuracy. The application of CSD and HW significantly reduced the residual autocorrelations as the seasonal component was extracted from the time series, resulting in comparable results for eight out of eleven PV systems, with a mean absolute percentage error (MAPE) of 6.22% between the performance loss rates calculated from each method. Finally, the optimal use of multiplicative ARIMA resulted in Gaussian white noise (GWN) residuals and produced higher performance loss rates than LR for all technologies, except the amorphous Silicon (a-Si) system. The LOESS non-parametric method produced directly comparable results to multiplicative ARIMA, with a MAPE of -2.04% between the performance loss rates calculated from each method, whereas LR, CSD and HW showed higher deviation from ARIMA, with MAPE of 25.14%, -13.71% and -6.39%, respectively.

Index Terms — degradation, photovoltaic systems, performance, statistical methods.

I. INTRODUCTION

Various outdoor performance studies of grid-connected photovoltaic (PV) systems demonstrated that both crystalline silicon (c-Si) and thin-film PV technologies exhibit strong seasonality and performance variations [1]–[3]. More specifically, for thin-film technologies the performance variations are further correlated with a tendency for performance loss with increasing outdoor exposure, with the Staebler-Wronski effect [4] and thermal annealing [5] which apply strictly to amorphous Silicon (a-Si) [6], and with the transient response of Cadmium Telluride (CdTe) technologies in regards to light soaking [7].

Although performance loss is usually directly observable from the monthly performance ratio, \( R_p \), time series, as a downward trend, further insight into the performance loss rate of each PV technology can be obtained by extracting the trend using statistical methods [8]. Model-based methods such as Linear Regression (LR), Classical Seasonal Decomposition (CSD) [9], [10], Holt-Winters (HW) exponential smoothing [11], [12] and Autoregressive Integrated Moving Average (ARIMA) [13] assume the specification of a stochastic time series model. The non-parametric filtering methods, such as LOcally wEighted Scatterplot Smoothing (LOESS) [14], [15] do not require specification of a model and are popular because of their simplicity.

In the past, CSD, LOESS and multiplicative ARIMA have been successfully used to forecast the \( R_p \) of grid-connected PV systems of different technologies [16] and LR, CSD and LOESS were used to calculate their performance loss rates [17], [18]. Due to the fact that the LR and CSD methods fit a fixed model, particular characteristics of each time series are therefore not captured and this results in significant autocorrelations in the model residuals. Another model-based method is HW, in which triple exponential smoothing is applied to the time series. Triple exponential smoothing takes into account seasonal changes, as well as trends, through the minimization of the squared one-step ahead prediction error, in contrast to CSD, which bases the calculation of trend, seasonal component and residuals on a centred moving average.

Another advanced model-based method used in this evaluation was multiplicative ARIMA, as the method further allowed the calculation of the optimal model orders for each time series. The ARIMA method is more flexible than classical methods since it can effectively deal with seasonal variations, random errors, outliers and level shifts and can therefore be used to specify a model which removes all autocorrelations in the model residuals. Previous investigations showed that multiplicative ARIMA could be used to model monthly PVUSA metrics [19], to calculate degradation on even a relatively short time series, reduce the impact of outliers and other errors on the degradation rate [20] and offer an improvement in forecasting the \( R_p \) of c-Si technologies over classical methods [16]. Due to the complex nature of ARIMA, it is entirely implemented in software [21], [22]. One of the most widely used software packages, X-13-ARIMA [23], developed by the U.S. Census Bureau, was used to apply the parametric X-11 ARIMA method [24], [25] on
the time series and to automatically detect additive outliers and level shifts in the data.

In this work, the performance loss rates from the first five years of operation of different technology grid-connected PV systems installed in Nicosia, Cyprus were evaluated using different methods of trend extraction applied on monthly DC $R_P$ time series for each system. The trend of the $R_P$ time series of each PV technology was extracted using LR, CSD, HW, LOESS and ARIMA and the resulting performance loss rates were calculated from the gradient of the trend. It is important to note that the linear, monthly performance loss rate at the DC side includes the modelling error of each technique, PV array loss factors such as module degradation, soiling, shading, mismatch and initial power stabilization. Additionally, since PV module production is continuously improved, the performance loss rates obtained from this study might be affected by the specific model and batch of modules.

II. METHODOLOGY

A. Outdoor PV system testing

High-resolution measurements from the first five years of operation of different technology, 1 kW$_p$ grid-connected PV systems were used to assess the performance loss rate of each technology. The PV systems were installed in June 2006 at the Photovoltaic Technology test site of the University of Cyprus, in Nicosia, Cyprus and are listed in Table I. Operational and meteorological measurements were recorded every second and averaged every fifteen minutes [26]. The PV technologies range from mono-c-Si, multi-c-Si, Heterojunction with Intrinsic Thin Layer (HIT), Edge defined Film-fed Growth (EFG), Multi-crystalline Advanced Industrial cells (MAIN) to a-Si, Cadmium Telluride (CdTe) and Copper Indium Gallium Diselenide (CIGS).

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Model</th>
<th>Technology</th>
<th>$P_{MPP}$ (kW$_p$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solon</td>
<td>P220/6+</td>
<td>multi-c-Si</td>
<td>1.540</td>
</tr>
<tr>
<td>Sanyo</td>
<td>HIP-205NHE1</td>
<td>mono-c-Si (HIT)</td>
<td>1.025</td>
</tr>
<tr>
<td>Atersa</td>
<td>A-170M 24V</td>
<td>mono-c-Si</td>
<td>1.020</td>
</tr>
<tr>
<td>SolarWorld</td>
<td>SW165</td>
<td>multi-c-Si</td>
<td>0.990</td>
</tr>
<tr>
<td>Sunpower</td>
<td>STM 200 FW</td>
<td>mono-c-Si (back-contact)</td>
<td>1.000</td>
</tr>
<tr>
<td>Schott Solar</td>
<td>ASE-260-DG-FT</td>
<td>multi-c-Si (EFG)</td>
<td>1.000</td>
</tr>
<tr>
<td>BP Solar</td>
<td>BP7185S</td>
<td>mono-c-Si (Saturn)</td>
<td>1.110</td>
</tr>
<tr>
<td>Schott Solar</td>
<td>ASE-165-GT-FT/MC</td>
<td>multi-c-Si (MAIN)</td>
<td>1.020</td>
</tr>
<tr>
<td>Würth Solar</td>
<td>WS 11007/75</td>
<td>CIGS</td>
<td>0.900</td>
</tr>
<tr>
<td>First Solar</td>
<td>FS60</td>
<td>CdTe</td>
<td>1.080</td>
</tr>
<tr>
<td>MHI</td>
<td>MA100T2</td>
<td>a-Si</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Fifteen minute average measurements of the DC power at the maximum power point (MPP), $P_{MPP}$, were used in conjunction with irradiation measurements from a calibrated Kipp & Zonen CM21 pyranometer, mounted on the plane-of-array (POA), in order to construct monthly $R_P$ ratings for the period June 2006 – June 2011. The $R_P$ is defined as the ratio of the final energy yield of the PV system, $Y_f$, and the reference yield, $Y_r$ [27]:

$$R_P = \frac{Y_f}{Y_r}$$

where $Y_f$ is defined as the monthly yield of the PV array normalized by the maximum power, $P_{max}$ of the PV array, provided by the manufacturer and $Y_r$ is calculated by dividing the total monthly POA irradiation by the reference irradiance of 1000 W/m$^2$.

The $R_P$ time series for all PV systems, over the period June 2006 – June 2011 is demonstrated in Fig. 1. The seasonal pattern of all technologies is evident, with differing amplitudes for each technology. Especially the First Solar CdTe and MHI a-Si systems show a weaker seasonal component than the c-Si and CIGS systems, mainly due to the lower temperature coefficients compared to c-Si technologies [28].

By applying temperature correction using the temperature coefficients, specified in the manufacturer datasheets [29], the seasonal component was reduced for all PV technologies, except the a-Si. Fig. 2 shows the monthly $R_P$ for each PV...
system, using temperature corrected fifteen minute average measurements of DC \(P_{MPP}\) and back sheet module temperature, \(T_{mod}\). Evidently, there were still seasonal variations on the \(R_P\), even after normalization of the two major seasonal factors: a) the irradiation, via the \(R_P\), and b) the module temperature via temperature correction. This fact signifies the need for trend extraction, even after performing irradiance and temperature corrections, to eliminate the impact of the seasonal variations on the performance loss rate of each PV technology.

\[\hat{y}_{n+l} = m_n + b_n + c_{n-S+l} \quad l = 1, 2, \ldots \]  
\[m_t = \alpha_0 (y_t - c_{t-S}) + (1 - \alpha_0)(m_{t-1} + b_{t-1}) \]  
\[b_t = \alpha_1 (m_t - m_{t-1}) + (1 - \alpha_1)b_{t-1} \]  
\[c_t = \alpha_2 \frac{y_t}{m_t} + (1 - \alpha_2)c_{t-S} \]

with \(\alpha_0, \alpha_1\) and \(\alpha_2\) lying between 0 and 1.

The general model for multiplicative ARIMA is given in (8) 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.

\[\Phi(T)\Phi_s(T^s)\nabla^d \nabla_s^D y_t = \theta(T)\theta_s(T^s)e_t \]  

In order to find the optimal ARIMA model, the time series was initially checked for stationarity and then transformed using differencing to achieve stationarity, as necessary. The lags \(p, q, P, Q\) of the model were determined from the autocorrelation function (ACF) and the partial autocorrelation function (PACF) [16]. The model selection procedure yielded multiple models that fit the data well. 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, the residuals were checked for Gaussian white noise (GWN) properties (i.e. un-correlated, normally distributed). The optimal multiplicative ARIMA models for the five year \(R_P\) time series of all systems are presented in Table II.

<table>
<thead>
<tr>
<th>PV system</th>
<th>ARIMA model ((p,d,q)(P,D,Q))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solon (P220/6+)</td>
<td>(0 0 0)(0 1 1)</td>
</tr>
<tr>
<td>Sanyo (HIP-205NHE1)</td>
<td>(0 1 1)(1 0 0)</td>
</tr>
<tr>
<td>Atersa (A-170M 24V)</td>
<td>(1 0 1)(0 1 1)</td>
</tr>
<tr>
<td>SolarWorld (SW165)</td>
<td>(0 1 1)(0 1 2)</td>
</tr>
<tr>
<td>Sunpower (STM 200 FW)</td>
<td>(1 0 1)(0 1 1)</td>
</tr>
<tr>
<td>Schott Solar (ASE-260-DG-FT)</td>
<td>(0 1 1)(1 1 0)</td>
</tr>
<tr>
<td>BP Solar (BP7185S)</td>
<td>(0 1 2)(0 1 1)</td>
</tr>
<tr>
<td>Schott Solar (ASE-165-GT-FT/MC)</td>
<td>(0 1 1)(0 1 2)</td>
</tr>
<tr>
<td>Würth Solar (WS 11007/75)</td>
<td>(0 1 1)(0 1 1)</td>
</tr>
<tr>
<td>First Solar (FS60)</td>
<td>(0 1 1)(0 1 1)</td>
</tr>
<tr>
<td>MHI (MA100T2)</td>
<td>(0 1 1)(0 1 1)</td>
</tr>
</tbody>
</table>
C. Non-parametric filtering methods

Non-parametric filtering methods are different than model-based methods in the sense that an explicit model is not specified. One such method is LOESS, which extracts the trend from locally weighted polynomial fitting. LOESS provides robust estimates of the trend and seasonal components that are not distorted by outliers and missing values [30]. For $R_p$ time series this is especially important since the ratings are prone to distortions due to system outages, sensor faults and uncertainty, missing values and irregular time series.

III. RESULTS

In the case of LR, the annual performance loss rates were calculated by multiplying the slope coefficient by 12 (12 months in a year). In the case of CSD, HW, ARIMA and LOESS, the rates were calculated by first applying linear regression on the extracted trend and then multiplying the gradient again by 12 to obtain the annual performance loss rate, with positive values defining a gradual performance loss.

![Fig. 3. Performance loss rates for each PV system using different trend extraction methods, over the period June 2006 – June 2011.](image)

Fig. 3 shows the distribution of calculated performance loss rates for all PV systems using the various methods. The results clearly demonstrate that different trend extraction methods provide different performance loss rates, especially in comparison to the simple LR. The manufacturer’s stated annual performance loss limit of 0.8–1.0 %/year was surpassed by all thin-film systems, which exhibited performance loss rates around 2.3 %/year on average for CIGS and CdTe and 1.72 %/year on average for a-Si, whereas the c-Si technologies exhibited lower performance loss rates, in the range of 0.25 to 1.37 %/year.

In addition, the results showed that the application of LR on the time series produced the lowest performance loss rates for nine out of eleven systems but with significant autocorrelations in the residuals, detected from the residual ACF and PACF plots, signifying model inadequacy. The application of CSD and HW significantly improved the residual autocorrelations as a seasonal component was extracted from the time series, resulting in comparable results between these two methods for eight out of eleven PV technologies, with a mean absolute percentage error (MAPE) of 6.22 % between the performance loss rates calculated from each method. The application of HW also improved the variation of the rates for the Solon system, but performed worse for the BP Solar system, which was partially shaded for short periods of time due to foliage, during the evaluation. Unlike LR, all other trend extraction methods produced very similar performance loss rates between them for each PV system, except for the BP Solar and Solon systems. This, along with the more accurate modelling results, in comparison to LR, validates the statistical significance and superiority of trend extraction approaches for calculating performance loss rates from the $R_p$ time series of outdoor exposed, grid-connected PV systems.

Finally, the optimal use of multiplicative ARIMA resulted in Gaussian white noise (GWN) residuals and the most accurate statistical model of the $R_p$ time series. ARIMA produced higher performance loss rates than LR for all technologies, except the a-Si system. The LOESS non-parametric method produced directly comparable results to multiplicative ARIMA, with a MAPE of -2.04 % between the performance loss rates calculated from each method, whereas LR, CSD and HW showed higher deviation from ARIMA, with MAPE of 25.14 %, -13.71 % and -6.39 %, respectively.

IV. CONCLUSIONS

In this work, trend extraction methods were presented in the context of calculating performance loss rates from monthly $R_p$ time series of grid-connected PV systems of different technologies. In contrast to the commonly used LR, trend extraction methods were able to separate the time series into their trend, seasonal and residual components. The complexity of each trend extraction method was directly correlated to its modelling accuracy and subsequently to its robustness in calculating statistically accurate performance loss rates. The exception to this rule was offered by the non-parametric
LOESS method, which, despite its easier usage, has shown nearly identical results to the more advanced ARIMA method.

Finally, although LR produced the lowest performance loss rates for nine out of eleven PV systems in this study, itsapplication was very inaccurate since there was a prevalent seasonal component in the $R_p$ time series, which could not be modelled with a linear fit. It has also been shown that temperature correction of the DC $P_{app}$ based on the manufacturer’s temperature coefficients and measurements of module temperature and then normalization with respect to irradiance could not remove the seasonal component in its entirety. Statistically accurate performance loss rates were then calculated using multiplicative ARIMA and the comparison between ARIMA and LOESS has shown that LOESS was also able to calculate statistically accurate results, despite the methods’ differing nature.

ACKNOWLEDGEMENTS

This work has been co-financed by the European Regional Development Fund and by the Republic of Cyprus in the framework of the project ‘Reliable Assessment of Degradation on new thin-film photovoltaic technologies’ with grant number TEXNOLOGIA/ENEPT/0311(BIE)/12.

REFERENCES


