Πανεπιστήμιο Κύπρου
University of Cyprus

*Master Thesis*

# Optimal methodology for forecasting performance loss rate of fielded photovoltaic systems using Robust Principal Component Analysis and Auto-Regressive Integrated Moving Average

Anna Michail

***Department of Mechanical and Manufacturing Engineering***

DECEMBER 2021

# UNIVERSITY OF CYPRUS

## Department of Mechanical and Manufacturing Engineering

***MASTER THESIS***

# Optimal methodology for forecasting performance loss rate of fielded photovoltaic systems using Robust Principal Component Analysis and Auto-Regressive Integrated Moving Average

**Anna Michail**

December 2021

Supervised by:
Ass. Prof. Andreas Kyprianou and Prof. George E. Georghiou

*The present Master dissertation was submitted in partial fulfilment of the requirements for the degree of Master of Mechanical Engineer of the University of Cyprus.*

Photovoltaic Technology Laboratory, FOSS Research Centre for Sustainable Energy,

University of Cyprus

# *Abstract*

In the last decades photovoltaic (PV) systems penetration in the energy mix is rapidly increasing with a total cumulative installed capacity of 758.9 GW in 2020, as reported by the International Energy Agency (IEA). For a further sustainable PV growth, it is crucial to assure the lifetime energy yield and the formulation of strict warranties to reduce investment risk. To achieve this, the accurate evaluation of the performance loss rate (PLR) of fielded PV systems is necessary. The PLR includes fault and loss events as well as PV module degradation and its estimation is technology and methodology dependent.

The aim of this thesis is to develop an optimal methodology for forecasting the annual PLR of fielded PV systems. Annual PLRs based on forecasted performance ratio (PR) were computed using a Seasonal Auto Regressive Integrating Moving Average (SARIMA) model and robust principal component analysis (RCPA). The forecasted PLRs were compared with actual PLRs obtained over an 8-year period for eleven different PV technologies installed in Nicosia, Cyprus.

The SARIMA method was used to model the PR time series of each PV system and the orders $p, d, q, P, D, Q$ and $s$ of each model were derived using four different methods. Three of them were based on three different information criterions, namely the Akaike Information Criterion (AIC), the modified AIC for small sample size (AICc) and the Bayesian Information Criterion BIC. The fourth method was based on an empirical evaluation of the autocorrelation function (ACF), partial ACF (PACF) and a statistical residuals analysis.

Each devised model was used to forecast the monthly PR for 3 years, using a 5-year period of monthly PR measurements as the train dataset. All methods showed good fitting for the forecasted PR, with the RMSE and MAE values being below 6%. The PLR was computed at the end of each forecasted year using the RPCA method. The results showed that the actual and forecasted PLR values were in agreement with the PLR values reported in literature for the c-Si and thin-film technologies, confirming the reliability of the RPCA method to calculate the PLR. Comparing the actual and forecasted PLR values, all methods manage to estimate the PLR sufficiently exhibiting absolute differences up to $1.2\ \%/yr$, with the BIC method achieving the lowest difference of $0.51\ \%/yr$.

The optimal methodology for estimating the PLR was derived based on a comparative analysis. Specifically, the comparative analysis was based on the forecasting accuracy of the four methods for both PR and PLR values, the statistical significance and their simplicity. The method that achieved the lowest total score was considered the preferred one. The models based on BIC had the lowest scores for the forecasting accuracy, but the statistical significance scores of the method was quite high. However, the BIC method achieved the most preferred score with a total score of 0.01807. Furthermore, the total scores based on AIC and AICc were slightly higher than the BIC, with the method based on AIC exhibit the second lower value with a total score of 0.01939 and the method based on AICc having a total score of 0.02049. Last but not least, the methodology based on ACF and PACF had the lowest score regarding the residual's behaviour showing that the residuals of this method are closest to white noise behaviour. Depside that, this methodology achieved the total score of 0.02181, which was the higher score among the methods, ranking it to the least preferred method.

# *Acknowledgements*

First and foremost, I would like to express my gratitude towards my supervisor Ass. Prof. Andreas Kyprianou and co-supervisor Prof. George E. Georghiou for the valuable advices they have offered me, for the trust they have shown in me and for giving me the opportunity to work on such an exciting project to deploy the present thesis.

I am also deeply grateful to Mr. Andreas Livera for his useful guidance and assistance through the fulfilment of this project.

Furthermore, I would like to thank my friends and family for their unconditional love and support all the time.

Finally, I would like to thank the Photovoltaic Technology Laboratory of FOSS Research Centre for Sustainable Energy for granting me access to their data and its staff for their continued support throughout my studies at University of Cyprus.

# *Contents*

# List of Figures

# List of Tables

# *Acronyms*

| | |
|---|---|
| Abs Dif | Absolute error of difference |
| AC | Alternating Current |
| ACF | Autocorrelation Function |
| AIC | Akaike Information Criterion |
| AICc | Akaike Information Criterion corrected for small sample size |
| ALM | Augmented Lagrange Multiplier |
| AR | Auto-Regressive |
| ARIMA | Auto-Regressive Integrating Moving Average |
| ARMA | Auto-Regressive Moving Average |
| a-Si | Amorphous Silicon cells |
| BIC | Bayesian Information Criterion |
| CdTe | Cadmium Telluride cells |
| CIGS | Copper Indium Gallium Selenide cells |
| CSD | Classical Seasonal Decomposition |
| c-Si | Crystalline Silicon cells |
| DC | Direct Current |
| Dif | Error of difference |
| EFG | Edge defined Film-fed Growth |
| HIT | Heterojunction with Intrinsic Thin layer cells |
| iALM | inexact Augmented Lagrange Multiplier |
| IEA | International Energy Agency |
| IV | Current-Voltage |
| KPI | Key Performance Indicator |
| LOESS | Locally Weighted Smoothing (LOcal regrESSion) |
| LR | Linear Regression |
| MA | Moving Average |
| MAE | Mean Absolute Error |
| MAIN | Multi-crystalline Advanced Industial cells |

| | |
|---|---|
| MHI | Mitsubishi Heavy Industries |
| Mono-c-Si | Mono-crystalline Silicon cells |
| MPP | Maximum Power Point |
| Multi-c-Si | Multi-crystalline Silicon cells |
| NREL | National Renewable Energy Laboratory |
| OLS | Ordinary Least-Squares |
| PACF | Partial Autocorrelation Function |
| POA | Plane of Array |
| PTC | PVUSA test conditions ($G_I = 1000\ W/m^2,\ T_{amb} = 20$ ℃ and $W_S = 1\ m/s^2$) |
| PV | Photovoltaic |
| PVUSA | Photovoltaics for Utility-Scale Applications |
| RMSE | Root Mean Square Error |
| RPCA | Robust Principle Component Analysis |
| SARIMA | Seasonal Auto-Regressive Integrating Moving Average |
| STC | Standard Test Conditions (1000 W/m^2 irradiance, AM1.5 air mass, 25°C cell temperature) |
| Thin-film | Thin Film cells |
| UCY | University of Cyprus |
| YoY | Year-over-Year |

# *Symbols*

| SYMBOL | DESCRIPTION | UNIT |
|---|---|---|
| $A$ | PV array area | $m^2$ |
| $E_A$ | PV array output energy (DC) | $kWh$ |
| $FF$ | Fill factor | % |
| $G_I$ | In-plane irradiance (POA) | $W \cdot m^{-2}$ |
| $G_{I_{STC}}$ | Reference in-plane irradiance (under STC), equal to $1000\ W/m^2$ | $W \cdot m^{-2}$ |
| $H_I$ | In-plane irradiation | $kWh \cdot m^{-2}$ |
| $I_A$ | Array current | $A$ |
| $I_{MPP}$ | Maximum Power Point Current | $A$ |
| $I_{SC}$ | Shot Circuit Current | $A$ |
| $P_0$ | Array power rating (DC) | $kW$ |
| $P_A$ | Array power (DC) | $W$ |
| PLR | Performance Loss Rate | $\%/year$ |
| $P_{MPP}$ | Maximum Power Point Power | $W$ |
| $P_{PTC}$ | Power extrapolated to PTC | $W$ |
| $PR$ | Performance Ratio | % |
| $P_{TU}$ | Power to the utility grid | $W$ |
| $RH$ | Relative humidity | % |
| $R_S$ | Series resistance | $Ohms$ |
| $R_{SH}$ | Shunt resistance | $Ohms$ |

| SYMBOL | DESCRIPTION | UNIT |
|---|---|---|
| $T_{amb}$ | Ambient air temperature | ℃ |
| $T_{mod}$ | Module temperature | ℃ |
| $V_A$ | Array voltage | $V$ |
| $V_{MPP}$ | Maximum Power Point Voltage | $V$ |
| $V_{OC}$ | Open Circuit Voltage | $V$ |
| $W_a$ | Wind direction | |
| $W_S$ | Wind speed | $m/s^2$ |
| $Y_A$ | PV array energy yield | $kWh/kW$ |
| $Y_F$ | Final system yield | $kWh/kW$ |
| $Y_r$ | Reference yield | $kWh/kW$ |
| $\gamma_{manuf}$ | Temperature coefficient of power from the module specifications | $\%/K$ |
| $\eta_{STC}$ | PV module efficiency rated under STC | % |

# Notation

| | |
|---|---|
| $\nabla^d$ | Difference operator of dth order |
| $a$ | Significance level |
| $B$ | Back-shift operator |
| $H_a$ | Alternative Hypothesis |
| $H_o$ | Null Hypothesis |

# *Chapter 1 – Introduction*

Over the last decades, a rapidly increasing penetration of photovoltaic (PV) systems in the energy mix is evidenced. Figure 1.1 illustrates the globally total cumulative installed capacity of PV systems over the years, reaching 758.9 GW by the end of 2020 as reported by the International Energy Agency (IEA) [1]. As indicated by IEA, PV is the fastest growing renewable energy technology in terms of global installation. A key factor that will enable the further uptake of the technology is the reduction of the PV electricity costs by increasing the lifetime output. This can be achieved by improving the reliability and service lifetime performance. In this sense, a main challenge in the quest for ensuring quality of operation especially for grid-connected PV systems is to safeguard reliability and good performance by identifying and quantifying accurately the factors behind the various performance loss mechanisms, while also detecting and diagnosing potential failures at early stages [2], [3].

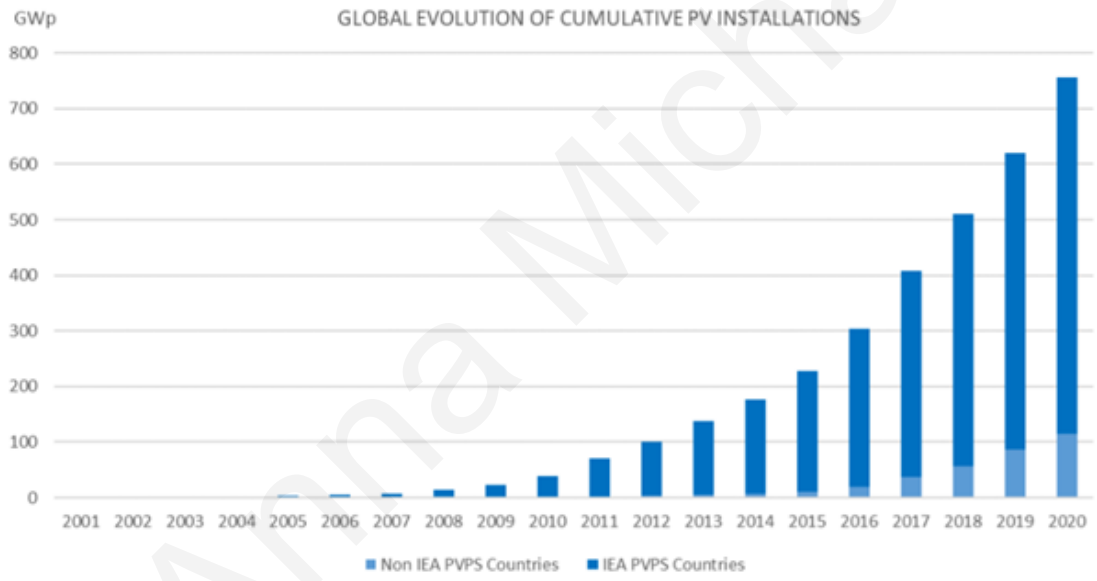


**Figure 1.1. Global evolution of cumulative PV installations** [1]**.**

## 1.1 Thesis motivation

PV systems over their lifetime develop degradation modes due to fault and loss events as well as PV module degradation. These degradation modes can be attributed to environmental and PV internal factors such as temperature, humidity, soiling (the accumulation of dust, dirt, and other particles that cover the surface of the module), solar irradiance, shading, module/cell cracks and module mismatches etc [4]. The aforementioned factors accelerate different degradation modes and impose significant stress over the lifetime of a PV module, resulting in the reduction of durability, which must be quantified through the estimation of the performance loss rate (PLR). Knowledge of the PLR is crucial for reducing uncertainties and financial risks [3]. It is important to

note that up to date, there is no standardized method for accurate calculating the PLR of fielded PV systems [5], [6].

For instance, warranties offered by the manufacturers of PV modules guarantee that over the 25 to 30 years' service lifetime of the modules the power output as calculated under standard test conditions (STC), $P_0$, will exhibit a maximum of $20\,\%$ degradation [4], [7]. More recently, the manufactures include in their guarantee a maximum linear degradation rate of the $P_0$ per year. Usually, guarantees declare the $2\,\%$ of power degradation rate during the first year and not higher than $0.4\,\%/yr$ to $0.8\,\%/yr$ thereafter, depending on the manufacturer [8]. These values are not carried out by testing the PV modules to the end of their lifetime in the field, as they are taken under laboratory conditions at STC.

In the case of fielded PV modules to measure their performance and compare it with the indications from the warranties, their operation is needed to be disturbed and measured under laboratory conditions at STC. This leads to high enough measurement uncertainties that difference lower than $2\,\% - 3\,\%$ from the $P_0$. On the other hand, by fully testing the module at STC, only a small part of field performance is tested, as such conditions rarely occur in outdoor environments. Consequently, the overall performance of the PV system in the field must be analysed to come to conclusions on the actual PLR.

The aforementioned reasons further urge the need for establishing a standardized methodology for accurately estimating the PLR of fielded PV modules and systems. This will assist to assure the lifetime energy yield and the formulation of strict warranties to reduce investment risk, leading to a further sustainable growth of PV technology.

## 1.2 Research objectives

The main focus of the present study is to establish an optimal methodology for estimating accurately the PLR of fielded PV systems. The PLR estimation is generally based on a performance metric which will assess the performance of the PV system and its value will reflect the losses on the PV rated output. In the present thesis the performance ratio (PR) was used as the performance metric. The main research objectives of the present thesis work are summarised as follows:

- Precise capture the behaviour of the PR time series over the years for each PV system using a seasonal auto-regressive integrating moving average (SARIMA) model and forecast future values of the PR time series.
- Estimate accurately the PLR for the upcoming years using the forecasted values of the PR time series.
- Evaluate all the methods proposed and discussed in this study to derive a generalized methodology for the estimation of the PLR for fielded PV systems.

## 1.3 Outline of the thesis

This thesis is divided and organized into five chapters. In particular, Chapter 1 is the introduction and elucidates the motivation and research objectives of this thesis.

Chapter 2 provides an overview of the PLR and the research work performed in the literature. It also provides a review of the methodologies used for PLR estimation as well as the selected method for PLR estimation, that is used in Chapter 3 and Chapter 4.

Chapter 3 describes the methodology and background theory. This chapter is divided into six subsections. The first and second subsections include a description of the experimental setup and the metric used for assessing the PV performance. The third and fourth subsection give the background theory of the SARIMA model and RPCA, which are the foundation methodologies in which the PLR estimation was based. The fifth subsection describes how the PLR values were calculated from the forecasted PR values from the SARIMA models after the implementation of the RPCA analysis. This chapter also includes the metrics in which the model adequacy assessment and the description of the comparative analysis used for the final selection of the optimal method.

Chapter 4 elucidates the PLR results obtained using eight years of monthly PR data for eleven PV systems. The final selection of the optimal method is also included in this chapter.

Finally, in the last chapter, Chapter 5, the conclusions and future work are presented.

# *Chapter 2 – Overview of the Performance Loss Rate*

## 2.1 Performance Loss Rate Definition

The accurate estimation of power production over time is of vital importance for the further growth of the PV sector. A precise quantification of the decreasing trend of power over time, also known as PLR, is crucial to all the PV relevant stakeholders, such as utility companies, integrators, investors and researchers alike [3], [6].

The issue of performance loss of PV has lead researchers across the world to investigate the underlying degradation mechanism achieving this way increment of the modules production and efficiency, preventing failures, as well as the introduction of new materials [4]. PV systems, especially the field operating ones, experience performance losses at all levels, i.e. cell, module, array and system. Different environmental factors and degradation mechanisms can cause such losses, including temperature, module soiling, humidity, snow, precipitation and solar irradiation, and to parameters relating to their constituent instruments.

At the cell level, the main mechanisms underlying degradation losses are corrosion, light-induced degradation, contact stability and cracks [6]. Additional to these, mismatches between modules, shading, glass breakage, busbar failure, diode failures, delamination, broken interconnects and hot-spots are some of the factors which can cause degradation to the module, array and generally the whole system [4], [6]. These factors induce degradation mechanisms causing significant stress over the lifetime of a PV module, resulting in the reduction of durability, which must be quantified through the estimation of the PLR [9]. For the quantification of long-term behaviour and lifetime of PV systems, the outdoor field testing is of crucial meaning, as is their designated operating environment and is realistic way to relate indoor accelerated testing to outdoor results with the aim to accurate forecast field performance.

The effect of the different degradation mechanisms on PVs depends on the technology, the operating topology, their cumulative history of field exposure and of course the location of installation [10]. Resulting, to different performance loss trends of the nominal power of different PV systems. This performance loss trend is expressed as PLR and defines the rate of nominal performance drop over time. Usually is expressed in $\%/yr$ and represents the reduction of the chosen performance metric in the field [11].

The main parameters which were related to the PLR of crystalline silicon (c-Si) technology were decreasing of the short-circuit current, $I_{SC}$ and fill factor, $FF$ [12], [13]. The $I_{SC}$ initial degradation was attributed to oxygen contamination in the bulk of the Si junction, whereas the slow long-term degradation correlated linearly with ultraviolet exposure [13]. A comprehensive review [3] published by National Renewable Energy Laboratory (NREL) reported that no statistical difference was found between mono-Si and multi-Si technologies. However, modules connected to inverter experience higher PLR than the ones under open-circuit conditions, due to thermomechanical fatigue of the interconnects. Additionally, the type of materials used in modules showed to have an influence on the PLR, with the EVA and polyvinyl butyral encapsulants having higher PLR than silicone encapsulant. Also, glass–glass modules exhibited larger PLR than glass–polymer modules. Concluding, their review study reported that high PLR were attributed to high losses in $FF$, i.e., significant increases in series resistance, $R_S$, while moderate PLR were

due to optical losses in $I_{SC}$. Sánchez-Friera et al. in their study [14] ascribed the large PLR, of almost $1\ \%/yr$, to the antireflective coating, in addition to front delamination and inherent junction degradation.

Jordan et al. in [15] aggregated and analysed more than 11000 annual PLRs of almost 200 studies from 40 different countries. Through their study they found the median and mean PLR for c-Si modules to be in the range $0.5-0.6\ \%/yr$ and $0.8-0.9\ \%/yr$, respectively. PV modules of thin-film technologies exhibited additional degradation mechanisms [16] and in general higher PLR of the $FF$ in comparison to the c-Si ones [17]. A total of 455 annual PLRs from studies of thin-film technologies were used in [15], resulting to median and mean value of the PLR greater than $1\ \%/yr$ and almost equal to $1.4\ \%/yr$, respectively. The overall mean PLR including all the studies was calculated at $0.91\ \%/yr$, with the thin-film technologies variating from $1\ \%/yr$ to $6\ \%/yr$, in contrast with c-Si technologies which were mainly concentrated around the mean.

To conclude, this noticeable variation of the estimation of PLR for the different technologies of modules, is attributed to a number of different factors. Such factors are the technology of the module under study, the operating topology, the different climate/operating conditions [3] their cumulative history of field exposure [10] and the methodology used for the PLR evaluation [4].

## 2.2 Methodologies for the estimation of the PLR

In literature estimation methods for PLR vary depending on the selected statistical method, utilised performance metric, data filtering technique and if the test conditions are under field operation or indoors at STC [3]. The most commonly used statistical methods for calculating the PLR includes the Linear Regression (LR), Year-over-Year (YoY) methodology, Classical Seasonal Decomposition (CSD), Auto-Regressive Integrated Moving Average (ARIMA) model and the LOcally wEighted Scatterplot Smoothing (LOESS) [4]. The implementation such methods requires the use of a performance metric, such as electrical parameters from IV curves, regression models as the Photovoltaics for Utility Scale Applications (PVUSA) or normalized ratings as the PR.

### 2.2.1 Test conditions

Performance metrics can depend on measurements acquired under field operation or indoors at STC. Testing at STC requires the use of solar simulators making this approach more time consuming and efficient only for application on small-scale PV plants [18]. Usually, for larger plants only a small sample of PV modules is tested indoors and used as an indication for the PLR of the whole PV plant. Thus, the PLR estimation using data from field operation are more accurate and representative on the actual loss of power which will affect the levelized cost of energy (LCoE). For these reasons, a multi-disciplinary methodology must be used which will overcome all the issues arising with fielded data.

To include the failures and degradation mechanisms occurred in a PV module/array/system and estimate the PLR an analysis of the performance of PV in the fields and the prevailing

meteorological conditions are needed. Typical parameters include acquired electrical measurements, such as the array current, $I_A$, voltage, $V_A$, and power $P_A$, power to the utility grid, $P_{TU}$ (for grid-connected systems), meteorological measurements, such as the global irradiance in plane of array (POA), $G_I$, ambient air temperature, $T_{amb}$, module temperature, $T_{mod}$, wind speed $W_S$, and relative humidity, $RH$ and additional parameters which are extracted from continuous current-voltage (IV) characterization modules and arrays in the field, such as the short-circuit current, $I_{SC}$, open circuit voltage $V_{OC}$, fill factor $FF$, series resistance $R_S$, and shunt resistance, $R_{SH}$ [19]. These parameters are then used to generate time series of performance metrics.

### 2.2.2 Performance metric

The most common used performance metrics are the electrical parameters from IV curves which recorded under outdoor or controlled indoor conditions and corrected to STC, the power extrapolated to Photovoltaics for Utility-Scale Applications (PVUSA) test conditions ($P_{PTC}$) [20], and the performance ratio (PR).

IV curves can be generated manually outdoors at fixed intervals or indoors at STC at sparse intervals with the modules operating at MPP between IV scans [21]. The PLR can be detected based on the electrical parameters of the IV curve, which can be indicative for the degradation mechanisms affecting the module. However, outdoor IV characterization is mostly used for research [12] and diagnostic purposes [22]. On the other hand, indoor IV characterization is less commonly used as it is time consuming and inefficient for field operating PVs. On top of that, damaging of the modules might occur during indoor IV characterization due to mishandling, dismounting and transportation.

To extrapolate field measurements, regression models are used, which exploit the linear relationship between the meteorological and PV operational parameters. A such model is the PVUSA [23]–[25]. The model assumes that $I_A$ is proportional to $G_I$ and that $V_A$ is proportional to $T_{mod}$, which also depend on the $G_I$, $T_{amb}$ and $W_S$. For its implementation filtering of the $G_I$ measurements are needed, i.e. use only the high values ($G_I \geq 800\, W/m^2$) and the use of training dataset for $P_A$ or $P_{TU}$, $G_I$, $T_{amb}$ and $W_S$ in order to identify the coefficients $c_1$, $c_2$, $c_3$ and $c_4$ using Eq. (2.1). The coefficients of the model are calculated monthly and then the PVUSA test conditions (PTC) ($G_I = 1000\, W/m^2$, $T_{amb} = 20$ ℃ and $W_S = 1\, m/s^2$) are used to estimate the $P_{MPP}$.

$$P_{MPP} = G_I(c_1 + c_2 \cdot G_I + c_3 \cdot T_{amb} + c_4 \cdot W_S) \quad (2.1)$$

The above model is inaccurate for thin-film technologies. For this reason a modification was proposed in [26] which includes an additional regression coefficient $c_5$, representing a loss factor (Eq. (2.2)). This modified model expresses the array yield, $Y_A$, rather than the $P_{MPP}$ and uses measurements for $G_I \geq 50\, W/m^2$.

$$Y_A = G_I(c_1 + c_2 \cdot G_I + c_3 \cdot T_{amb} + c_4 \cdot W_S) - c_5 \quad (2.2)$$

One of the most widely used, both in industry and research, performance metrics is the PR. The PR is a key performance indicator (KPI) that provides the ability to compare the performance of different PV technologies, with different capacities and geographical locations [27]. The PR can be expressed as the ratio of the $Y_A$ of the PV array (or the final yield, $Y_f$ of the PV plant) and the reference yield ($Y_r$) [19]. Usually the PR is calculated monthly or annually, with the annual ratings providing quick insights regarding the permanent performance loss of the PV. However, weekly and daily values of the PR are useful for identifying failure and measurement outliers. The monthly values of the PR, on the other hand, are a good indication of soiling losses and seasonality in comparison with the PVUSA method. Jordan et al. in [28] have reported that the variability of the PLR estimation was increased with the use of smaller time intervals of these metrics. In general, PR due to its calculation formula achieves normalization with irradiance, which can reflect in the metric values the overall effect of losses, such as soiling, reflection losses, system shutdown and component failures.

In literature several studies have compared the PLR using PVUSA or PR as the performance metric. Marion et al. in [29] found similar results for both metrics using different PV technologies, however other studies [28], [30] found considerable difference between them. Additionally, the use of temperature corrected PR as formulated in [31] showed higher PLR than when using the regular PR [28]. This can be attributed to the non-linear behaviour of the instant temperature coefficient in field operation [28].

The performance assessment in the present study was based on monthly PR time series, as this metric provides the ability to compare the performance of different PV technologies, with different capacities and geographical locations, and its value reflects the overall effect of losses which is needed for the accurate evaluation of the PLR.

### 2.2.3 Time series analysis

Phinikarides et al. in [32] showed that the PLR estimation is highly depended on the utilised statistical method. Generally, the different statistical analysis methods are used in order to identify the behaviour / trend of the PV performance over time and then evaluate the rate of performance change, i.e. the annual PLR. The main methods used in literature can be divided in model-based and non-parametric-methods. Model-based methods require the identification of a stochastic time series model and include the LR, CSD and ARIMA. On the other hand, for non-parametric methods there is not needed to define a model and they are popular due to their robustness and simplicity, such methods include the LOESS and YoY.

The LR is one of the easiest to apply methods and that made it the most commonly used method for estimating PLR in literature. The LR method fits the $\hat{y}$ values in Eq. (2.3) to describe the PV performance metric by identifying the slope of the trend, $c_1$ and the intercept, $c_2$ using ordinary least-squares (OLS) [33]. However, LR method appears to have high uncertainty due to its high sensitivity to outliers and seasonal variations [32].

$$\hat{y} = c_1 t + c_2 \tag{2.3}$$

The YoY methodology was proposed by the SunPower for estimating the PLR [34] and is a comparative data point method, which can be used as an alternative to regressive models. The YoY calculates the rate of change between two points at the same time in subsequent years [35]. The data points can be months, weeks or days of the performance metric, and the rate of change is calculated for each point individually. Then the YoY method calculates the median value of the rate of change over 1-year period, which represents the PLR of the PV system for that year.

Seasonal fluctuations due to weather changes closely affect the value of the PR. In order to consider the seasonal behaviour of the PR and accurate estimate the PLR seasonalized techniques such as CSD and ARIMA are needed.

CSD is a simple method [36] used to estimate their PLR of PV systems [4], [5], [33]. The CSD can be expressed as an additive model (Eq. (2.4)) or as a multiplicative model (Eq. (2.5)) depending on the seasonal component.

$$\hat{y} = T_t + S_t + e_t \tag{2.4}$$

$$\hat{y} = T_t \cdot S_t \cdot e_t \tag{2.5}$$

where $\hat{y}$ are the fitted values, $T_t$ the trend, $S_t$ the seasonal component and $e_t$ the residual component. However, the CSD assumes a stable $S_t$ for all years and does not make any amendments to the model based on the residuals. These can cause correlated residuals in the model, which comes to a confrontation with the most basic assumption of uncorrelated residuals in stochastic models. Lastly, the assumption of stable $S_t$ does not reflect at all the behaviour of a-Si technologies which are experiencing periods of lower and higher efficiency due to the Staebler–Wronski effect and thermal annealing, respectively.

ARIMA method, on the other hand, and specifically the seasonal ARIMA (SARIMA) is more adaptive in comparison with CSD. SARIMA model [37] due to its nature can effectively handle seasonal behaviour, random errors, outliers and level shift as it is eliminating all the correlated residuals of the model. The linear SARIMA was used in the present study to capture the behaviour of the PR and forecast future values, more details about the model are given in Section 3.3.

# *Chapter 3 – Background Theory and Methodology*

Figure 3.1 illustrates a flowchart of the overall methodology followed for the identification of the optimal method to estimate the PLR.

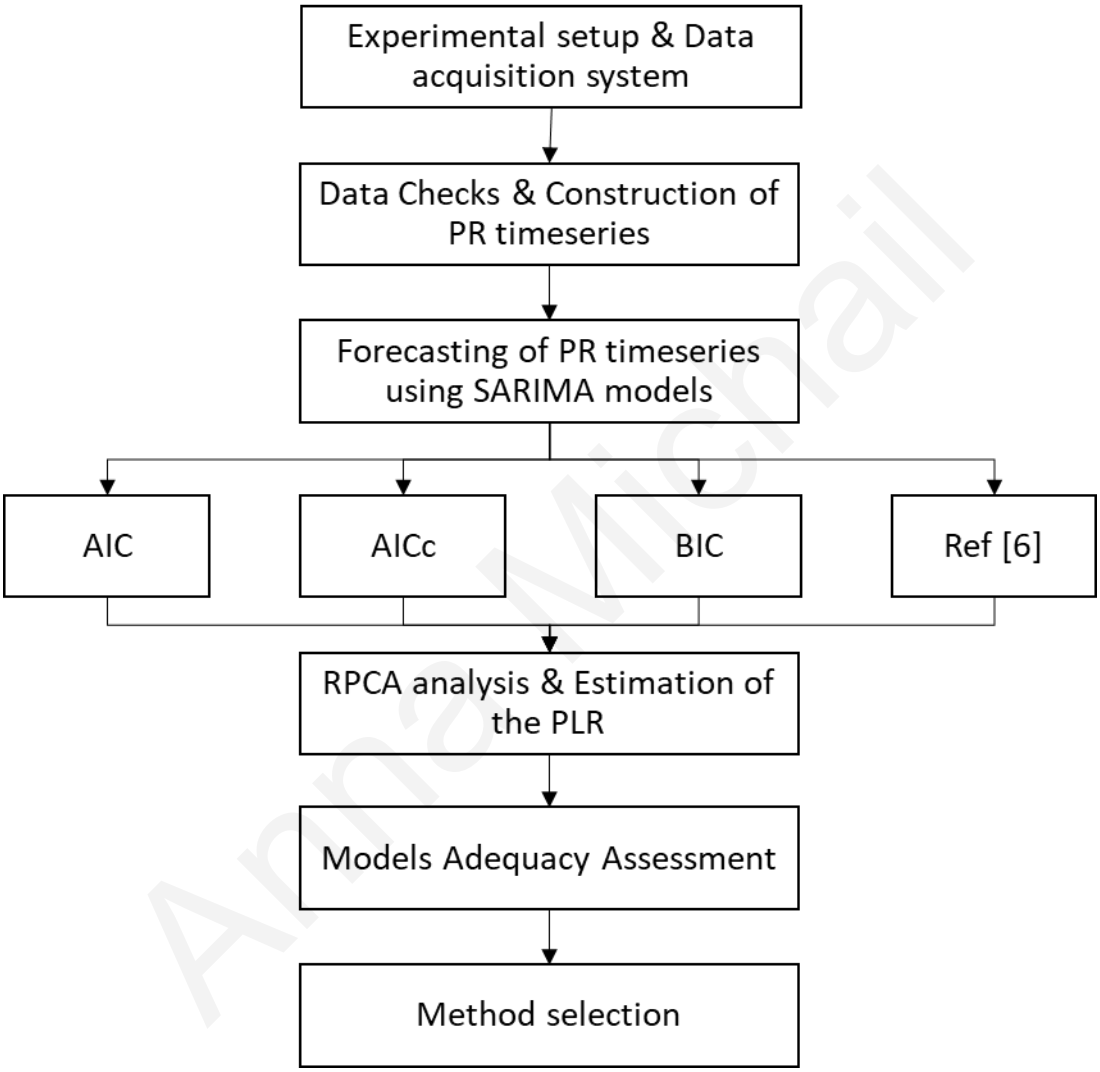


**Figure 3.1. Overall methodology of the present study.**

## 3.1 Experimental Setup and Data Acquisition System

At the outdoor test facility of the University of Cyprus (UCY) (see Figure 3.2), grid-connected PV systems of different technologies (mono-c-Si, multi-c-Si and thin films) and approximately 1 $kW_p$ capacity each, were installed in May 2006. The PV systems were installed side-by-side in an open-field arrangement due to South at the optimum annual energy angle of 27.5°. The specifications of the installed PV modules are presented in Table 3.1 [38], [39].

**Figure 3.2. Outdoor test facility of the UCY in Nicosia, Cyprus.**

**Table 3.1. Manufacturer datasheet specifications of installed PV modules and rated power, $P_0$, of the PV system.**

| a/a | Manufacturer | Module type | Technology | $\eta_{STC}$ (%) | $V_{OC}$ $(V)$ | $I_{SC}$ $(A)$ | $I_{MPP}$ $(A)$ | $V_{MPP}$ $(V)$ | $P_{MPP}$ $(W)$ | $A$ $(m^2)$ | $P_0$ $(kW_p)$ | $\gamma_{manuf}$ $\left(\frac{\%}{K}\right)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 01 | Solon | P220/6þ | Multi-c-Si | 13.40 | 36.50 | 8.25 | 7.62 | 28.90 | 220.00 | 1.64 | 1.540 | -0.430 |
| 02 | Sanyo | HIP-205NHE1 | Mono-c-Si (HIT-cell) | 16.40 | 50.30 | 5.54 | 5.05 | 40.70 | 205.00 | 1.25 | 1.025 | -0.300 |
| 03 | Atersa | A-170M 24V | Mono-c-Si | 12.90 | 44.00 | 5.10 | 4.75 | 35.80 | 170.00 | 1.32 | 1.020 | -0.370 |
| 04 | Suntechnics | STM 200 FW | Mono-c-Si (back-contact cell) | 16.10 | 47.80 | 5.40 | 5.00 | 40.00 | 200.00 | 1.24 | 1.000 | -0.380 |
| 05 | Schott Solar | ASE-260-DG-FT | Multi-c-Si (EFG) | 11.70 | 70.90 | 4.91 | 4.55 | 57.10 | 250.00 | 2.14 | 1.000 | -0.470 |
| 06 | BP Solar | BP7185S | Mono-c-Si (11mplan-cell) | 14.80 | 44.80 | 5.50 | 5.10 | 36.50 | 185.00 | 1.25 | 1.110 | -0.500 |
| 07 | SolarWorld | SW165 poly | Multi-c-Si | 12.70 | 43.90 | 5.10 | 4.60 | 35.50 | 165.00 | 1.30 | 0.990 | -0.470 |
| 08 | Schott Solar | ASE-165-GT-FT/MC | Multi-c-Si (MAIN-cell) | 13.00 | 44.00 | 5.25 | 4.71 | 36.00 | 170.00 | 1.31 | 1.020 | -0.470 |
| 09 | Würth Solar | WS 11007/75 | CIGS | 10.30 | 45.50 | 2.50 | 2.22 | 36.00 | 75.00 | 0.73 | 0.900 | -0.360 |
| 10 | First Solar | FS60 | CdTe | 8.30 | 90.00 | 1.14 | 0.94 | 64.00 | 60.00 | 0.72 | 1.080 | -0.250 |
| 11 | MHI | MA100T2 | a-Si (single cell) | 6.40 | 141.00 | 1.17 | 0.93 | 108.00 | 100.00 | 1.57 | 1.000 | -0.200 |

During the 8 years evaluation period (from June 2006 to May 2014), the performance of the Solon multi-c-Si and BP solar mono-c-Si was affected by partial shading during the second, third and fourth years [38], [39].

The performance of each PV system and the prevailing meteorological conditions were recorded according to the IEC 61724 [19], at a resolution of 1 second and stored as 1- and 15-min averages with the use of a monitoring platform. The acquired meteorological data include the in-plane irradiance $G_I$ from pyranometers, ambient temperature $T_{amb}$, back surface module temperature $T_{mod}$, wind speed $W_S$, and direction $W_a$, whereas the electrical data include the array dc current $I_A$, voltage $V_A$ and power $P_A$, and ac power to the utility grid $P_{TU}$.

The monitoring platform consists of different sensors summarised in Table 3.2. Periodic calibrations and inspections of the sensors were performed to ensure top quality measurements and to account for any sensor drifts. Additionally, all sensor cabling and connection terminals were periodically checked for moisture intrusion, damage, and loose connections [38], [39].

**Table 3.2. Data acquisition equipment and sensors.**

| Instrument | Manufacturer | Model |
|---|---|---|
| Data logger | Delphin | Topmessage |
| Ambient temperature | Theodor Friedrichs | 2030 |
| Module temperature | Heraeus | PT100 |
| Total irradiance | Kipp & Zonen | CM 21 e CV 2 |
| DC voltage | Custom made | Voltage divider |
| DC current | Custom made | Shunt resistor |
| AC energy | NZR | AAD1D5F |
| Wind speed | Theodor Friedrichs | 4034 |
| Wind direction | Theodor Friedrichs | 4122 |

## 3.2 PV Performance Analysis

The estimated PLR can be affected considerable by the amount of corrupted and missing data in the performance metric used [5]. For this reason, the measurements of the $P_A$ and $G_I$ were prior checked for invalid data and outliers as described in [40]. Specifically, for the correction of downtimes of the system and sensor, set thresholds and maintenance logs were used. Additionally, downtimes of short outage periods (less than a day), were corrected using past measurements. To avoid introducing bias to the datasets, longer outage periods were not corrected. The obtained datasets were considered the complete datasets of observed values, which were used to construct the monthly PR time series using the following formula [19]:

$$PR = \frac{Y_A}{Y_r} = \frac{{E_A}/{P_0}}{{H_I}/{G_{I_{STC}}}} \tag{3.1}$$

where $Y_A$ is the monthly array energy yield defined as the PV array output energy, ($E_A$), per rated kW of installed PV array ($P_0$), and $Y_r$ is the reference yield which can be calculated by dividing the total in-plane irradiation ($H_I$) by the module's reference plane of array irradiance ($G_{I_{STC}}$).

In this work, the monthly PR was constructed (see Figure 3.3) using the 15-min average outdoor data acquired over a period of 8 years from the beginning of their operation.

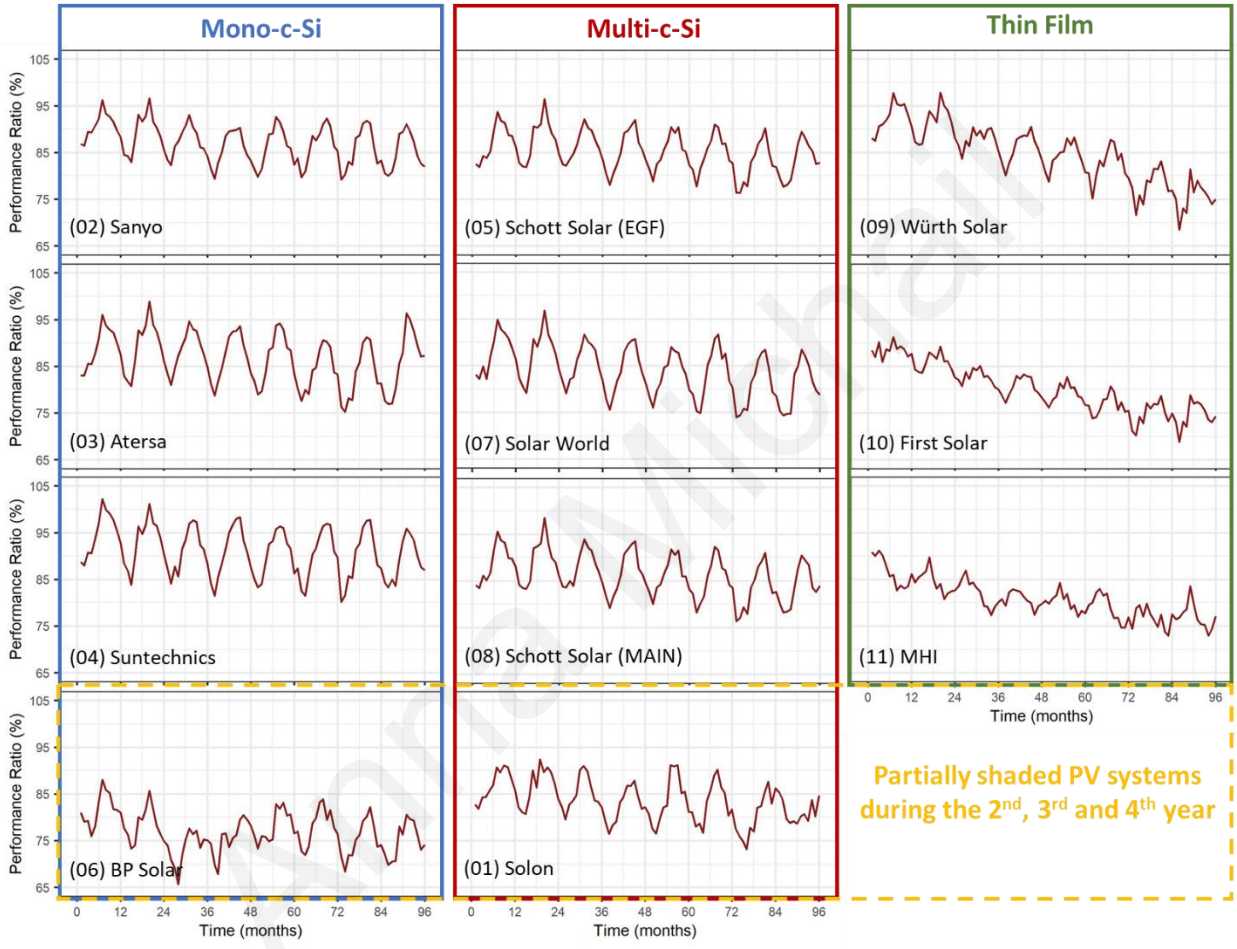


**Figure 3.3. PR time series the eleven PV systems categorized by the three main technologies, mono-c-Si, multi-c-Si and thin film with the additional sub- category for the two partially shaded PV systems. The numbering corresponds to the numbering used in Table 3.1.**

Figure 3.4 shows the PR time series of the Atersa mono-c-Si, Schott Solar (MAIN) multi-c-Si and First Solar CdTe thin film system per year. It can be observed that the PR time series exhibit seasonality and its value decreases over the years. More specifically, the seasonality is evident in Figure 3.3 and Figure 3.4 as a periodicity every 12 months and a repetitive behaviour of the graphs over the years, respectively. The decreasing of the PR value over the years is revealed as a downward trend of the mean PR value over the years. This change of the mean value over

time in statistical analysis can be expressed as non-stationary of the statistical behaviour of the time series. The optimal statistical model to describe times series that exhibit both seasonality and non-stationarity is the SARIMA model.

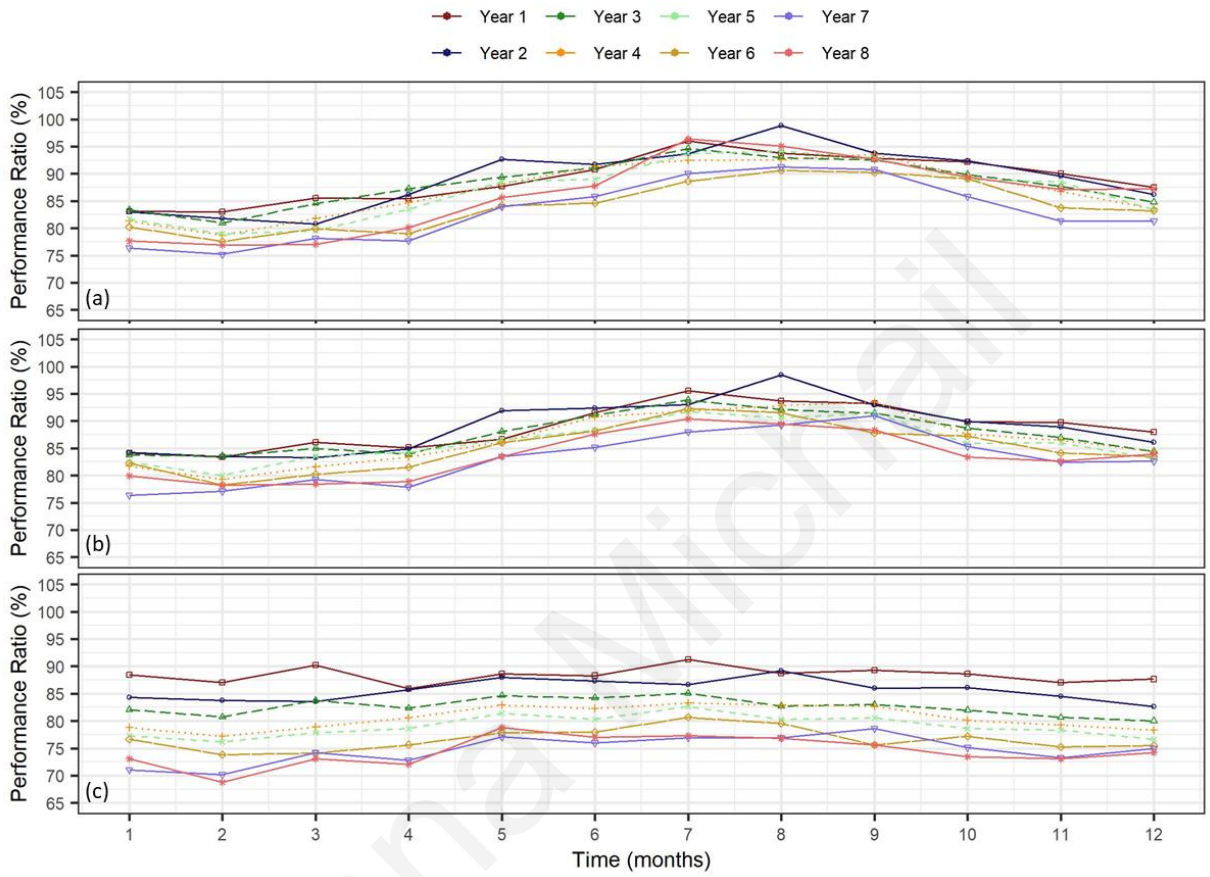


**Figure 3.4. PR time series divided per year of the (a) Atersa mono-c-Si (b) Schott Solar (MAIN) multi-c-Si and (c) First Solar CdTe thin film system.**

Makrides et al. in [38] showed that a minimum period of 3 to 5 years is required for accurate PLR estimation. Especially, for the c-Si technologies the PLR converges to a steady state value after 5 years, whereas more time might be required for thin-film technologies [38]. For these reasons, the first 5 years of the monthly PR time series were used as the train dataset to identify and estimate the SARIMA models for each PV system. Monthly PR forecasts were then generated for the next 3 years using the estimated SARIMA models. The 3-year forecast time series were then combined with the 5-year training dataset to create an 8-year time series of monthly PR for each PV system, later mentioned as forecasted data.

The robust principle component analysis (RPCA) proposed in [41] was then used to obtain the PLR at the end of the 6$^{th}$, 7$^{th}$ and 8$^{th}$ year for the actual and forecasted data of each PV system. To implement the RPCA methodology, the forecasted and actual data both consisted by 96

monthly PR time series were divided into eight 12-month time series. Then the eight 12-month time series were collected in a matrix $\boldsymbol{D}$ of dimension 8x12, called data matrix, for each PV system to be analysed using the RPCA.

## 3.3 Seasonal Auto-Regressive Integrating Moving Average model (SARIMA)

SARIMA is composed by three different basic time series operators, namely the autoregressive, the moving average and the integrating operators.

### 3.3.1 Auto-Regressive model

An auto-regressive (AR) model is a representation of a type of random process; as such, it is used to describe certain time-varying processes in nature, economics, etc. The auto-regressive model specifies that the output variable depends linearly on its own previous values and on a stochastic term (Eq. (3.2)). The AR model is not always stationary, as it may contain unit roots (generalized AR operator).

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-1} + a_t \tag{3.2}$$

where $\varphi_i, \dots, \varphi_p$ are the parameters of the model and $a_t$ is a white noise.

For the time series analysis the backshift operator ($B$) must be introduced, which operates on an element of a time series to produce the previous element as given from the following equation

$$BX_t = X_{t-1} \tag{3.3}$$

where $X_t$ current and $X_{t-1}$ previous element of the time series. Additionally, the first difference operator ∇ (Eq. (3.4)) will be present in the following models and is operating on $X_t$ as follows

$$\nabla X_t = X_t - X_{t-1} \leftrightarrow \nabla X_t = (1 - B)X_t \tag{3.4}$$

The second difference operator $\nabla^2$ is given by

$$\begin{gathered} \nabla(\nabla X_t) = \nabla X_t - \nabla X_{t-1} \leftrightarrow \nabla^2 X_t = (1 - B)\nabla X_t \\ \leftrightarrow \nabla^2 X_t = (1 - B)(1 - B)X_t \leftrightarrow \nabla^2 X_t = (1 - B)^2 X_t \end{gathered} \tag{3.5}$$

Following the above approach, $\nabla^i$ is given as

$$\nabla^{\mathrm{i}} X_t = (1 - B)^{\mathrm{i}} X_t \tag{3.6}$$

Then the Eq. (3.2) can be rewritten in terms of the backshift operator $B$ as

$$X_t = \sum_{i=1}^{p} \varphi_i B^i X_t + a_t \tag{3.7}$$

An auto-regressive model can thus be represented as a superposition of its own previous values whose input is a stochastic term (white noise).

$$\varphi(B) X_t = a_t \tag{3.8}$$

where,

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p \tag{3.9}$$

is the $p^{\mathrm{th}}$ order polynomial in the operator $B$, known as the auto-regressive operator.

### 3.3.2 Moving Average model

In time series analysis, the moving-average (MA) model, also known as moving-average process, is a common approach for modelling univariate time series. The moving-average model specifies that the output variable depends linearly on the current and past values of a stochastic term,

$$X_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{3.10}$$

where the $\theta_1, \ldots, \theta_q$ are the parameters of the model and the $a_t, a_{t-1}, \ldots, a_{t-q}$ are white noise error terms. The value of $q$ is called the order of the MA model. This can be equivalently written in terms of the backshift operator $B$ as

$$X_t = \left(1 - \theta_1 B - \cdots - \theta_q B^q\right) a_t \tag{3.11}$$

and in an operator form

$$X_t = \theta(B) a_t \tag{3.12}$$

Thus, a moving-average model is conceptually a linear regression of the current value of the series against current and previous (observed) white noise error terms or random shocks. The random shocks at each point are assumed to be mutually independent and to come from the same distribution, typically a normal distribution.

$$\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q \tag{3.13}$$

where the $q^{\mathrm{th}}$ order polynomial, is known as the moving average operator.

### 3.3.3 ARMA model

To achieve greater flexibility in fitting of actual time series, it is sometimes advantageous to include both AR and MA terms in the model. This leads to the mixed autoregressive-moving average (ARMA) model [37]:

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{3.14}$$

or

$$\varphi(B) X_t = \theta(B) a_t \tag{3.15}$$

The model employs $p + q + 2$ unknown parameters $\mu, \varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q, \sigma_a^2$, that are estimated from the data, where $\mu$ expresses the mean and $\sigma_a^2$ the variance of the model. In practice, it is frequently true that an adequate representation of actually occurring stationary time series can be obtained with autoregressive, moving average, or mixed models, in which $p$ and $q$ are not greater than 2 and often less than 2.

### 3.3.4 Integrating operator and the ARIMA model

Many time series encountered in industry or business exhibit nonstationary behaviour and in particular do not vary about a fixed mean. Such series may nevertheless exhibit homogeneous behaviour over time of a kind, e.g. the behaviour of the PR. In particular, although the general level about which fluctuations are occurring may be different at different times, the broad behaviour of the series, when differences in level are allowed for, may be similar over time. Such behaviour may often be represented by a model in terms of a generalized autoregressive operator $\psi(B)$, in which one or more of the zeros of the polynomial $\psi(B)$ (i.e., one or more of the roots of the equation $\varphi(B) = 0$) lie on the unit circle. If there are $d$ unit roots and all other roots lie outside the unit circle, the operator $\psi(B)$ can be written

$$\psi(B) = \varphi(B)(1 - B)^d \tag{3.16}$$

where $\varphi(B)$ is a stationary autoregressive operator. Thus, a model that can represent homogeneous nonstationary behaviour is of the form

$$\psi(B) X_t = \varphi(B)(1 - B)^d X_t = \theta(B) a_t \tag{3.17}$$

that is,

$$\varphi(B) w_t = \theta(B) a_t \tag{3.18}$$

where

$$w_t = (1 - B)^d X_t = \nabla^d X_t \tag{3.19}$$

Thus, homogeneous nonstationary behaviour can sometimes be represented by a model that calls for the $d^{\text{th}}$ difference of the process to be stationary. In practice, $d$ is usually 0, 1, or at most 2, with $d = 0$ corresponding to stationary behavior.

The process defined by Eq. (3.18) and (3.19) provides a powerful model for describing stationary and nonstationary time series and is called an autoregressive integrated moving average process, of order ($p$, $d$, $q$), or ARIMA($p$, $d$, $q$) process. The process is defined by

$$w_t = \varphi_1 w_{t-1} + \cdots + \varphi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_p a_{t-p} \tag{3.20}$$

The relationship, which is the inverse to (3.19), is $X_t = \boldsymbol{S}^d w_t$, where $\boldsymbol{S}^d = \nabla^d = (1-B)^{-d} = 1 + B^{-1} + \cdots + B^{-d}$ is the summation (or integration) operator defined by

$$\boldsymbol{S} w_t = \sum_{j=0}^{\infty} w_{t-j} = w_t + w_{t-1} + w_{t-2} + \cdots \tag{3.21}$$

Thus, the general ARIMA process may be generated by summing or ‘‘integrating’’ the non-stationary process $w_t d$ times. A special form of the model in Eq. (3.20) can be employed to represent seasonal time series, known as the SARIMA model.

### 3.3.5 SARIMA model

The SARIMA model is a multiplicative model, which models seasonal data $X_t$ of period $s$ by combining the operators from Eq. (3.9), (3.13) and (3.19) in the following manner:

$$\varphi_p(B)\Phi_P(B^s)\nabla^d \nabla_s^D X_t = \theta_q(B)\Theta_Q(B^s)a_t \tag{3.22}$$

where $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ are the seasonal AR and MA polynomials in $B^s$ of orders $P$ and $Q$ respectively and $\nabla_s^D = (1-B^s)^D$ is the seasonal differencing operator. Th model of Eq. (3.22) is usually labelled as $(p,d,q)(P,D,Q)_s$. The orders $p, d, q, P, D$ and $Q$ and the coefficients of the polynomials $\varphi_p(B), \Phi_P(B^s), \theta_q(B)$ and $\Theta_Q(B^s)$ are estimated from the available data.

The model of Eq. (3.22) can provide crucial information of the time, statistical, and spectral behaviour of the time series, $X_t$. This information can be exploited to forecast the future values of the time series $X_t$. It can be shown, that the forecast value, $\hat{X}_t(h)$, at an instant of $h$ time steps ahead from a time instant $t$ is the conditional expectation,

$$\hat{X}_t(h) = E\left[X_{t+h} \middle| \theta_q, \varphi_p, \Theta_Q, \Phi_P, X_t, X_{t-1}\right] \tag{3.23}$$

where $X_{t+h}$ is the random variable generated by the SARIMA model (3.22) at the future time instant $t+h$ and $\theta, \varphi, \theta_s, \varphi_s$ are the coefficients of the polynomials $\theta(B)$, $\varphi(B)$, $\theta_s(B^s)$, $\varphi_s(B^s)$ respectively, and $X_t, X_{t-1}, \ldots$ are actual PR values up to the time $t$.

The identified SARIMA model for each PV system using the initial 5 years training dataset, were utilised to forecast the monthly PR values for the last 3 years. These models are linear, thus it can be said that the estimation of the PLR is based on a linear assumption.

## 3.4 Robust Principal Component Analysis (RPCA)

The quality of outdoor measurements can be affected unpredictably by the environment, operating conditions, and the uncertainty of the sensors. Even, thought the measurements were cleaned from outliers and missing data before the construction of the PR time series the aforementioned effects can still affect the data and consequently the estimation of the PLR. Kyprianou et al. in [41] have proposed the use of the robust principal component analysis (RPCA) in order to further eliminate these effects.

Every actual and forecasted dataset of each of the PV systems was used to create a rectangular $(m, n)$ data matrix $\boldsymbol{D}$, where $m$ and $n$ indicate the 8 years and the 12 months of each year, respectively. Then an RPCA analysis on $\boldsymbol{D}$ is carried out, where the data matrix $\boldsymbol{D}$ is expressed as:

$$\boldsymbol{D} = \boldsymbol{K} + \boldsymbol{E} \tag{3.24}$$

where, $\boldsymbol{K}$ is a low rank data matrix and $\boldsymbol{E}$ an unknown sparse perturbation matrix. RPCA extracts $\boldsymbol{K}$ by solving the following optimization formula:

$$\begin{aligned} &minimize \|k\|_s + \mu \|E\|_1 \\ &\text{subjected to } \boldsymbol{D} = \boldsymbol{K} + \boldsymbol{E} \end{aligned} \tag{3.25}$$

where $\|k\|_s$ is a norm defined by the sum of the singular values of $\boldsymbol{K}$ and $\|E\|_1$ is a norm defined by the summation of the absolute values of $\boldsymbol{E}$ and is a weighting parameter that trades off between the two norms. The RPCA program was solved using the function “rrpca” of the package “rsvd” in R Project for Statistical Computing [42], which decomposes the matrix using the method of inexact Augmented Lagrange Multiplier (iALM) as given in Eq. (3.25) [43].

## 3.5 Definition of Performance Loss Rate

The PLR of the PV systems was estimated using the RPCA actual and forecasted data. The PLR for the $i^{th}$ year was calculate using the Eq. (3.26) for both RPCA actual and forecasted data. The first year of the PR of each PV system was used as a reference for the PLR estimation, as the highest PLR of the lifetime of PV usually happens during this year.

$$PLR_i = \frac{A_{1-i}/A_1}{i} \tag{3.26}$$

where $A_1 = \int_1^{12}(PR_1)dt$ is area beneath the curve of the 1st year of the RPCA data and $A_{1-i} = \int_1^{12}(PR_1 - PR_i)dt$ is the area between the curve of the 1st and $i^{th}$ year. In Figure 3.5 the shaded region of each graph represents the $A_{1-8}$, which indicates the 8-year PR decay of three different PV system for the RPCA actual and forecasted data.

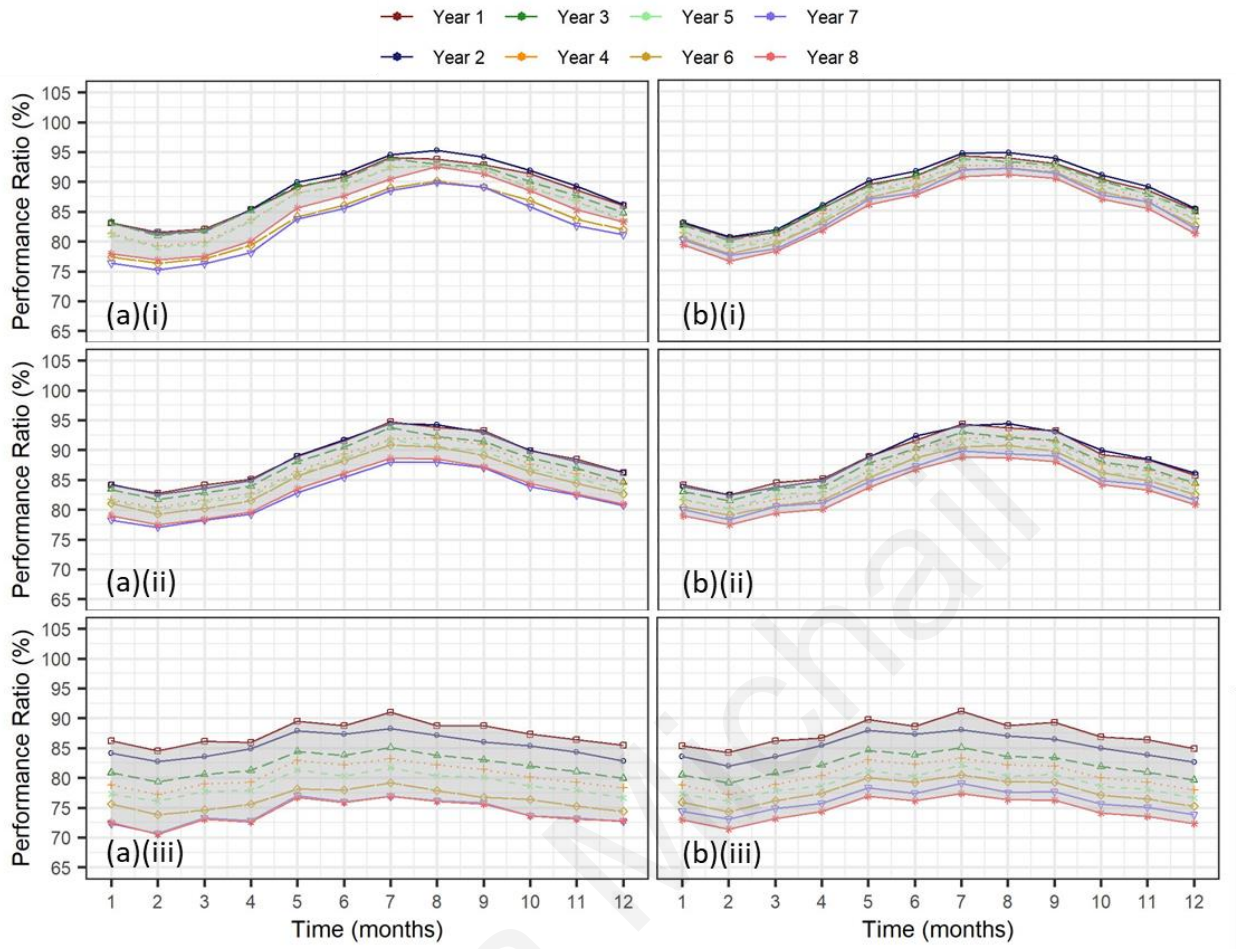


**Figure 3.5. The monthly PR observations divided by year of the (a) actual and (b) forecasted RPCA data for the (i) Atersa mono-c-Si (ii) Schott Solar (MAIN) multi-c-Si and (iii) First Solar CdTe thin film system. Shaded region represents the 8-year PR decay of the data.**

## 3.6 Model Adequacy Assessment

### 3.6.1 Model selection based on Information Criteria

Forecast accuracy measures such as root mean squared error (RMSE) and mean absolute error (MAE) can be used for selecting a model for a given set of data, provided the errors are computed from data in a validation set and not from the same data as were used for model estimation (training set). However, there are often too few validation-samples errors to draw reliable conclusions [44]. Furthermore, the sample autocorrelation (ACF) and partial autocorrelation (PACF) functions are extremely useful in model identification, but sometimes there are cases involving mixed models where they can provide ambiguous results. This of course can be solved with the model being subjected to further examination, diagnostic checking, and modification, if necessary [37].

Nevertheless, the use of a penalized method based on the training-sample fit can solve the aforementioned issues easier. Model specification can be based on model selection criteria such as AIC, AICc and BIC. AIC represents Akaike's information criterion, AICc is a modification of the AIC for small sample size, which was proposed to minimize the possibilities to select an overfitting model, and lastly, BIC is the Bayesian information criterion due to Schwarz. These criteria are likelihood based and include under normality the determinant of the innovations covariance matrix that reflects the goodness of fit of the model. A second term is a function of the number of fitted parameters and penalizes models that are unnecessarily complex.

The formulas of the AIC, AICc and BIC information criteria are given by the Eq. (3.27), (3.28) and (3.29), respectively [37]:

$$AIC = \frac{-2\ln(L) + 2r}{n} \tag{3.27}$$

$$AICc = AIC + \frac{2r^2 + 2r}{n - r - 1} \tag{3.28}$$

$$BIC = \frac{-2\ln(L) + 2r\ln(n)}{n} \tag{3.29}$$

where $n$ is the number of the training-sample data, $r \approx p + q + P + Q$ is the number of estimated parameters in the model and $L$ is the maximized likelihood of the model fitted to the differenced data $\nabla^d \nabla_s^D X_t$.

It can be observed that AICc and BIC imposes a greater ''penalty factor'' for the number of estimated parameters than does AIC. These criteria can be used to compare models fitted using maximum likelihood and the model that gives the lowest value for a given criterion should be selected. Hence, since the AICc and BIC criteria imposes a greater penalty for the number of estimated model parameters than does AIC, the use of minimum AICc or BIC for model selection would always result in a chosen model whose number of parameters ($r$) is no greater than that chosen under AIC [37].

In the present study all three information criterions were used for the identification of the SARIMA models and were compared with each other and with the models proposed in [6] for the identification of the optimal methodology for estimating the PLR. The SARIMA models in [6] were derived based on the ACF and PACF and verified using statistical analysis of the model's residuals.

### 3.6.2 Residuals Behaviour – Portmanteau Lack-of-Fit test

After the SARIMA models are identified diagnostic checks are needed to be applied to the fitted models in order to assure the quality of fitting. The most used and appropriate checks are applied to the residuals of the fitted models, which can indicate necessary modifications to the selected model. The residuals ($\hat{a}_t$) in a time series model are the differences between the simulated results of the fitted model ($\hat{y}_t$) and the actual observations ($y_t$),

$$\hat{a}_t = y_t - \hat{y}_t \tag{3.30}$$

The main properties that the residuals of a good forecasting method should have are that they must be uncorrelated and have zero mean [45]. If the residuals are correlated, it indicates that there is information in the residuals which the model should use for computing forecasts. Also, non-zero mean of the residuals shows that the forecasts are biased, i.e. constant difference between the forecasted and actual values. Additional properties which are not as essential as the two above, suggests that the residuals have a constant variance and are normally distributed. For an adequate model, it can be shown that the following equation is valid [37]

$$\hat{a}_t = a_t + \boldsymbol{O}\left(\frac{1}{\sqrt{n}}\right) \tag{3.31}$$

where $a_t$ is white noise, $n$ is the number of the model's residuals (same as the number of training-sample data) and $\boldsymbol{O}(1/\sqrt{n})$ represents order of the $1/\sqrt{n}$. Eq. (3.31) shows that as the series length increases, the $\hat{a}_t$'s become close to the $a_t$'s.

Box et al. in [37] proposed two diagnostic checks for the model's residuals, the autocorrelation check and the cumulative periodogram of the residuals. Autocorrelation check refers to the visual check of the ACF graphs of the residuals, which should be performed for all the obtained SARIMA models. The approximate upper and lower bounds used for the standard error of a single autocorrelation is $\pm 1/\sqrt{n}$, if the residuals fall within these bounds then the model is assumed to be adequate. In order to quantify the autocorrelations of the residuals the second check was performed as well, which was based on the portmanteau lack-of-fit test proposed by Ljung and Box.

$$\tilde{Q} = n(n+2)\sum_{k=1}^{K}(n-k)^{-1}r_k^2(\hat{a}) \tag{3.32}$$

where $n$ the number of data used to fit the model, $r_k(\hat{a})$ autocorrelations of the residuals $\hat{a}$ and $K$ the number of the first $r_k^2(\hat{a})$ used. The null hypothesis $H_0$ is that the mean $E[\tilde{Q}]$ is approximately distributed as $\chi^2(K-p-q-P-Q)$, which indicates that the model is appropriate fitted. For the execution of the Eq. (3.33) the "Box.test" function was used from the package "stats" in R Project for Statistical Computing. The $p$-values of the portmanteau statistic $\tilde{Q}$ were examined, a lack of fitting was indicated if their values were at or near the significance level ($\alpha = 0.05$). The $p$-values represent the probability of obtaining the observed results, when the $H_0$ of a study question is true. As a result, higher values of the $p$-values indicate that the model approximates better the $H_0$, and contrary $p$-values lower than $\alpha$ indicates that the $H_0$ is rejected meaning that the fitted model does not follow a $\chi^2(K-p-q-P-Q)$ distribution.

However, it is urged that the portmanteau statistic should not be considered as the only diagnostic check for the fitted model, but a careful examination must be performed of the residuals and their individual autocorrelation coefficients. The Ljung-Box portmanteau lack-of-fit test was mostly used in the present thesis for quantification of the fitting ability of the developed models.

### 3.6.3 Performance Assessment Metrics

The performance accuracy of the developed models was assessed based on performance metrics. Such metrics commonly used in PV forecasting applications include the MAE which measures the difference between the measured and simulated data and the RMSE which describes the standard deviation of the prediction errors. Additionally, the error of difference (Dif) and absolute error of difference (Abs Dif) where used in the present study specially to quantify the difference between the actual (based on observed measurements) and forecasted (based on simulated values) of the PR and PLR values. The metrics used to analyse the performance of the developed performance models are based on the difference between the measured (actual) and simulated (forecasted) PR or PLR values and are calculated as follows:

$$\begin{aligned} Dif_i &= y_i - \hat{y}_i \\ Abs\ Dif_i &= |y_i - \hat{y}_i| \end{aligned} \tag{3.33}$$

$$MAE = \frac{1}{n} \times \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{3.34}$$

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{3.35}$$

where $y_i$ and $\hat{y}_i$ are the observed and forecasted value (for PR or PLR) respectively.

### 3.6.4 Selection of Optimal Method

The SARIMA models based on AIC, AICc, BIC and Ref [6] were compared to identify the optimal method for estimating the PR time series and thereafter the PLR. The performed comparative analysis was based on the forecasting accuracy for both PR and PLR, statistical significance and method's simplicity (the results are summarized in section 4.3).

Specifically, for the forecasting accuracy for the PR time series the RMSE metric and the information criterions were used. The ratio $RMSE_{for}/RMSE_{ref}$ express the mean RMSE of the SARIMA models developed by each methodology based on the information criterions ($RMSE_{for}$) divided with the mean RMSE using the SARIMA models as in Ref [6] ($RMSE_{ref}$).

Likewise, the $AIC_{ref}/AIC_{for}$ express the mean AIC value of the using the SARIMA models as in Ref [6] ($AIC_{ref}$), divided by the mean AIC value of the SARIMA models developed by each methodology based on the information criterions ($AIC_{for}$), same for the AICc and BIC, respectively. The value of the information criterion is preferred to have the smallest value (largest negative value), that's why these ratios are reversed in comparison with the one for RMSE.

The forecasting accuracy for the PLR was based on the mean value of Abs Dif between the actual and forecasted PLR for the last 3 years. Moreover, four different values of the mean Abs Dif were used for each PV technology, i.e. mono-c-Si, multi-c-Si and thin films, and separated for the partially shaded PV systems.

For evaluating the statistical significance, the adequacy of the models of each method was evaluated based the residual behaviour by checking their individual autocorrelation coefficients and quantified based on the Ljung-Box portmanteau lack-of-fit test.

The methodology followed in Ref [6] was assumed to be a more empirical approach and a good understanding of the ARIMA theory was needed to identify the SARIMA models. In contrast, with the methods based on the three different information criterions, simple and automatic functions were utilised for the identification of the SARIMA models for each PV system. However, even if automatic functions can identify the SARIMA models, a “manual” supervision of the results is needed to verify the statistical adequacy of the models.

The total score of each methodology is the product of the score for each value mentioned above, with the preferred methodology be the one with the lowest total score.

# *Chapter 4 – Results*

## 4.1 Arima Modelling and Forecasting of the PR time series

As has been stated in section 3.3 forecasting of a given time series $X_t$ requires the availability of a model that captures its time, statistical and spectral behaviour. To develop such a model, the orders and the coefficients of the operators must be identified, and the statistical diagnostic checks should be performed to assure the statistical significance of the model. The lags $p, d, q, P, D, Q$ and $s$ of the SARIMA model were derived using the "auto.arima" function from the "forecast" package in R Project for Statistical Computing [44].

### 4.1.1 Identification of SARIMA model orders

The "auto.arima" function in R returns the best ARIMA or SARIMA model according to the chosen information criterion (e.g. AIC, AICc or BIC) [44]. As aforementioned, the PR time series exhibit seasonality and a downward tendency of the average PR value over periods of 12-month. Therefore, as the dataset are consisted from monthly values the seasonality $s$ and the seasonal differencing $D$ were set equal to 12 and 1, respectively, for all the systems in the "auto.arima" function. The $d$ wasn't set equal to a constant value, its values were chosen automatic from the algorithm to be equal to 0 or 1, depending on PR time series of each PV system.

Each devised model was used to forecast the monthly PR for 3 years, using a 5-year period of monthly PR measurements as the train dataset with the help of the "sarima.for" function of the "astsa" package in R. Table 4.1 summarizes the identified SARIMA models based on the different information criterions AIC, AICc and BIC respectively, as well as the SARIMA models proposed in [6].

**Table 4.1. Identified SARIMA models $(p,d,q),(P,D,Q)_s$ for the PV grid connected systems based on the AIC, AICc and BIC value and as proposed in** [6] **. (*Partially shaded systems)**

| a/a | Manufacturer | AIC | AICc | BIC | Ref [6] |
|---|---|---|---|---|---|
| 01 | Solon * | $(1,0,0),(1,1,0)_{12}$ | $(1,0,0),(1,1,0)_{12}$ | $(1,0,0),(1,1,0)_{12}$ | $(1,1,1),(1,1,1)_{12}$ |
| 02 | Sanyo | $(0,1,1),(1,1,1)_{12}$ | $(0,1,1),(1,1,1)_{12}$ | $(0,1,1),(1,1,1)_{12}$ | $(2,1,1),(1,1,1)_{12}$ |
| 03 | Atersa | $(0,0,0),(1,1,0)_{12}$ | $(0,0,0),(1,1,0)_{12}$ | $(0,0,0),(1,1,0)_{12}$ | $(2,1,1),(1,1,1)_{12}$ |
| 04 | Suntechnics | $(0,0,0),(1,1,0)_{12}$ | $(0,0,0),(1,1,0)_{12}$ | $(0,0,0),(1,1,0)_{12}$ | $(3,1,1),(1,1,1)_{12}$ |
| 05 | Schott Solar (EGF) | $(0,0,0),(1,1,1)_{12}$ | $(0,0,0),(1,1,1)_{12}$ | $(0,0,0),(1,1,1)_{12}$ | $(3,1,1),(1,1,1)_{12}$ |
| 06 | BP Solar * | $(0,1,2),(1,1,0)_{12}$ | $(2,1,0),(1,1,0)_{12}$ | $(0,1,0),(1,1,0)_{12}$ | $(3,1,2),(1,1,0)_{12}$ |
| 07 | SolarWorld | $(0,0,0),(1,1,1)_{12}$ | $(0,0,0),(1,1,1)_{12}$ | $(0,0,0),(1,1,0)_{12}$ | $(3,1,1),(1,1,1)_{12}$ |
| 08 | Schott Solar (MAIN) | $(0,0,0),(1,1,1)_{12}$ | $(0,0,0),(1,1,1)_{12}$ | $(0,0,0),(1,1,0)_{12}$ | $(3,1,1),(1,1,1)_{12}$ |
| 09 | Würth Solar | $(1,0,0),(1,1,0)_{12}$ | $(0,0,0),(1,1,0)_{12}$ | $(0,0,0),(1,1,0)_{12}$ | $(3,1,1),(1,1,1)_{12}$ |
| 10 | First Solar | $(1,1,0),(1,1,0)_{12}$ | $(1,1,0),(1,1,0)_{12}$ | $(1,1,0),(1,1,0)_{12}$ | $(3,1,1),(1,1,1)_{12}$ |
| 11 | MHI | $(0,1,1),(1,1,0)_{12}$ | $(0,1,1),(1,1,0)_{12}$ | $(0,1,1),(1,1,0)_{12}$ | $(3,1,1),(1,1,1)_{12}$ |

The SARIMA models identified with the three different information criterions yielded identical results expect for the BP solar, SolarWorld, Schott Solar (MAIN) and Wurth Solar PV systems. However, none of the obtained models based on the information criterions were in agreement with the ones proposed in [6].

### 4.1.2 Forecasting of the PR time series

The SARIMA models were then used to generate 36 future values of the monthly PR time series. However, it should be noted that these models are purely statistical models and their probability limits should be considered. Usually, the upper and lower 50% and 95% probability limits are used to express the accuracy of the forecasts [37]. The probability limits express that the realized value of the time series, when it eventually occurs, will be included within these limits with the stated probability. Figure 4.1 illustrates 60 actual (black line) the 36 forecasted values (red line) of the monthly PR including the upper and lower 50% and 95% probability limits for three of the eleven PV systems, of different technologies. It can be observed that the probability limits are closed to the forecasted values and do not diverge over time.

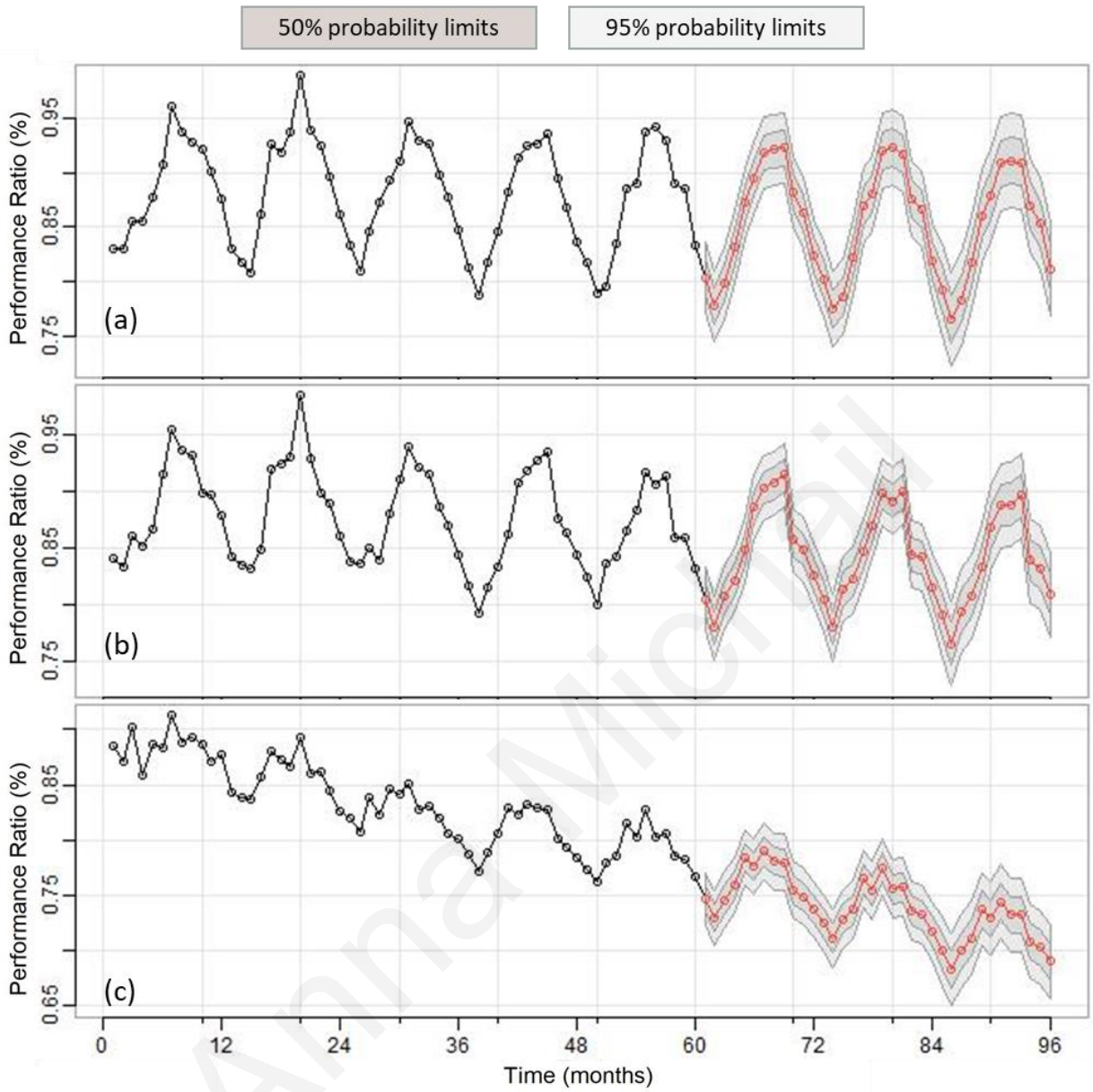


**Figure 4.1. Upper and lower 50% and 95% probability limits for the (a) Atersa mono-c-Si (b) Schott Solar (MAIN) multi-c-Si and (c) First Solar CdTe thin film system using the SARIMA models based on the AIC.**

Figure 4.2 shows the actual and forecasted values of the PR time series over the 8 years for three of the eleven PV systems, of different technologies. The forecasted values are closed to the actual ones for the presented PV systems. Further analysis regarding the forecasting accuracy is performed in the next subsection for all PV systems and methods. The figures with the probability limits and the comparison of the actual and forecasted data for the remaining PV systems and methods are included in Appendix A.

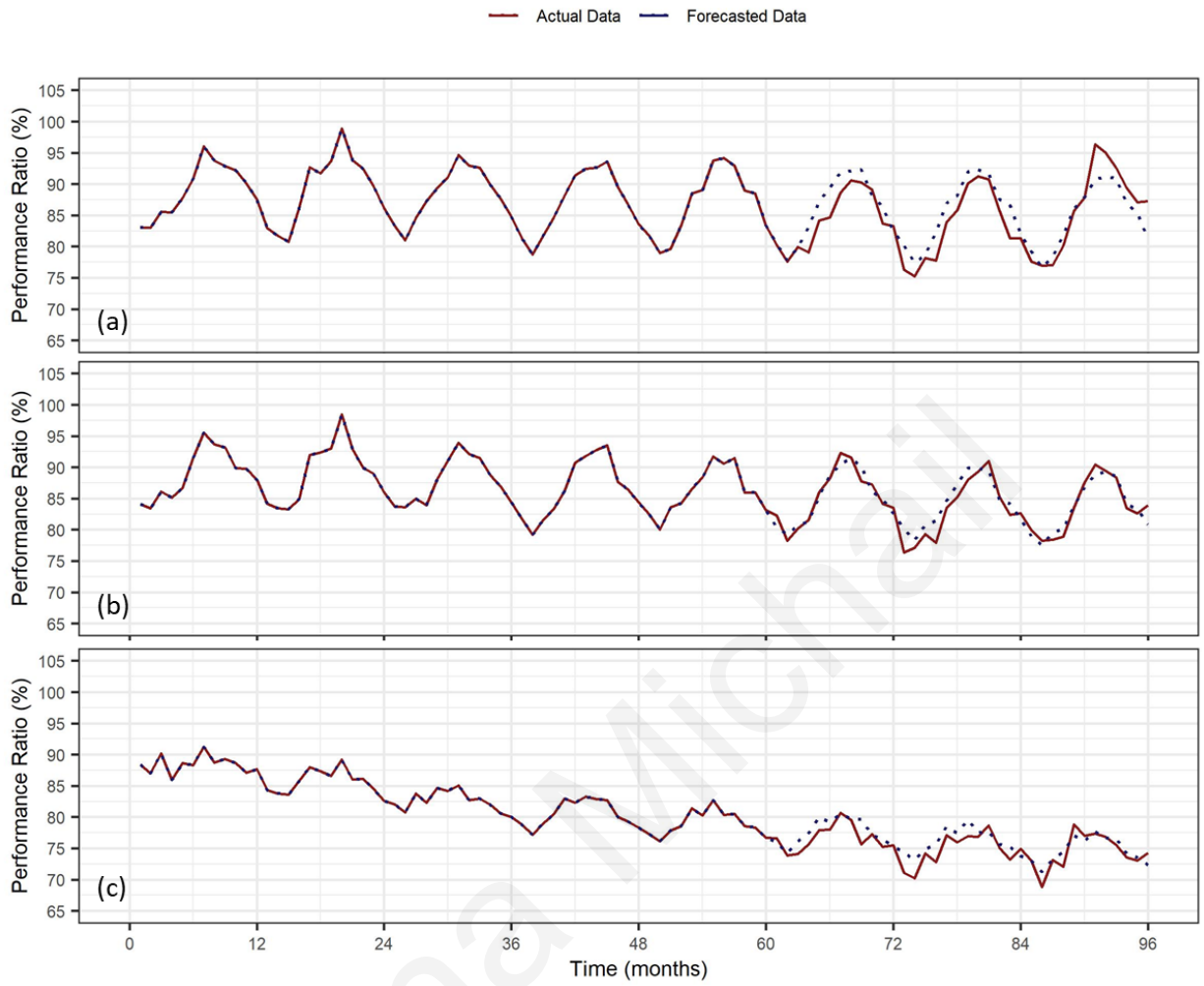


**Figure 4.2. Three years of forecasted PR, with five years of training using the SARIMA models based on the AIC for (a) Atersa mono-c-Si (b) Schott Solar (MAIN) multi-c-Si and (c) First Solar CdTe thin film system.**

During the development of the SARIMA models, the performance loss of the PV systems was modelled as non-stationary, however this is purely statistical property which allows to the model to change, decrease or increase, the yearly and/or monthly average value of the PR. Generally, in terms of planning and policy making is should be considered that the PR values will not increase intensely over time. For this reason, the upper 50% and 95% curve should not be considered. The forecast data is the most likely expected future PR values with the lower 95% PR curve being the statistical worst-case scenario.

### 4.1.3 Robustness of the identified SARIMA model

In this study the forecasted PR values were used to forecast the yearly PLR, for this reason is vital to capture as best as possible the real behaviour of the PR time series of each PV system. For the evaluation of the goodness of fit of the SARIMA models, different parameters were considered, including performance metrics (RMSE and MAE), the values of the information criterions and the behaviour of the residuals.

#### *4.1.3.1 Performance Metrics*

The actual and forecasted PR data (in %), i.e. for the last 3 years, were used to calculate the RMSE and MAE (Figure 4.3) for each PV system based on the AIC, AICc, BIC information criterion or on the method followed in [6].

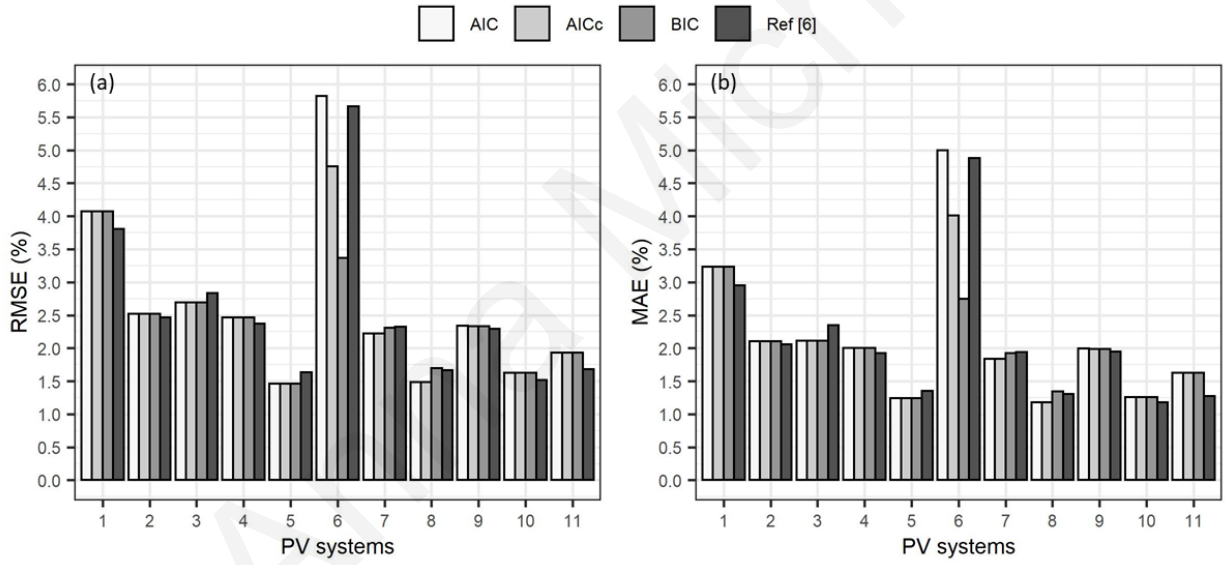


**Figure 4.3. (a) RMSE and (b) MAE between the actual and forecasted PR data (in %) for each PV system based on the AIC, AICc, BIC information criterion or on the method followed in** [6] **for identifying the SARIMA model of each system.**

The values of RMSE and MAE for all methods used for the PV systems under investigation were lower than 6 %, indicating good fitting. The 1st and 6th PV system were affected by partial shading during the years (2nd, 3rd and 4th year) that the training sets were obtained, causing unexpected lower values of PR. This can explain the high values of errors, as the SARIMA models obtained for these PV systems included the effect of partial shading in the behaviour of their PR time series.

Generally, comparing the RMSE and MAE values of each method no sufficient conclusions can be obtain for which method had the best fitting, as their results were very close. Additionally, none

of the methods had the lower values of errors for all the PV systems. However, the RMSE and MAE values using the BIC for the 6th PV system had up to 2.45% and 2.25% difference with the other methods, respectively and with the AIC method having the highest errors for this PV system.

#### *4.1.3.2 Information Criterion*

As aforementioned in Section 3.6.1 information criterions reflect the goodness of fit of the model and the model with the lowest (or higher negative) value of the chosen criterion is selected as optimal model. However, diagnostic checks must be applied afterwards to verify the statistical significance of the model. Even if some methods were based on specific information criterions for the identification of the SARIMA model, calculation of the value of other information can be calculated. Figure 4.4 compares the values of the AIC, AICc and BIC for the identified SARIMA model of each system based on the AIC, AICc, BIC information criterion or on the method followed in [6].

The most striking observation from the comparison presented in the graphs below is that the values of the three information criterions for the models obtained in [6] are in almost all cases higher (smaller negative). This was expected, as the identification of these models was not based on any information. However, there is not a dramatic difference compared to the other methods, which indicates that even a more empirical approach can achieve good fitted models.

Additionally, counterintuitive results can be observed on the AICc and BIC graphs for the models of the 6th PV system. It was expected that the model based on the AICc will have the highest negative value compared to the other model and similarly for the model based on the BIC. However, for the specific PV system the model based on the AIC seems to have the highest value for all three information criterions. This might be attributed to the greater ‘‘penalty factor’’ for the number of estimated parameters which AICc and BIC have in comparison to the AIC and random behaviour of the PR of the system due to the partial shading effects.

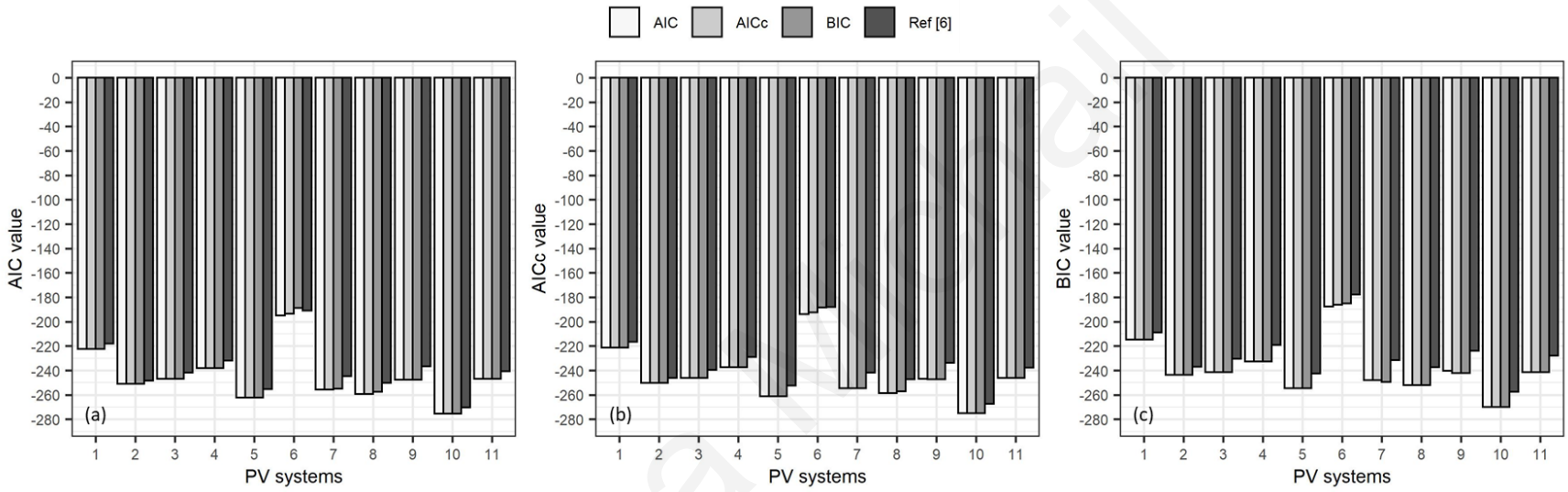


**Figure 4.4. (a) AIC, (b) AICc, and (c) BIC value for the identified SARIMA model of each system based on the AIC, AICc, BIC information criterion or on the method followed in** [6]**.**

*4.1.3.3 Residuals behaviour – Lack of fit test*

The results discussed in the above subsection are not sufficient to indicate the SARIMA model which will best describe the behaviour of the PR time series of a specific PV system. However, they constituted a good first impression of the fitting ability of the model. To assure the quality of the fitting of diagnostic checks must be applied as discussed in Section 3.6.2.

Figure 4.5 illustrates the time plot, ACF, histogram of the residuals and the Q-Q plot that displays the sample residuals with "o" and the theoretical-normal quantiles with a solid line obtained using the different methods to identify the SARIMA model of the BP Solar mono-c-Si PV system. From the graph it can be observed that the residuals of the models based on AIC, AICc and Ref [6] fall within the boundaries of the ACF. Additionally, the residuals of these methods seem to follow at an acceptable level the line of the theoretical quantiles, indicating that the residuals follow the normal distribution. On the contrary, the model based on the BIC has an autocorrelation lag at 2 $\left(r_2^2(\hat{a})\right)$ which have greater value than the lower ACF limit which indicates that the model might missing some crucial information regarding the behaviour of the PR time series. This can be corrected manually by changing the MA order of the model ($q$). However, in the present study no such corrections were performed in order to better compare the ability of the methods developed to capture the behaviour of the modelled time series. Detailed results of the residuals behaviour of the identified models for all PV systems can be found in Appendix A.

Table 4.2 provides a synopsis of the number of lags following out of the ACF boundaries for the each obtained model obtained using the different methods. It can be observed the models based on the BIC have the highest mean value of out-falling lags, of 0.9091. On the contrary, the models proposed in [6] resulted in a significant lower value of out-falling lags with a mean value of 0.2727.

**Table 4.2. Number (No.) and position of lags following out of the residual ACF boundaries for the each obtained model based on the AIC, AICc, BIC and Ref** [6]**.**

| a/a | Manufacturer | AIC | | AICc | | BIC | | Ref [6] | |
|---|---|---|---|---|---|---|---|---|---|
| | | No. | Position | No. | Position | No. | Position | No. | Position |
| 1 | Solon * | 0 | - | 0 | - | 0 | - | 0 | - |
| 2 | Sanyo | 1 | $r_5^2(\hat{a})$ | 1 | $r_5^2(\hat{a})$ | 1 | $r_5^2(\hat{a})$ | 1 | $r_5^2(\hat{a})$ |
| 3 | Atersa | 1 | $r_3^2(\hat{a})$ | 1 | $r_3^2(\hat{a})$ | 1 | $r_3^2(\hat{a})$ | 0 | - |
| 4 | Suntechnics | 1 | $r_3^2(\hat{a})$ | 1 | $r_3^2(\hat{a})$ | 1 | $r_3^2(\hat{a})$ | 0 | - |
| 5 | Schott Solar (EGF) | 1 | $r_3^2(\hat{a})$ | 1 | $r_3^2(\hat{a})$ | 1 | $r_3^2(\hat{a})$ | 0 | - |
| 6 | BP Solar * | 0 | - | 0 | - | 1 | $r_2^2(\hat{a})$ | 0 | - |
| 7 | SolarWorld | 1 | $r_3^2(\hat{a})$ | 1 | $r_3^2(\hat{a})$ | 1 | $r_3^2(\hat{a})$ | 0 | - |
| 8 | Schott Solar (MAIN) | 1 | $r_3^2(\hat{a})$ | 1 | $r_3^2(\hat{a})$ | 1 | $r_3^2(\hat{a})$ | 0 | - |
| 9 | Würth Solar | 0 | - | 1 | $r_4^2(\hat{a})$ | 1 | $r_4^2(\hat{a})$ | 0 | - |
| 10 | First Solar | 0 | - | 0 | - | 0 | - | 0 | - |
| 11 | MHI | 2 | $r_7^2(\hat{a}), r_{13}^2(\hat{a})$ | 2 | $r_7^2(\hat{a}), r_{13}^2(\hat{a})$ | 2 | $r_7^2(\hat{a}), r_{13}^2(\hat{a})$ | 2 | $r_7^2(\hat{a}), r_{13}^2(\hat{a})$ |
| **Mean** | | **0.7273** | | **0.8182** | | **0.9091** | | **0.2727** | |

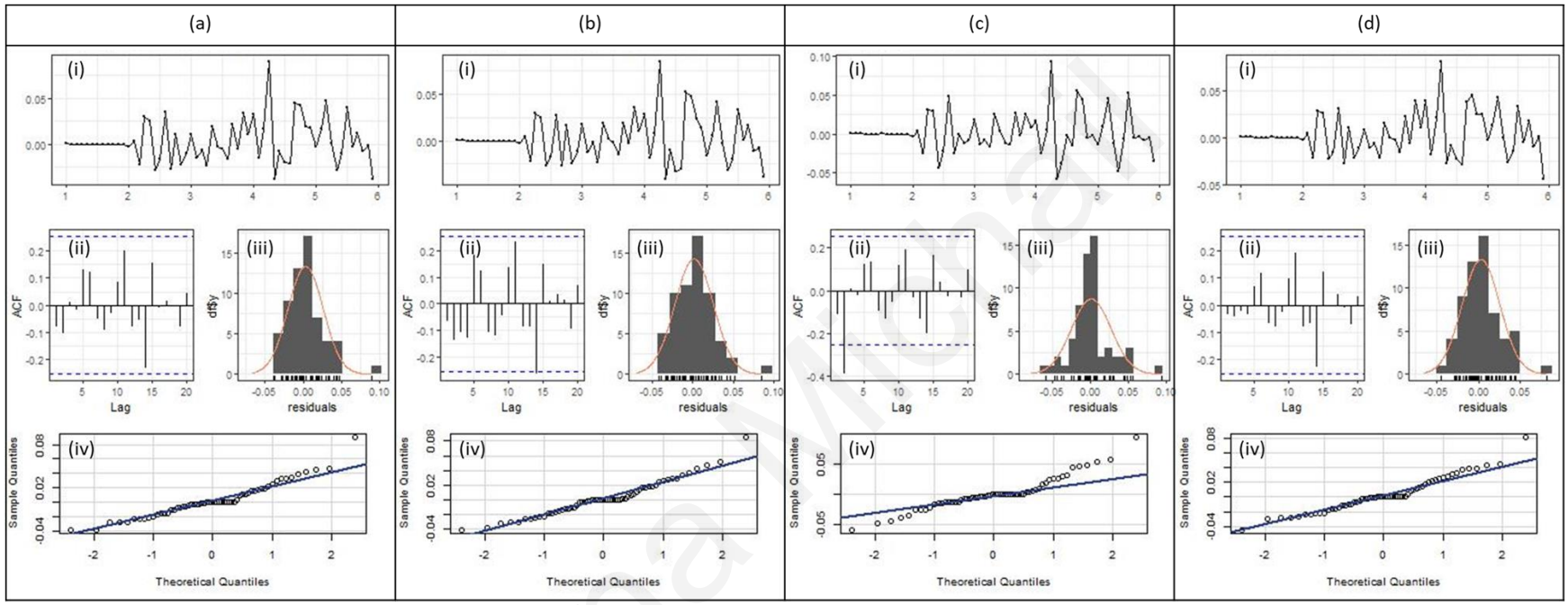


**Figure 4.5. (i) Time plot, (ii) ACF, (iii) histogram of the residuals and (iv) Q-Q plot that displays the sample residuals with "o" and the theoretical-normal normal quantiles with a solid line obtained using (a) AIC, (b) AICc, (c) BIC information criterion or (d) the method followed in** [6] **for identifying the SARIMA model of the BP Solar mono-c-Si PV system.**

Additionally, Table 4.3 shows the results of the Ljung-Box portmanteau lack-of-fit test for the four different methods used to obtain the SARIMA model for each PV system. As discussed before for the specific test, the $p$-value must not be lower than the significance level ($\alpha = 0.05$) in order for the residuals to meet the criteria of a white noise behaviour.

**Table 4.3. Ljung-Box portmanteau lack-of-fit test using the command "Box.test" for the SARIMA models of each PV system using the AIC, AICc, BIC or Ref.** [6] **method.**

| a/a | $H_a = AIC$ | | $H_a = AICc$ | | $H_a = BIC$ | | $H_a = Ref$ [6] | |
|---|---|---|---|---|---|---|---|---|
| | $\widetilde{Q}$ | $p$-value | $\widetilde{Q}$ | $p$-value | $\widetilde{Q}$ | $p$-value | $\widetilde{Q}$ | $p$-value |
| 1 | 6.1402 | 0.9088 | 6.1402 | 0.9088 | 6.1402 | 0.9088 | 4.0659 | 0.9822 |
| 2 | 12.8676 | 0.3787 | 12.8676 | 0.3787 | 12.8676 | 0.3787 | 15.8905 | 0.1963 |
| 3 | 15.3008 | 0.2254 | 15.3008 | 0.2254 | 15.3008 | 0.2254 | 9.6461 | 0.6470 |
| 4 | 17.4354 | 0.1339 | 17.4354 | 0.1339 | 17.4354 | 0.1339 | 3.7766 | 0.9871 |
| 5 | 17.0974 | 0.1460 | 17.0974 | 0.1460 | 17.0974 | 0.1460 | 6.5053 | 0.8885 |
| 6 | 8.0798 | 0.7789 | 14.5828 | 0.2650 | 18.9065 | 0.0908 | 6.4213 | 0.8934 |
| 7 | 10.8927 | 0.5381 | 10.8927 | 0.5381 | 11.8817 | 0.4552 | 4.5933 | 0.9702 |
| 8 | 13.2040 | 0.3544 | 13.2040 | 0.3544 | 18.7786 | 0.0940 | 3.9506 | 0.9843 |
| 9 | 10.4405 | 0.5774 | 16.3902 | 0.1740 | 16.3902 | 0.1740 | 7.8072 | 0.8000 |
| 10 | 8.7974 | 0.7201 | 8.7974 | 0.7201 | 8.7974 | 0.7201 | 9.1364 | 0.6912 |
| 11 | 16.2636 | 0.1795 | 16.2636 | 0.1795 | 16.2636 | 0.1795 | 17.5518 | 0.1300 |
| $\frac{1}{Mean}$ | **0.0806** | **2.2262** | **0.0738** | **2.7336** | **0.0688** | **3.1370** | **0.1231** | **1.3463** |

Figure 4.6 depicts the $p$-value of the Ljung-Box portmanteau lack-of-fit test for the SARIMA models of each PV system using the AIC, AICc, BIC or Ref [6] method. The p-values for all the models were higher than the significance level of $a = 0.05$. This indicates that the SARIMA models fitted are adequate and do not lack of fit. However, it can be seen that the $p$-values of the models proposed in [6] are higher in most cases compared with the other methods, indicating that behaviour of their residuals has higher possibility to follow a white noise behaviour. On the other hand, the $p$-values of the models based on the BIC exhibit the lower values.

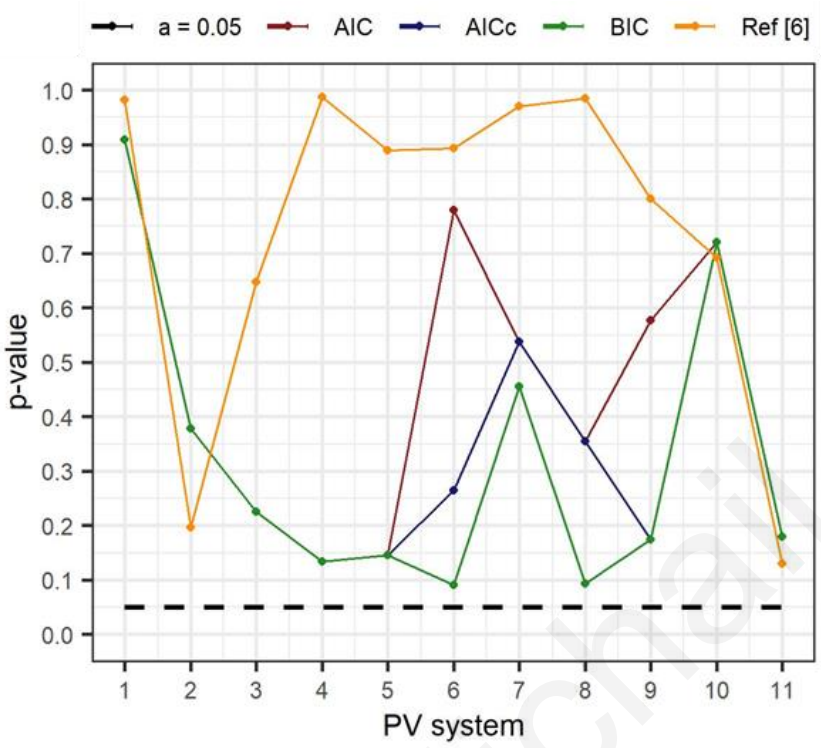


**Figure 4.6. p-values from the Ljung-Box portmanteau lack-of-fit test for the SARIMA models of each PV system using the AIC, AICc, BIC or Ref.** [6] **method, the significance level ($\alpha = 0.05$) appears as a dashed line.**

## 4.2 Estimation of the PLR

The RPCA data is used to define the PLR of the PV systems. Figure 4.7 shows the effect of the RPCA analysis on the actual dataset. In particular, (a) indicates the raw measurements and (b) indicates the actual dataset after procced with RPCA for the three PV systems, Atersa mono-c-Si, Schott Solar (MAIN) multi-c-Si and First Solar thin film. It can be observed that the RPCA remove the outliers in the yearly time series of each system by extracting the data matrix $\boldsymbol{K}$, that is the matrix after the removal of matrix $\boldsymbol{E}$, which causes the outliers from the original data matrix $\boldsymbol{D}$.

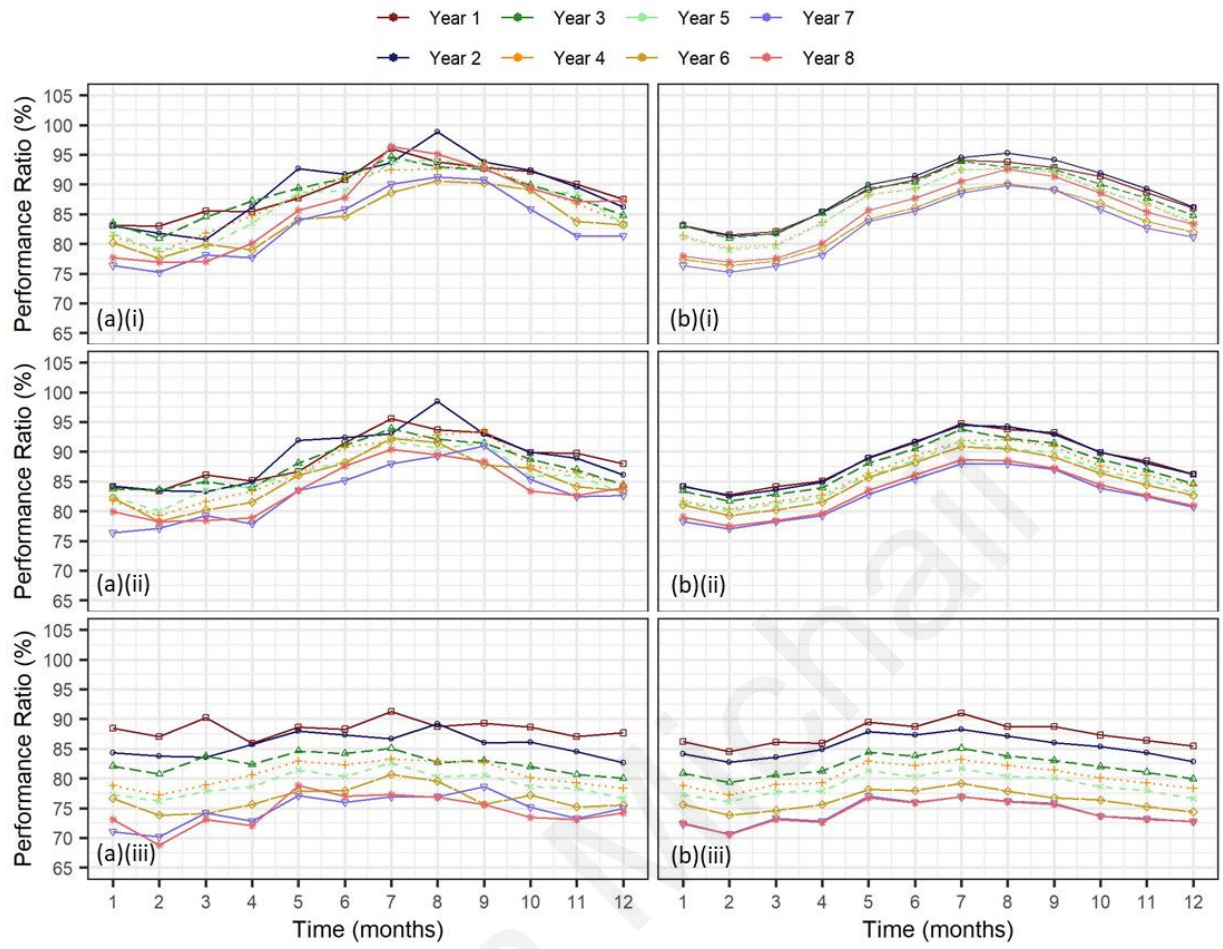


**Figure 4.7. The monthly PR observations (a) before and (b) after implementing RPCA for the (i) Atersa mono-c-Si (ii) Schott Solar (MAIN) multi-c-Si and (iii) First Solar CdTe thin film system.**

Following the procedure described in Section 3.5, and in particular using Eq. (3.26), the PLR was calculated for all the SARIMA models using the different methods at the end of the 6$^{th}$, 7$^{th}$ and 8$^{th}$ year for all the eleven PV systems. Figure 4.8 shows the PLR calculated from the RPCA actual dataset for the (a) 6$^{th}$, 7$^{th}$ and 8$^{th}$ years of the eleven PV systems and (b) the maximum, mean, median and minimum value over the last 3 years categorized by PV technology. The median and mean PLR values fall within the reported values in literature, i.e. median and mean value between $0.5-0.6\ \%/yr$ and $0.8-0.9\ \%/yr$, respectively for c-Si and with the median and mean value for thin-film greater than $1\ \%/yr$ and almost equal to $1.4\ \%/yr$, respectively. These results confirming the reliability of the RPCA analysis for calculate the PLR.

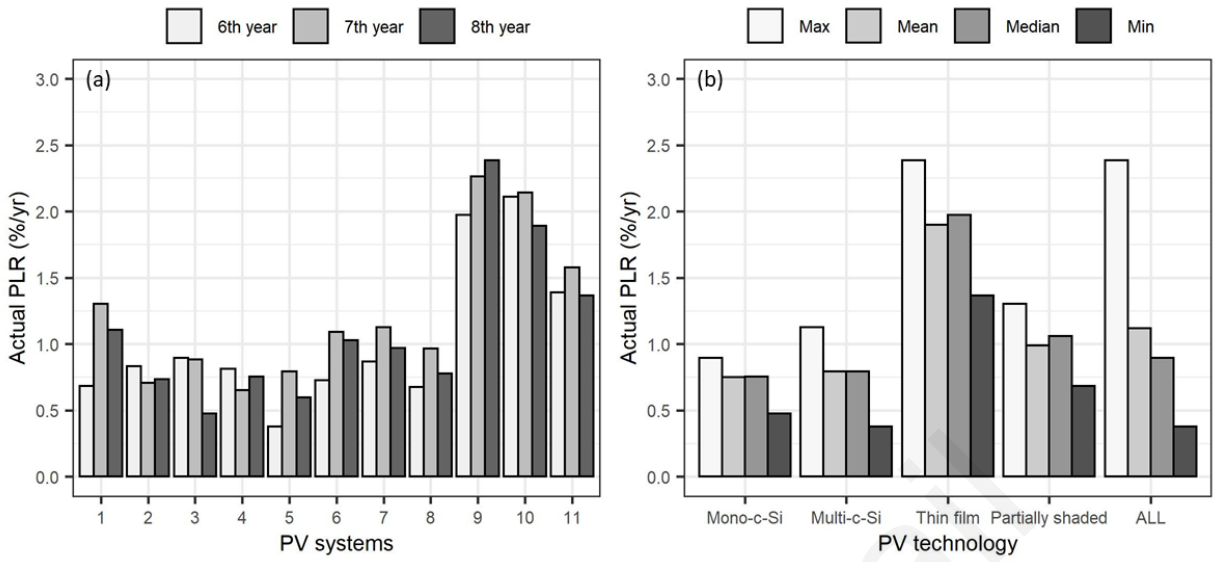


**Figure 4.8. PLR for the RPCA actual datasets (a) for the 6th, 7th and 8th year for each PV system and (b) maximum, mean, median and minimum values over the last 3 years categorized by PV technology.**

The forecasted values of the PLR are presented in Figure 4.9 for the 6th, 7th and 8th years of the eleven PV systems using the different methods based on the AIC, AICc and BIC information criterion and the method proposed in [6]. From a quick comparison between the Figure 4.8.a and the graphs in Figure 4.9 it can be observed that the AIC, AICc and Ref [6] methods have surprisingly very lower value for the PLR of the 6th PV system in comparison with the actual and the forecast PLR value based on the BIC. Figure 4.10 compares the mean and median values of each method categorized by PV technology where is evident again that for the partial shaded PV systems the forecasted values of the PLR were quite low compared to the actuals.

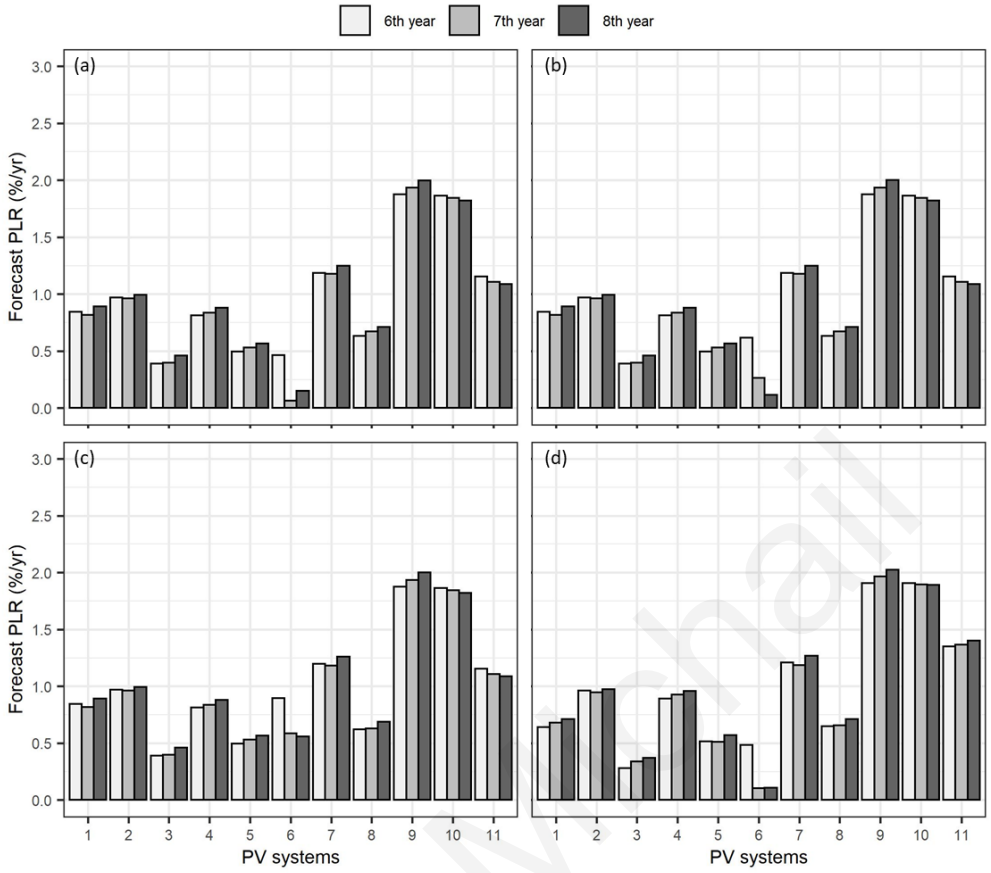


**Figure 4.9. PLR for the RPCA forecasted datasets for the 6th, 7th and 8th year for each PV system using the SARIMA models based on (a) AIC, (b) AICc, (c) BIC and (d) Ref** [6]**.**

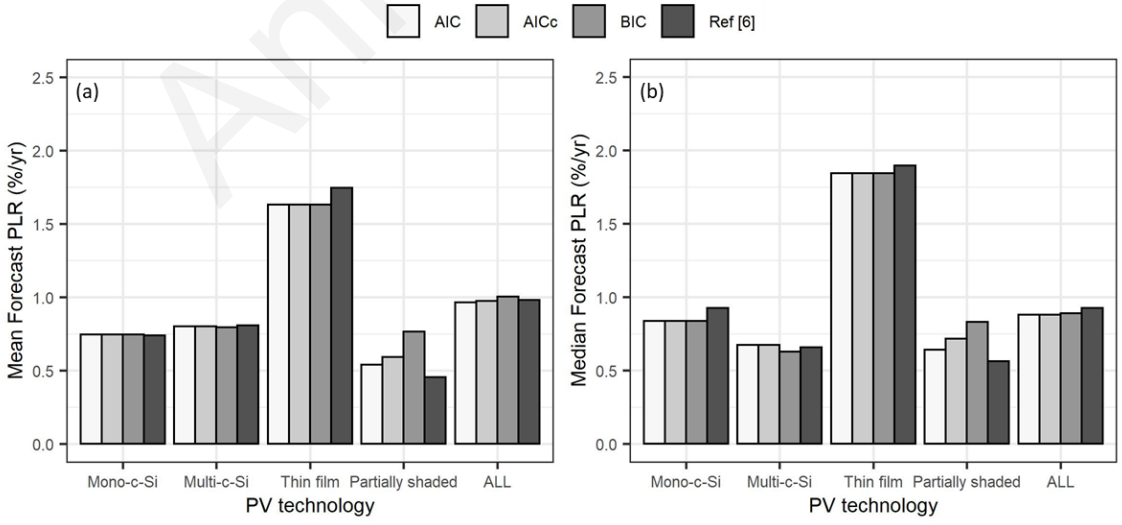


**Figure 4.10. (a) Mean and (b) median values of the abs dif. over the last 3 years for the PLR for the RPCA forecasted datasets categorized by PV technology and method (AIC, AICc, BIC and Ref** [6]**).**

To further compare the forecasted PLR with the actual PLR and quantify the ability of each methodology the Abs Dif (Eq. (3.33)) was calculated. The results of the Abs Dif are illustrated in Figure 4.11 per PV system over the last three years of operation, and the mean and median values of Abs Dif categorised by PV technology are showed in Figure 4.12. These results show that all methods manage to estimate the PLR sufficiently with Abs Dif up to $1.2\ \%/yr$. Also, the results based on the BIC achieve Abs Dif values below of $0.51\ \%/yr$, especially due to the better estimations obtained for the 6th PV system. Generally, all the methods have the higher Abs Dif of the PLR for the partially shaded PV systems. The results obtained based on the proposed models of Ref [6] shows better results for the thin-film PV systems, however for the partially shaded PV systems their mean and median Abs Dif is higher than the other methods.

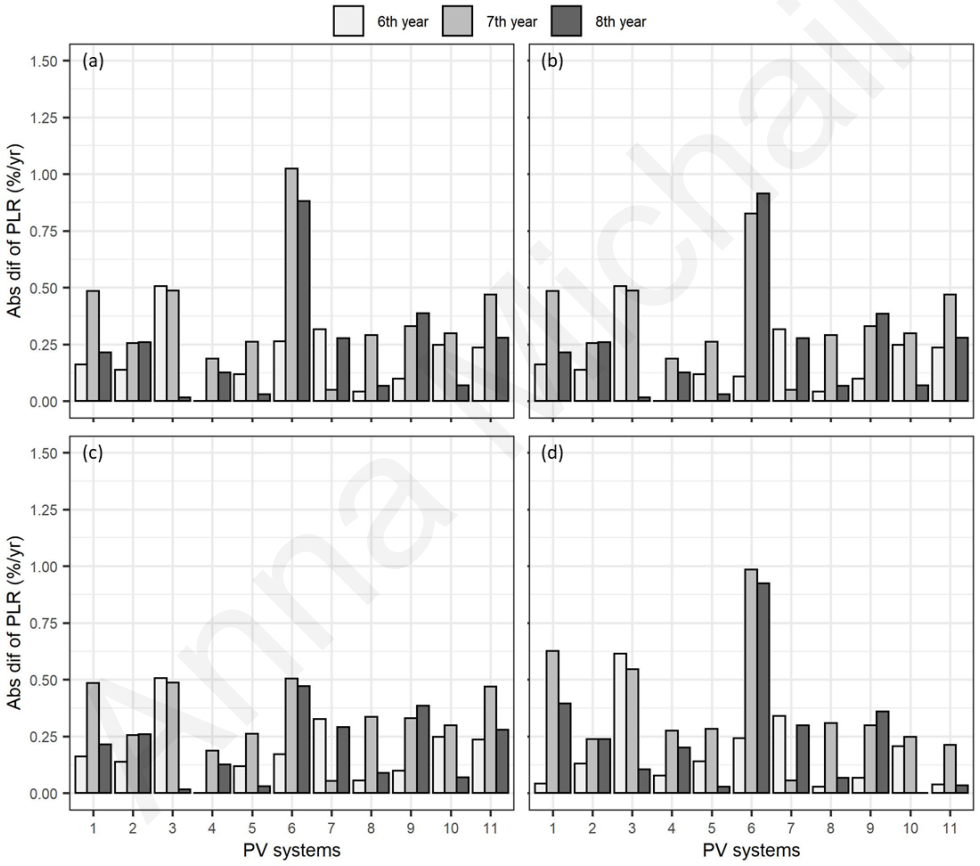


**Figure 4.11. PLR for the Abs Dif between the RPCA actual and forecasted datasets for the 6th, 7th and 8th year for each PV system using the SARIMA models based on (a) AIC, (b) AICc, (c) BIC and (d) Ref** [6]**.**

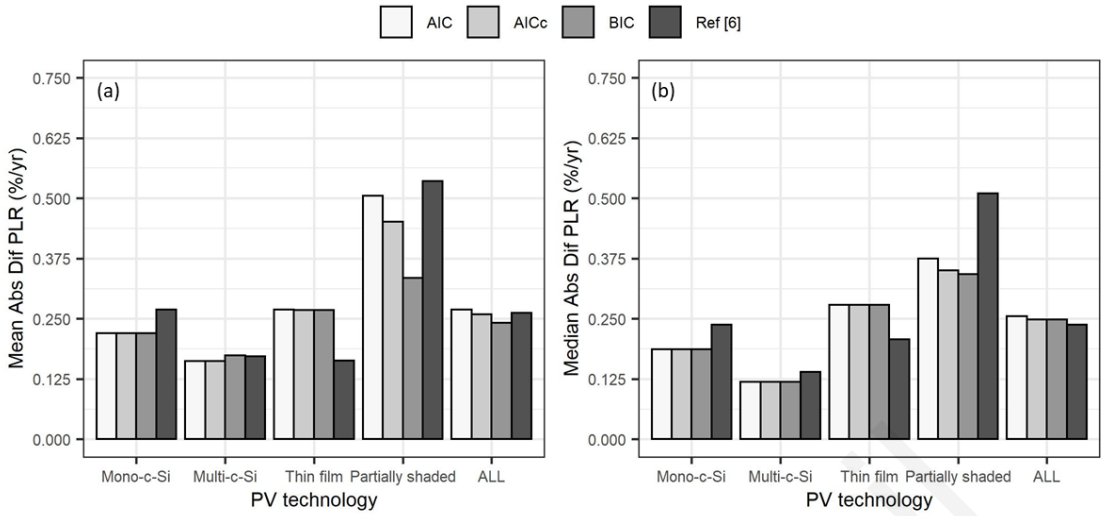


**Figure 4.12. (a) Mean and (b) median values of the Abs Dif over the last 3 years between the PLR for the RPCA actual and forecasted datasets categorized by PV technology and method (AIC, AICc, BIC and Ref** [6]**).**

## 4.3 Comparison of methods

In the above subsections, different performance metrics and statistical analysis were performed for the assessment of the goodness of the methods used to estimate the PLR. However, the individual results of each analysis were not sufficient in order to indicate the optimal methodology for this scope. For this reason, a comparative analysis was performed to combine all the individual parameters as stated in Section 3.6.4. Table 4.4 summarises the results of the comparative analysis. The optimal method for forecasting the PLR of PV systems is the one achieving the lowest score.

The values of the total score of all methods were very closed in general. The models based on BIC had the lowest scores for the forecasting accuracy, but the statistical significance scores of the method was quite high. However, the method achieved the most preferable score with a total score of 0.01807.

Furthermore, the total scores based on AIC and AICc were slightly higher than the BIC, with the method based on AIC exhibit the second lower value with a total score 0.01939 and the method based on AICc having a total score of 0.02049.

It is important to note that the method followed in Ref. [6] was assumed to be a more empirical approach, and for this reason its simplicity score was set equal to 4 and for the other three methods equal to 2. However, this method achieved to have the lowest score regarding the residual's behaviour, showing that the residuals of this method are closest to a white noise behaviour. Depside that, this method achieved a total score of 0.02181, which was the higher score among the methods, ranking it to the least preferred method.

**Table 4.4. Comparative analysis of the AIC, AICc, BIC and Ref.** [6] **methods to identify the SARIMA models and the total score of each method.**

| **Method** | **Forecasting accuracy** | | | | | | | | **Statistical significance** | **Simplicity** | **Total score** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **PR** | | | | **PLR** | | | | **Residuals Behaviour** $H_0 = theoretical\ residuals$ | | |
| | **Performance metric** | **Information Criterion** | | | **Mono-c-Si** | **Multi-c-Si** | **Thin Film** | **Partially Shaded** | **Ljung-Box portmanteau lack-of-fit test** | | |
| | $\frac{RMSE_{for}}{RMSE_{ref}}$ | $\frac{AIC_{ref}}{AIC_{for}}$ | $\frac{AICc_{ref}}{AICc_{for}}$ | $\frac{BIC_{ref}}{BIC_{for}}$ | $Mean(Abs\ Dif\ of\ PLR)$ | | | | $\frac{1}{Mean(p\text{-}values)}$ | | |
| **Score Range** | $0^+$ high – 5 low skill of forecasting in comparison with Ref. [6]<br>(1 = no difference) | $0^+$ high – 5 low skill of capturing information by the model in comparison with Ref. [6]<br>(1 = no difference) | | | $0^+$ low – 5 high error between the actual and forecasted data | | | | 1 white noise behaviour – >20 rejection of the $H_0$ | 1 Simple – 5 Complex | **Prefer the lowest value** |
| **Ref.** [6] | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.2693 | 0.1722 | 0.1631 | 0.5354 | 1.3463 | 4 | **0.02181** |
| **AIC** | 1.0135 | 0.9733 | 0.9658 | 0.9495 | 0.2196 | 0.1618 | 0.2685 | 0.5047 | 2.2262 | 2 | **0.01939** |
| **AICc** | 0.9755 | 0.9739 | 0.9663 | 0.9494 | 0.2196 | 0.1618 | 0.2682 | 0.4514 | 2.7336 | 2 | **0.02049** |
| **BIC** | 0.9373 | 0.9765 | 0.9684 | 0.9494 | 0.2196 | 0.1737 | 0.2682 | 0.3347 | 3.1370 | 2 | **0.01807** |

# *Chapter 5 – Conclusions and Future Work*

## 5.1 Summary

No standardised method is available for PLR estimation in the literature [3], [4]. Therefore, the current study's proposal was to determine an optimal methodology for estimating accurately the PLR of fielded PV systems.

In this investigation, field measurements were considered for the evaluation of different statistical methods. In particular, electrical and meteorological data from eleven grid-connected PV systems of different technologies located in Nicosia, Cyprus were used to construct the monthly PR time series over 8 years.

The SARIMA model was then used for the identification of a statistical model which is able to capture the time, statistical, and spectral behavior of the PR time series and accurate forecast future values. Four different methods were used for the identification of the SARIMA models, three of them based on three different information criterions, namely the AIC, AICc and BIC, and the fourth based on the ACF and PCF as proposed in [6]. Five years of the monthly PR time series were used as the train dataset to identify and estimate the SARIMA models for each PV system, which was considered as a minimum training period for the investigated PV technologies. The PLR of the PV systems was estimated using the forecasted data from four different methods after the RPCA implementation.

The optimal method was chosen based on a comparative analysis. Specifically, the comparative analysis was based on the forecasting accuracy for both PR and PLR, statistical significance and the simplicity of the methods. For the forecasting accuracy of the PR time series the RMSE metric and the information criterions were used. On the other hand, the forecasting accuracy of the PLR was based on the mean value of absolute difference between the actual and forecasted PLR, for each PV technology and separated for the partially shaded PV systems. Moreover, the statistical significance and the adequacy of the models of each method was evaluated based on residual analysis. Lastly, regarding the methods' simplicity the method proposed in [6] was assumed to be a more empirical approach, where a more deeply understanding of the ARIMA theory was needed, was assumed to be a more empirical approach were a more deeply understanding of the ARIMA theory. On the contrary, the information criterions -based methods were considered easiest to developed open access functions were used, with the obtained SARIMA models exhibiting good results.

## 5.2 Conclusions

The results for the SARIMA models showed that the three different information criterions exhibited identical results for 7 PV systems. However, none of the obtained models using on the information criterions were in agreement with the ones proposed in [6].

During the model adequacy assessment, all the methods showed good fitting of the forecasted PR values, with the RMSE and MAE values being below 6%. Nevertheless, no sufficient conclusion could be taken based on the RMSE and MAE values regarding the optimal PLR method, as their results were very close. However, the RMSE and MAE values using the BIC for the 6$^{th}$ PV system had up to 2.45% and 2.25% difference with the other methods, respectively and with the AIC method having the highest errors for this PV system.

Comparing the AIC, AICc and BIC for the identified SARIMA models of each method, it was observed that models obtained in [6] had in almost all cases the least preferable values. Though, there was not a dramatic difference compared to the other methodologies. This was expected, as the identification of these models was not based on any information criterion. On the other hand, counterintuitive results were observed on the values of the AICc and BIC for the methods based on these values. Specifically, for the models of the 6$^{th}$ PV system the model based on the AICc had a less preferable AICc value compared to the model based on AIC, and similarly for the model based on the BIC. The model based on AIC exhibited the most preferable value of all three information criterions for the 6$^{th}$ PV system. This might be attributed to the greater ''penalty factor'' for the number of estimated parameters which AICc and BIC have in comparison to the AIC and random behaviour of the PR of the system due to the partial shading effects.

The analysis of the residuals, and the inspection of the residuals' autocorrelation coefficients showed that the models based on the BIC had the highest mean value, 0.9091 of out-falling lags per model. On the contrary, the models proposed in [6] resulted in a significant lower value of out-falling lags per model with a mean value of 0.2727. Additionally, the Ljung-Box portmanteau lack-of-fit test showed results that were consistent with the residuals' autocorrelation coefficients. In particular, the $p$-values of the models proposed in [6] were higher in most cases compared with the other methods, indicating that behaviour of their residuals has higher possibility to follow a white noise behaviour. On the other hand, the $p$-values of the models based on the BIC exhibit the lower values. However, the p-values for all the models were higher than the significance level of $a = 0.05$, indicating that the fitted SARIMA models are adequate and do not lack of fit.

Moving to the results of the PLR, the actual and forecasted PLR values were in agreement with the PLR values reported in literature for the c-Si and thin-film technologies, confirming the reliability of the RPCA method to calculate the PLR. Comparing the actual and forecasted PLR values it was observed that all methods manage to estimate the PLR sufficiently with absolute difference up to $1.2\ \%/yr$. Also, the results based on the BIC achieve absolute difference values below of $0.51\ \%/yr$, especially due to the better estimations obtained for the 6$^{th}$ PV system. Generally, all the methods exhibit higher values of PLR absolute difference for the partially shaded PV systems. The results obtained based on the proposed models of Ref [6] showed better results for the thin-film PV systems, however for the partially shaded PV systems their mean and median absolute difference were higher than the other methods.

Finally, the comparative analysis performed showed that the values of the total score of all methods were very closed in general. The models based on BIC had the lowest scores for the forecasting accuracy, but the statistical significance scores of the method was quite high. However, this method achieved the most preferable score with a total score of 0.01807. Furthermore, the total scores based on AIC and AICc were slightly higher than the BIC, with the method based on AIC exhibit the second lower value with a total score of 0.01939 and the method based on AICc having a total score of 0.02049. Last but not least, the method proposed in [6] had the lowest score regarding the residuals' behaviour showing that the residuals of this method are closest to white noise behaviour. Depside that, this method achieved a total score of 0.02181, which was the higher score among the methods, ranking it to the least preferred method.

## 5.3 Future Work

The need for a statistically significant model was urge to accurate capture the behaviour of the chosen performance metric with the ultimate goal the estimation of the PLR. However, the results of this study indicated that even if the method proposed in [6] having the optimal residual's behaviour for all models, the values of RMSE and other metrics were not always superior in comparison with the other methods. This indicates that further investigation needs to be performed regarding the statistical significance of the models to fully understand its impact on the PLR estimations.

Moreover, modification of the identified models could be performed based on the autocorrelation coefficients of the residuals, especially for the ones based on the information criterion. This could further improve the efficacy of the models and ensure the high statistical significance of the models.

Lastly, the methodology proposed in this thesis should be tested on different PV systems topologies and on different locations to test its robustness.

# References


[1] (IEA) International Energy Agency, “Snapshot of Global PV Markets 2021,” 2021. [Online]. Available: http://www.iea-pvps.org/fileadmin/dam/public/report/technical/PVPS_report_-_A_Snapshot_of_Global_PV_-_1992-2014.pdf.

[2] M. Theristis *et al.*, “Nonlinear Photovoltaic Degradation Rates: Modeling and Comparison Against Conventional Methods,” in *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)*, Jun. 2020, vol. 2020-June, no. 4, pp. 0208–0212, doi: 10.1109/PVSC45281.2020.9300388.

[3] D. C. Jordan and S. R. Kurtz, “Photovoltaic Degradation Rates - An Analytical Review,” *Prog. Photovoltaics Res. Appl.*, vol. 21, no. 1, pp. 12–29, Jan. 2013, doi: 10.1002/pip.1182.

[4] A. Phinikarides, N. Kindyni, G. Makrides, and G. E. Georghiou, “Review of photovoltaic degradation rate methodologies,” *Renew. Sustain. Energy Rev.*, vol. 40, pp. 143–152, 2014, doi: 10.1016/j.rser.2014.07.155.

[5] I. Romero-Fiances *et al.*, “Impact of duration and missing data on the long-term photovoltaic degradation rate estimation,” *Renew. Energy*, vol. 181, pp. 738–748, Jan. 2022, doi: 10.1016/j.renene.2021.09.078.

[6] E. Pieri, A. Kyprianou, A. Phinikarides, G. Makrides, and G. E. Georghiou, “Forecasting degradation rates of different photovoltaic systems using robust principal component analysis and ARIMA,” *IET Renew. Power Gener.*, vol. 11, no. 10, pp. 1245–1252, 2017, doi: 10.1049/iet-rpg.2017.0090.

[7] A. Vikram, “What to know about a solar panel warranty,” *Energy Sage*, 2021. https://news.energysage.com/shopping-solar-panels-pay-attention-to-solar-panels-warranty/ (accessed Nov. 14, 2021).

[8] A. Phinikarides, “Degradation Rate Estimation in Photovoltaics,” University of Cyprus, 2017.

[9] U.S. Department of Energy, Sandia, and NREL, “Accelerated Aging Testing and Reliability in Photovoltaics,” *Sol. Energy Technol. Progr. Work. II*, 2008.

[10] Köntges M. *et al.*, *IEA-PVPS T13-01 2014 Review of Failures of Photovoltaic Modules Final*, no. July. 2014.

[11] S. Pulver, D. Cormode, A. Cronin, D. Jordan, S. Kurtz, and R. Smith, “Measuring degradation rates without irradiance data,” in *2010 35th IEEE Photovoltaic Specialists Conference*, Jun. 2010, no. February, pp. 001271–001276, doi: 10.1109/PVSC.2010.5614208.

[12] A. M. Reis, N. T. Coleman, M. W. Marshall, P. A. Lehman, and C. E. Chamberlin, “Comparison of PV module performance before and after 11-years of field exposure,” *Conf. Rec. Twenty-Ninth IEEE Photovolt. Spec. Conf.*, pp. 1432–1435, 2002, doi: 10.1109/PVSC.2002.1190878.

[13] C. R. Osterwald, A. Anderberg, S. Rummel, and L. Ottoson, “Degradation analysis of weathered crystalline-silicon PV modules,” in *Conference Record of the Twenty-Ninth IEEE Photovoltaic Specialists Conference*, pp. 1392–1395, doi: 10.1109/PVSC.2002.1190869.

[14] P. Sánchez-Friera, M. Piliougine, J. Peláez, J. Carretero, and M. Sidrach de Cardona,

"Analysis of degradation mechanisms of crystalline silicon PV modules after 12 years of operation in Southern Europe," *Prog. Photovoltaics Res. Appl.*, vol. 19, no. 6, pp. 658–666, Sep. 2011, doi: 10.1002/pip.1083.

[15] D. C. Jordan, S. R. Kurtz, K. VanSant, and J. Newmiller, "Compendium of photovoltaic degradation rates," *Prog. Photovoltaics Res. Appl.*, vol. 24, no. 7, pp. 978–989, Jul. 2016, doi: 10.1002/pip.2744.

[16] C. R. Osterwald, T. J. McMahon, and J. A. del Cueto, "Electrochemical corrosion of SnO2:F transparent conducting layers in thin-film photovoltaic modules," *Sol. Energy Mater. Sol. Cells*, vol. 79, no. 1, pp. 21–33, Aug. 2003, doi: 10.1016/S0927-0248(02)00363-X.

[17] C. Radue and E. E. van Dyk, "A comparison of degradation in three amorphous silicon PV module technologies," *Sol. Energy Mater. Sol. Cells*, vol. 94, no. 3, pp. 617–622, Mar. 2010, doi: 10.1016/j.solmat.2009.12.009.

[18] D. C. Jordan and S. R. Kurtz, "Field Performance of 1.7 GW of Photovoltaic Systems," *IEEE J. Photovoltaics*, vol. 5, no. 1, pp. 243–249, Jan. 2015, doi: 10.1109/JPHOTOV.2014.2361667.

[19] IEC 61724-1:2017, "Photovoltaic system performance - Part 1: Monitoring," vol. 32, pp. 2846–2855, 2017.

[20] A. Kimber *et al.*, "Improved test method to verify the power rating of a photovoltaic (PV) project," in *2009 34th IEEE Photovoltaic Specialists Conference (PVSC)*, Jun. 2009, no. July, pp. 000316–000321, doi: 10.1109/PVSC.2009.5411670.

[21] D. Jordan, "Methods for Analysis of Outdoor Performance Data (Presentation)," Golden, CO (United States), Feb. 2011. doi: 10.2172/1009680.

[22] S. Spataru, D. Sera, T. Kerekes, and R. Teodorescu, "Diagnostic method for photovoltaic systems based on light I-V measurements," *Sol. Energy*, vol. 119, pp. 29–44, 2015, doi: 10.1016/j.solener.2015.06.020.

[23] D. L. King, W. E. Boyson, and J. A. Kratochvil, "Photovoltaic array performance model.," Albuquerque, NM, and Livermore, CA, Aug. 2004. doi: 10.2172/919131.

[24] C. M. Whitaker *et al.*, "Application and validation of a new PV performance characterization method," in *Conference Record of the Twenty Sixth IEEE Photovoltaic Specialists Conference - 1997*, 1997, pp. 1253–1256, doi: 10.1109/PVSC.1997.654315.

[25] S. Smith, T. Townsend, C. Whitaker, and S. Hester, "Photovoltaics for utility-scale applications: project overview and data analysis," *Sol. Cells*, vol. 27, no. 1–4, pp. 259–266, Oct. 1989, doi: 10.1016/0379-6787(89)90034-3.

[26] S. J. Ransome and J. H. Wohlgemuth, "Predicting kWh/kWp performance for amorphous silicon thin film modules," in *Conference Record of the Twenty-Eighth IEEE Photovoltaic Specialists Conference*, 2000, pp. 1505–1508, doi: 10.1109/PVSC.2000.916180.

[27] B. Zinßer, G. Makrides, W. Schmitt, G. Georghiou, and J. Werner, "Annual Energy Yield of 13 Photovoltaic Technologies in Germany and in Cyprus," *22nd EU-PVSEC*, no. January, 2007.

[28] D. C. Jordan and S. R. Kurtz, "The Dark Horse of Evaluating Long-Term Field Performance—Data Filtering," *IEEE J. Photovoltaics*, vol. 4, no. 1, pp. 317–323, Jan. 2014, doi: 10.1109/JPHOTOV.2013.2282741.

[29] B. Marion *et al.*, “Performance parameters for grid-connected PV systems,” in *Conference Record of the Thirty-first IEEE Photovoltaic Specialists Conference, 2005.*, 2005, pp. 1601–1606, doi: 10.1109/PVSC.2005.1488451.

[30] G. Makrides, B. Zinsser, G. E. Georghiou, M. Schubert, and J. H. Werner, “Degradation of different photovoltaic technologies under field conditions,” in *2010 35th IEEE Photovoltaic Specialists Conference*, Jun. 2010, pp. 002332–002337, doi: 10.1109/PVSC.2010.5614439.

[31] T. Dierauf, A. Growitz, S. Kurtz, J. L. B. Cruz, E. Riley, and C. Hansen, “Weather-Corrected Performance Ratio,” Golden, CO (United States), Apr. 2013. doi: 10.2172/1078057.

[32] A. Phinikarides, G. Makrides, N. Kindyni, and G. E. Georghiou, “Comparison of trend extraction methods for calculating performance loss rates of different photovoltaic technologies,” in *2014 IEEE 40th Photovoltaic Specialist Conference (PVSC)*, Jun. 2014, pp. 3211–3215, doi: 10.1109/PVSC.2014.6925619.

[33] A. Frick, G. Makrides, M. Schubert, M. Schlecht, and G. E. Georghiou, “Degradation Rate Location Dependency of Photovoltaic Systems,” pp. 1–20, 2020.

[34] E. Hasselbrink *et al.*, “Validation of the PVLife model using 3 million module-years of live site data,” in *2013 IEEE 39th Photovoltaic Specialists Conference (PVSC)*, Jun. 2013, pp. 0007–0012, doi: 10.1109/PVSC.2013.6744087.

[35] D. C. Jordan, C. Deline, S. R. Kurtz, G. M. Kimball, and M. Anderson, “Robust PV Degradation Methodology and Application,” *IEEE J. Photovoltaics*, vol. 8, no. 2, pp. 525–531, Mar. 2018, doi: 10.1109/JPHOTOV.2017.2779779.

[36] D. M. Miller and D. Williams, “Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy,” *Int. J. Forecast.*, vol. 19, no. 4, pp. 669–684, Oct. 2003, doi: 10.1016/S0169-2070(02)00077-8.

[37] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis Forecasting and Control*, 5th ed. John Wiley & Sons Inc., Hoboken, New Jersey, 2015.

[38] G. Makrides, B. Zinsser, M. Schubert, and G. E. Georghiou, “Performance loss rate of twelve photovoltaic technologies under field conditions using statistical techniques,” *Sol. Energy*, vol. 103, no. May, pp. 28–42, 2014, doi: 10.1016/j.solener.2014.02.011.

[39] A. Phinikarides, G. Makrides, B. Zinsser, M. Schubert, and G. E. Georghiou, “Analysis of photovoltaic system performance time series: Seasonality and performance loss,” *Renew. Energy*, vol. 77, pp. 51–63, 2015, doi: 10.1016/j.renene.2014.11.091.

[40] A. Phinikarides, G. Makrides, and G. E. Georghiou, “Estimation of the Degradation Rate of Fielded Photovoltaic Arrays in the Presence of Measurement Outages,” in *32nd European PV Solar Energy Conference and Exhibition*, pp. 1754–1757.

[41] A. Kyprianou, A. Phinikarides, G. Makrides, and G. E. Georghiou, “Definition and Computation of the Degradation Rates of Photovoltaic Systems of Different Technologies with Robust Principal Component Analysis,” *IEEE J. Photovoltaics*, vol. 5, no. 6, pp. 1698–1705, 2015, doi: 10.1109/JPHOTOV.2015.2478065.

[42] N. B. Erichson, S. Voronin, S. L. Brunton, and J. N. Kutz, “Randomized Matrix Decompositions Using R,” *J. Stat. Softw.*, vol. 89, no. 11, 2019, doi: https://doi.org/10.18637/jss.v089.i11.

[43] Z. Lin, M. Chen, and Y. Ma, “The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices,” vol. 3, 2013, doi: https://arxiv.org/abs/1009.5055v3.

[44] Rob J. Hyndman and Yeasmin Khandakar, “Automatic Time Series Forecasting: The forecast Package for R,” *J. Stat. Softw.*, vol. 27, no. 3, p. 22, 2008, [Online]. Available: http://www.jstatsoft.org/%0Ahttp://www.jstatsoft.org/v27/i03/paper.

[45] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, 2nd ed. Melbourne, Australia: OTexts, 2018.

# *Appendix A – Results of Simulations*

## A.1 Results of SARIMA models based on the AIC information criterion

**Table A.1. Identified SARIMA models for the PV grid connected systems based on the AIC value. (*Partially shaded systems)**

| a/a | Manufacturer | $(p,d,q),(P,D,Q)_s$ | RMSE (%) | MAE (%) | AIC | AICc | BIC |
|---|---|---|---|---|---|---|---|
| 01 | Solon * | $(1,0,0),(1,1,0)_{12}$ | 4.07 | 3.23 | -222.33 | -221.40 | -214.85 |
| 02 | Sanyo | $(0,1,1),(1,1,1)_{12}$ | 2.52 | 2.11 | -251.09 | -250.13 | -243.69 |
| 03 | Atersa | $(0,0,0),(1,1,0)_{12}$ | 2.69 | 2.11 | -246.87 | -246.32 | -241.25 |
| 04 | Suntechnics | $(0,0,0),(1,1,0)_{12}$ | 2.47 | 2.00 | -238.11 | -237.57 | -232.50 |
| 05 | Schott Solar (EGF) | $(0,0,0),(1,1,1)_{12}$ | 1.47 | 1.25 | -262.09 | -261.16 | -254.61 |
| 06 | BP Solar * | $(0,1,2),(1,1,0)_{12}$ | 5.82 | 5.00 | -194.92 | -193.97 | -187.52 |
| 07 | SolarWorld | $(0,0,0),(1,1,1)_{12}$ | 2.22 | 1.84 | -255.64 | -254.71 | -248.16 |
| 08 | Schott Solar (MAIN) | $(0,0,0),(1,1,1)_{12}$ | 1.48 | 1.18 | -259.47 | -258.54 | -251.99 |
| 09 | Würth Solar | $(1,0,0),(1,1,0)_{12}$ | 2.34 | 1.99 | -247.77 | -246.84 | -240.29 |
| 10 | First Solar | $(1,1,0),(1,1,0)_{12}$ | 1.63 | 1.26 | -275.45 | -274.90 | -269.90 |
| 11 | MHI | $(0,1,1),(1,1,0)_{12}$ | 1.93 | 1.62 | -246.89 | -246.33 | -241.34 |

**Table A.2. RMSE and MAE between the actual and forecasted (using the SARIMA models based on the AIC) PR values after the application of the RPCA methodology for the 6th, 7th and 8th year. (*Partially shaded systems).**

| | | 6th year | | 7th year | | 8th year | |
|---|---|---|---|---|---|---|---|
| a/a | Manufacturer | RMSE (%) | MAE (%) | RMSE (%) | MAE (%) | RMSE (%) | MAE (%) |
| 01 | Solon * | 1.86 | 1.70 | 3.06 | 2.35 | 2.13 | 1.66 |
| 02 | Sanyo | 1.10 | 0.94 | 1.78 | 1.73 | 2.02 | 1.98 |
| 03 | Atersa | 2.51 | 2.35 | 2.86 | 2.70 | 1.13 | 0.90 |
| 04 | Suntechnics | 1.12 | 0.97 | 1.83 | 1.62 | 1.77 | 1.51 |
| 05 | Schott Solar (EGF) | 1.00 | 0.83 | 1.59 | 1.50 | 0.56 | 0.45 |
| 06 | BP Solar * | 2.74 | 2.16 | 6.01 | 5.69 | 7.87 | 7.52 |
| 07 | SolarWorld | 1.57 | 1.35 | 0.95 | 0.70 | 1.88 | 1.73 |
| 08 | Schott Solar (MAIN) | 0.40 | 0.36 | 1.71 | 1.66 | 0.50 | 0.38 |
| 09 | Würth Solar | 1.07 | 0.88 | 2.12 | 1.90 | 2.82 | 2.61 |
| 10 | First Solar | 1.40 | 1.27 | 1.90 | 1.83 | 0.69 | 0.54 |
| 11 | MHI | 1.13 | 0.96 | 2.31 | 2.19 | 1.43 | 1.31 |

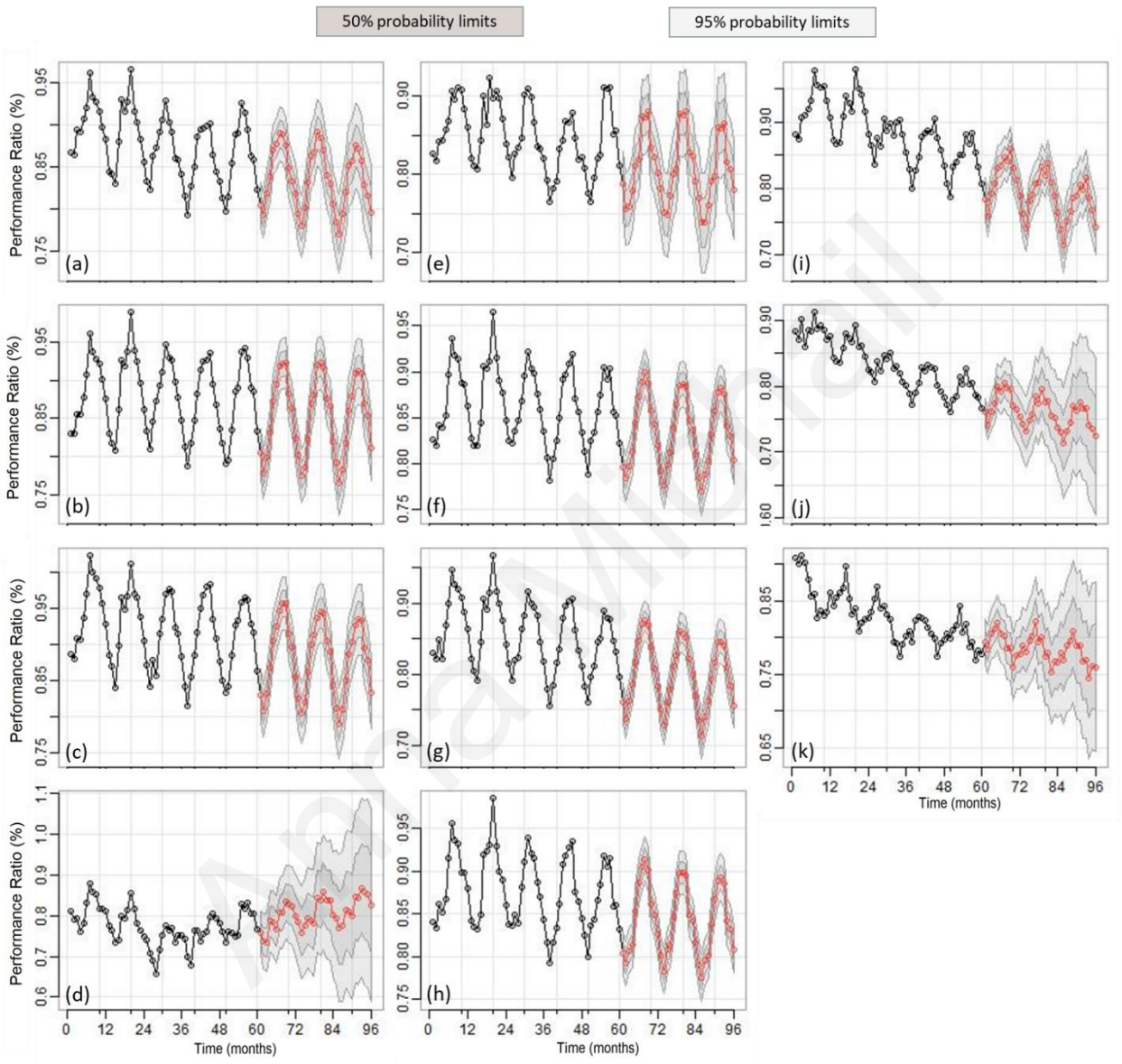


**Figure A.1. Upper and lower 50% and 95% probability limits using the SARIMA models based on the AIC. Mono-c-Si systems: (a) Sanyo (b) Atersa (c) Suntechnics and (d) BP Solar*. Multi-c-Si systems: (e) Solon*(f) Schott Solar (EGF) (g) Solar World and (h) Schott Solar (MAIN). Thin film systems: (i) Würth Solar (j) First Solar (EGF) and (k) MHI. (* Partially shaded systems)**

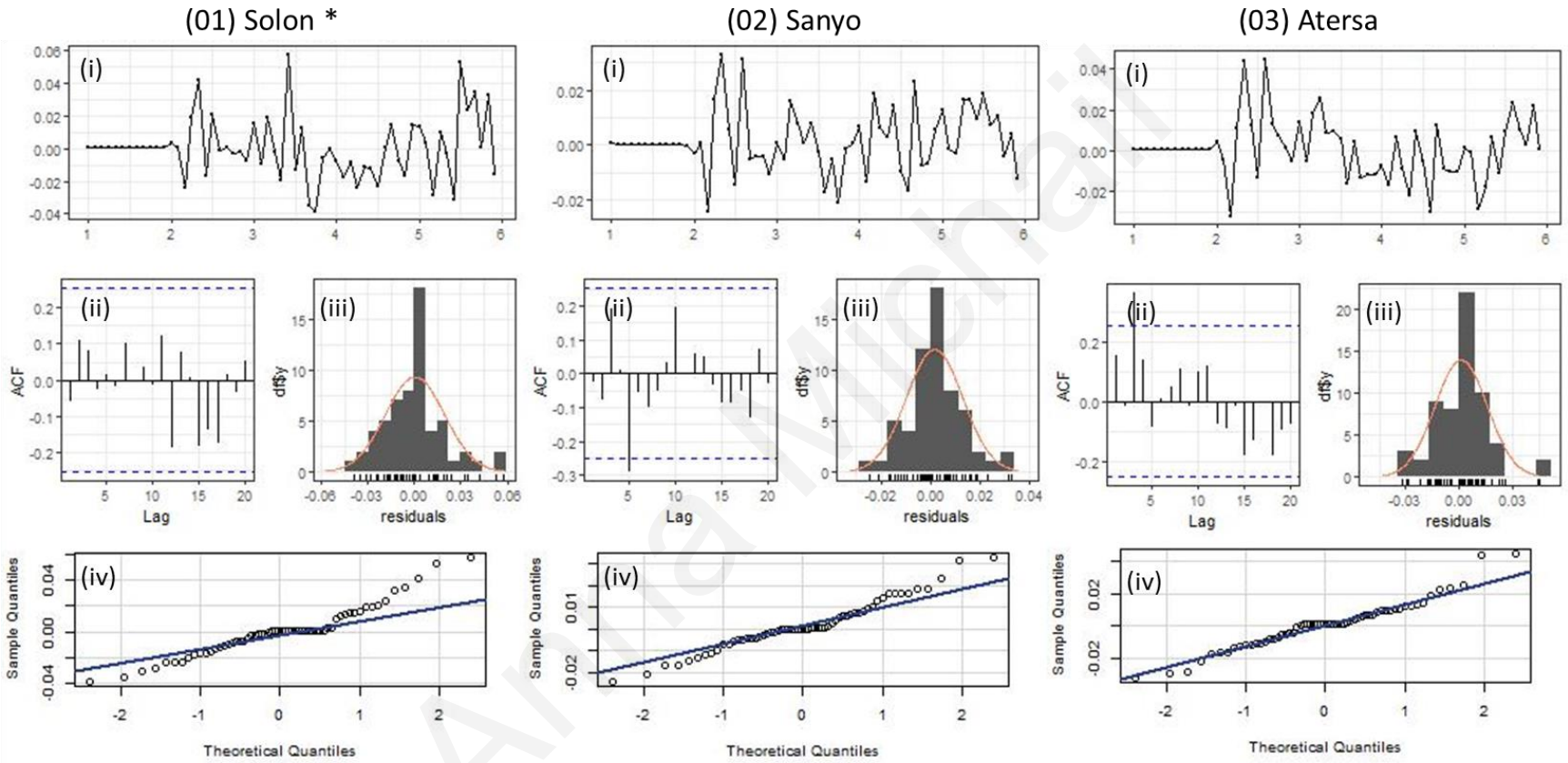


**Figure A.2. (i) Time plot, (ii) ACF, (iii) histogram of the residuals and (iv) Q-Q plot that displays the sample residuals with "o" and the theoretical-normal normal quantiles with a solid line obtained using the AIC information criterion for identifying the SARIMA model for the 1st, 2nd and 3rd PV system. (*Partially shaded systems)**

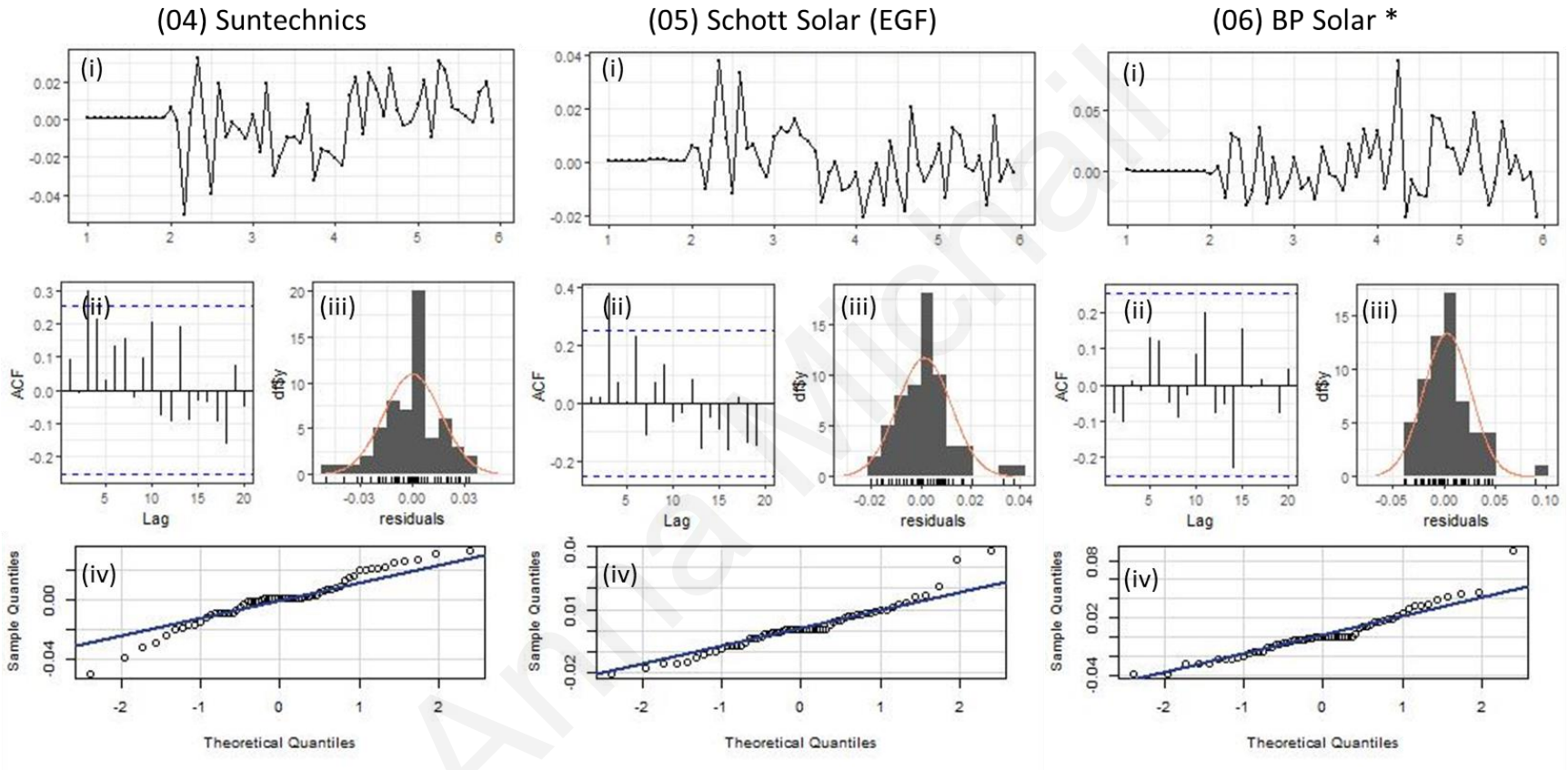


**Figure A.3. (i) Time plot, (ii) ACF, (iii) histogram of the residuals and (iv) Q-Q plot that displays the sample residuals with "o" and the theoretical-normal normal quantiles with a solid line obtained using the AIC information criterion for identifying the SARIMA model for the 4th, 5th and 6th PV system. (*Partially shaded systems)**

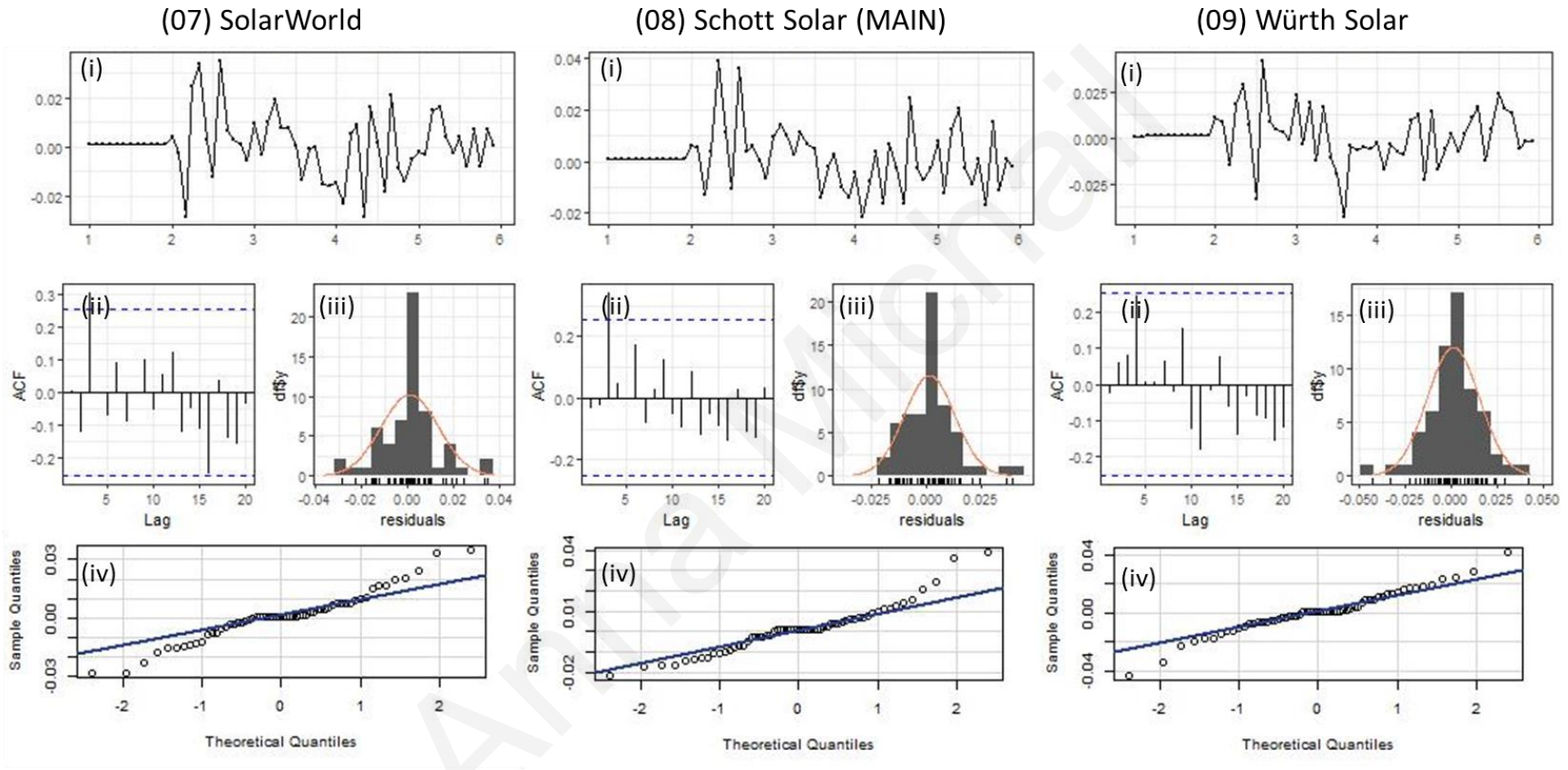


**Figure A.4. (i) Time plot, (ii) ACF, (iii) histogram of the residuals and (iv) Q-Q plot that displays the sample residuals with "o" and the theoretical-normal normal quantiles with a solid line obtained using the AIC information criterion for identifying the SARIMA model for the 7th, 8th and 9th PV system.**

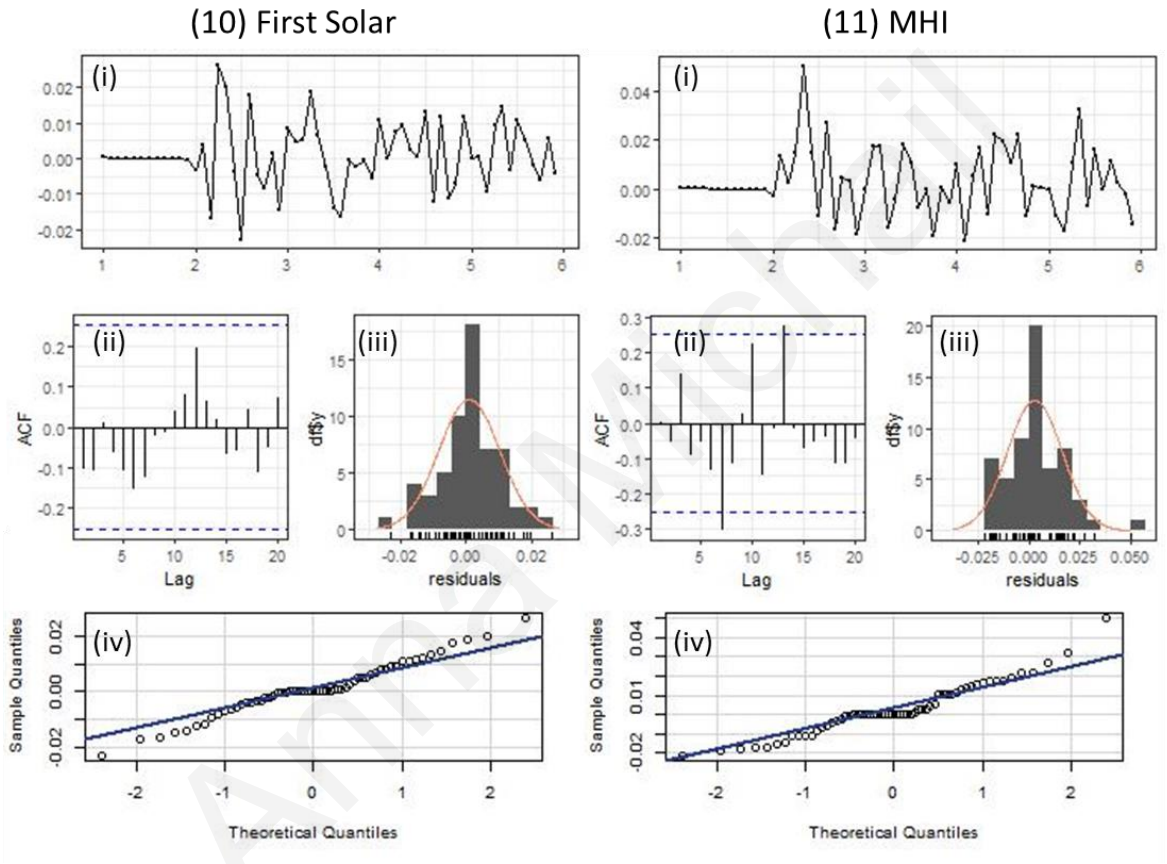


**Figure A.5. (i) Time plot, (ii) ACF, (iii) histogram of the residuals and (iv) Q-Q plot that displays the sample residuals with "o" and the theoretical-normal normal quantiles with a solid line obtained using the AIC information criterion for identifying the SARIMA model for the 10$^{th}$ and 11$^{th}$ PV system.**

**Table A.3. Actual and forecast performance loss rate per PV system as calculated using the SARIMA models based on the AIC for each system and then the RPCA methodology. (*Partially shaded systems)**

| | | 6th year | | | 7th year | | | 8th year | | |
|---|---|---|---|---|---|---|---|---|---|---|
| a/a | Manufacturer | Actual (%/yr) | Forecast (%/yr) | Dif. (%/yr) | Actual (%/yr) | Forecast (%/yr) | Dif. (%/yr) | Actual (%/yr) | Forecast (%/yr) | Dif. (%/yr) |
| 01 | Solon * | 0.69 | 0.85 | 0.16 | 1.30 | 0.82 | -0.49 | 1.11 | 0.89 | -0.21 |
| 02 | Sanyo | 0.83 | 0.97 | 0.14 | 0.71 | 0.96 | 0.25 | 0.73 | 0.99 | 0.26 |
| 03 | Atersa | 0.90 | 0.39 | -0.51 | 0.88 | 0.40 | -0.49 | 0.47 | 0.46 | -0.02 |
| 04 | Suntechnics | 0.81 | 0.81 | 0.00 | 0.65 | 0.84 | 0.19 | 0.76 | 0.88 | 0.13 |
| 05 | Schott Solar (EGF) | 0.38 | 0.50 | 0.12 | 0.79 | 0.53 | -0.26 | 0.60 | 0.57 | -0.03 |
| 06 | BP Solar * | 0.73 | 0.46 | -0.26 | 1.09 | 0.07 | -1.02 | 1.03 | 0.15 | -0.88 |
| 07 | SolarWorld | 0.87 | 1.18 | 0.32 | 1.13 | 1.18 | 0.05 | 0.97 | 1.25 | 0.28 |
| 08 | Schott Solar (MAIN) | 0.68 | 0.63 | -0.04 | 0.97 | 0.67 | -0.29 | 0.78 | 0.71 | -0.07 |
| 09 | Würth Solar | 1.97 | 1.88 | -0.10 | 2.26 | 1.93 | -0.33 | 2.39 | 2.00 | -0.39 |
| 10 | First Solar | 2.11 | 1.86 | -0.25 | 2.14 | 1.85 | -0.30 | 1.89 | 1.82 | -0.07 |
| 11 | MHI | 1.39 | 1.15 | -0.24 | 1.58 | 1.11 | -0.47 | 1.37 | 1.09 | -0.28 |

**Table A.4. Maximum, minimum, mean and median value of the forecast performance loss rate per PV technology and separated for the two partially shaded PV system using the SARIMA models based on the AIC.**

| | | Forecast PLR (%/yr) | | | |
|---|---|---|---|---|---|
| a/a | PV technology | Max | Min | Mean | Median |
| 01 | Mono-c-Si | 0.99 | 0.39 | 0.75 | 0.84 |
| 02 | Multi-c-Si | 1.25 | 0.50 | 0.80 | 0.67 |
| 03 | Thin-Film | 2.00 | 1.09 | 1.63 | 1.85 |
| 04 | Partially shaded PV systems | 0.89 | 0.07 | 0.54 | 0.64 |
| 05 | For all the PV systems | 2.00 | 0.07 | 0.97 | 0.88 |

**Table A.5. Maximum, minimum, mean and median value of the absolute difference between the actual and forecast performance loss rate per PV technology and separated for the two partially shaded PV system using the SARIMA models based on the AIC.**

| | | Absolute Difference (%/yr) | | | |
|---|---|---|---|---|---|
| a/a | PV technology | Max | Min | Mean | Median |
| 01 | Mono-c-Si | 0.51 | 0.00 | 0.22 | 0.19 |
| 02 | Multi-c-Si | 0.32 | 0.03 | 0.16 | 0.12 |
| 03 | Thin-Film | 0.47 | 0.07 | 0.27 | 0.28 |
| 04 | Partially shaded PV systems | 1.02 | 0.16 | 0.50 | 0.37 |
| 05 | For all the PV systems | 1.02 | 0.00 | 0.27 | 0.25 |

## A.2 Results of SARIMA models based on the AICc information criterion

**Table A.6. Identified SARIMA models for the PV grid connected systems based on the AICc value. (*Partially shaded systems)**

| a/a | Manufacturer | $(p,d,q),(P,D,Q)_s$ | RMSE (%) | MAE (%) | AIC | AICc | BIC |
|---|---|---|---|---|---|---|---|
| 01 | Solon * | $(1,0,0),(1,1,0)_{12}$ | 4.07 | 3.23 | -222.33 | -221.40 | -214.85 |
| 02 | Sanyo | $(0,1,1),(1,1,1)_{12}$ | 2.52 | 2.11 | -251.09 | -250.13 | -243.69 |
| 03 | Atersa | $(0,0,0),(1,1,0)_{12}$ | 2.69 | 2.11 | -246.87 | -246.32 | -241.25 |
| 04 | Suntechnics | $(0,0,0),(1,1,0)_{12}$ | 2.47 | 2.00 | -238.11 | -237.57 | -232.50 |
| 05 | Schott Solar (EGF) | $(0,0,0),(1,1,1)_{12}$ | 1.47 | 1.25 | -262.09 | -261.16 | -254.61 |
| 06 | BP Solar * | $(2,1,0),(1,1,0)_{12}$ | 4.76 | 4.01 | -193.45 | -192.50 | -186.05 |
| 07 | SolarWorld | $(0,0,0),(1,1,1)_{12}$ | 2.22 | 1.84 | -255.64 | -254.71 | -248.16 |
| 08 | Schott Solar (MAIN) | $(0,0,0),(1,1,1)_{12}$ | 1.48 | 1.18 | -259.47 | -258.54 | -251.99 |
| 09 | Würth Solar | $(0,0,0),(1,1,0)_{12}$ | 2.33 | 1.99 | -247.63 | -247.08 | -242.02 |
| 10 | First Solar | $(1,1,0),(1,1,0)_{12}$ | 1.63 | 1.26 | -275.45 | -274.90 | -269.90 |
| 11 | MHI | $(0,1,1),(1,1,0)_{12}$ | 1.93 | 1.62 | -246.89 | -246.33 | -241.34 |

**Table A.7. RMSE and MAE between the actual and forecasted (using the SARIMA models based on the AICc) PR values after the application of the RPCA methodology for the 6th, 7th and 8th year. (*Partially shaded systems).**

| | | 6th year | | 7th year | | 8th year | |
|---|---|---|---|---|---|---|---|
| a/a | Manufacturer | RMSE (%) | MAE (%) | RMSE (%) | MAE (%) | RMSE (%) | MAE (%) |
| 01 | Solon * | 1.86 | 1.70 | 3.06 | 2.35 | 2.13 | 1.66 |
| 02 | Sanyo | 1.10 | 0.94 | 1.78 | 1.73 | 2.02 | 1.98 |
| 03 | Atersa | 2.51 | 2.35 | 2.86 | 2.70 | 1.13 | 0.90 |
| 04 | Suntechnics | 1.12 | 0.97 | 1.83 | 1.62 | 1.77 | 1.51 |
| 05 | Schott Solar (EGF) | 1.00 | 0.83 | 1.59 | 1.50 | 0.56 | 0.45 |
| 06 | BP Solar * | 2.43 | 1.91 | 4.95 | 4.56 | 6.23 | 5.80 |
| 07 | SolarWorld | 1.57 | 1.35 | 0.95 | 0.70 | 1.88 | 1.73 |
| 08 | Schott Solar (MAIN) | 0.40 | 0.36 | 1.71 | 1.66 | 0.50 | 0.38 |
| 09 | Würth Solar | 1.06 | 0.88 | 2.11 | 1.90 | 2.80 | 2.59 |
| 10 | First Solar | 1.40 | 1.27 | 1.90 | 1.83 | 0.69 | 0.54 |
| 11 | MHI | 1.13 | 0.96 | 2.31 | 2.19 | 1.43 | 1.31 |

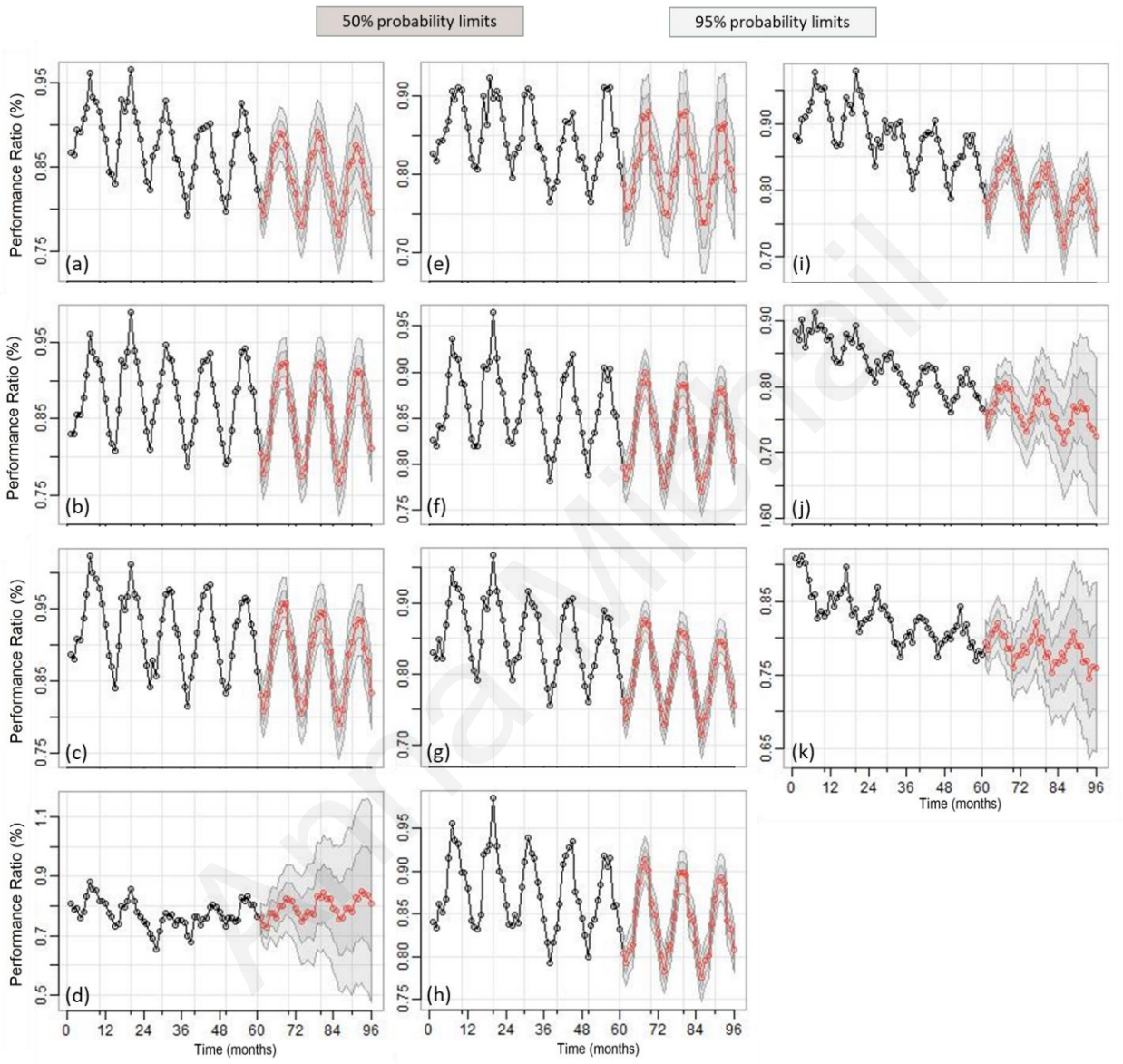


**Figure A.6. Upper and lower 50% and 95% probability limits using the SARIMA models based on the AICc. Mono-c-Si systems: (a) Sanyo (b) Atersa (c) Suntechnics and (d) BP Solar*. Multi-c-Si systems: (e) Solon*(f) Schott Solar (EGF) (g) Solar World and (h) Schott Solar (MAIN). Thin film systems: (i) Würth Solar (j) First Solar (EGF) and (k) MHI. (* Partially shaded systems)**

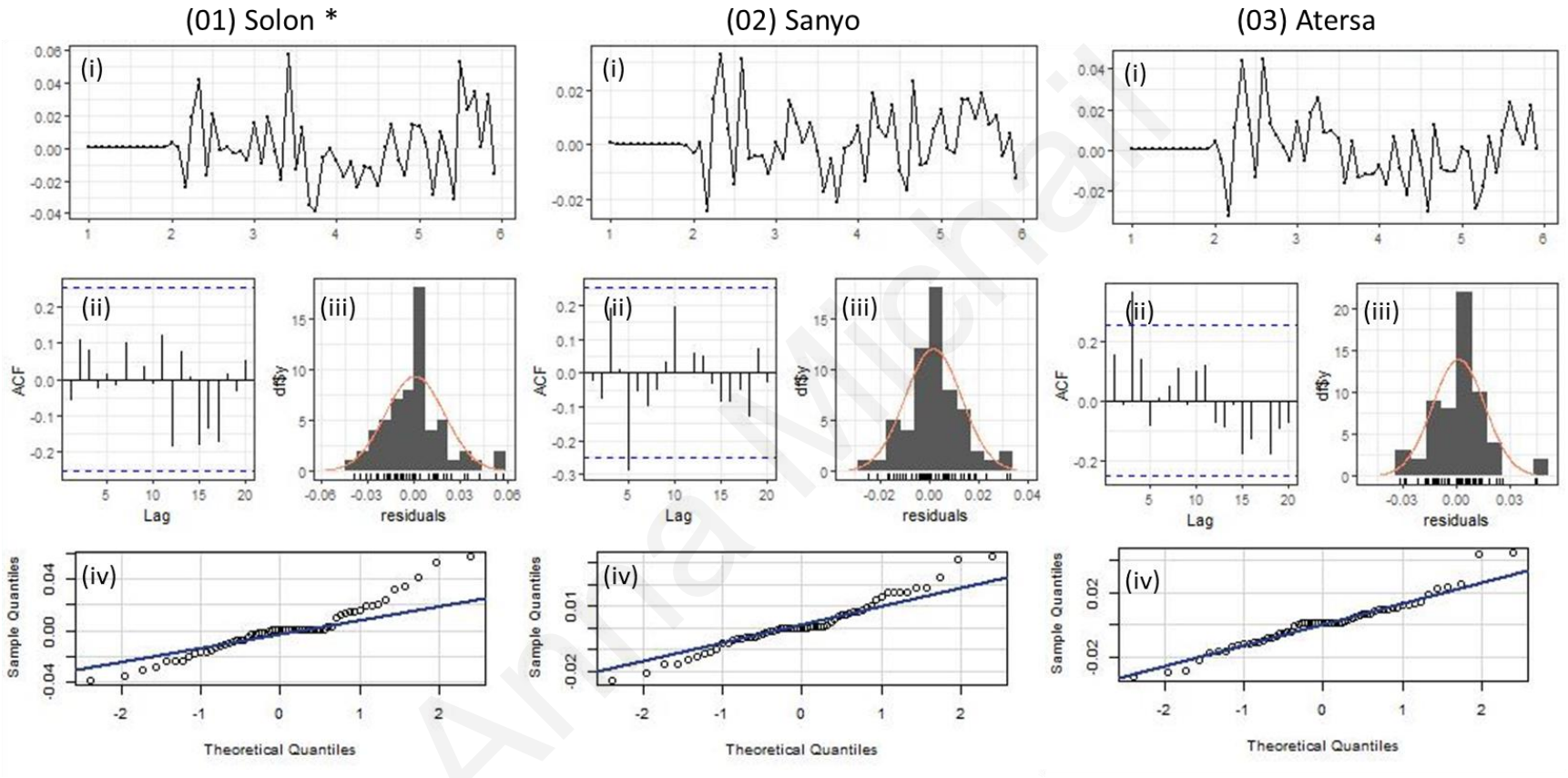


**Figure A.7. (i) Time plot, (ii) ACF, (iii) histogram of the residuals and (iv) Q-Q plot that displays the sample residuals with "o" and the theoretical-normal normal quantiles with a solid line obtained using the AICc information criterion for identifying the SARIMA model for the 1st, 2nd and 3rd PV system. (*Partially shaded systems)**

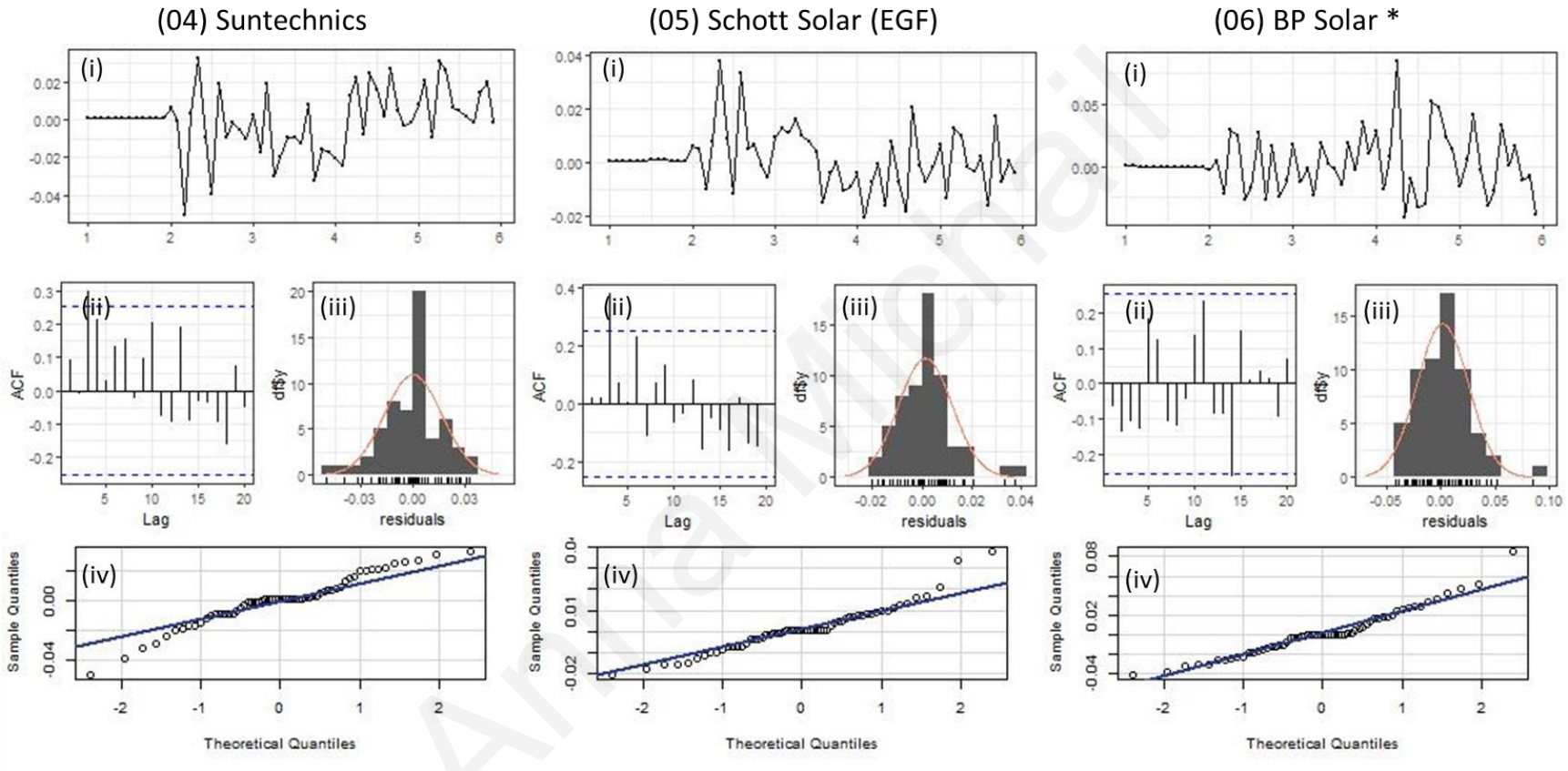


**Figure A.8. (i) Time plot, (ii) ACF, (iii) histogram of the residuals and (iv) Q-Q plot that displays the sample residuals with "o" and the theoretical-normal normal quantiles with a solid line obtained using the AICc information criterion for identifying the SARIMA model for the 4$^{th}$, 5$^{th}$ and 6$^{th}$ PV system. (*Partially shaded systems)**

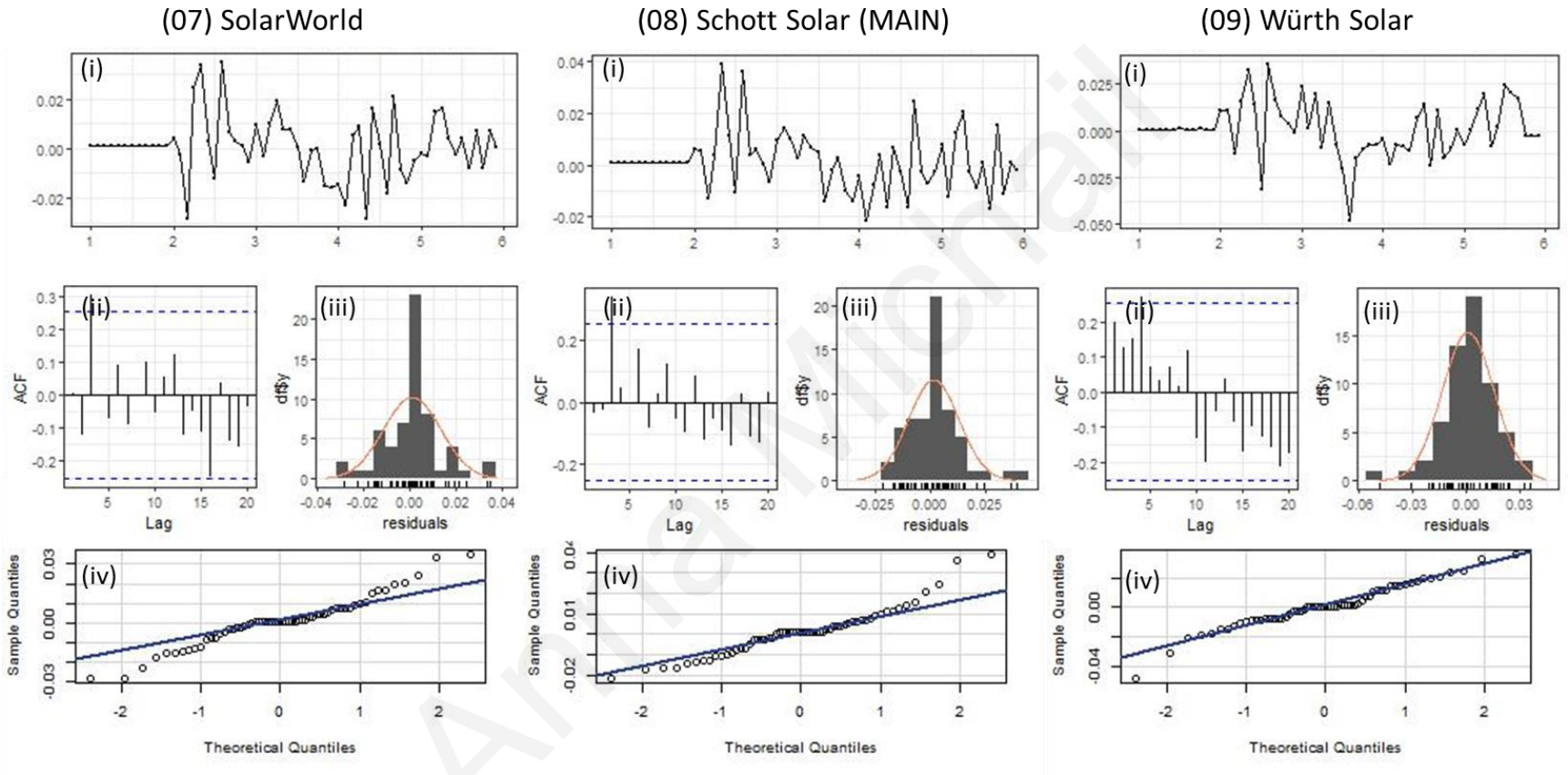


**Figure A.9. (i) Time plot, (ii) ACF, (iii) histogram of the residuals and (iv) Q-Q plot that displays the sample residuals with "o" and the theoretical-normal normal quantiles with a solid line obtained using the AICc information criterion for identifying the SARIMA model for the 7th, 8th and 9th PV system.**

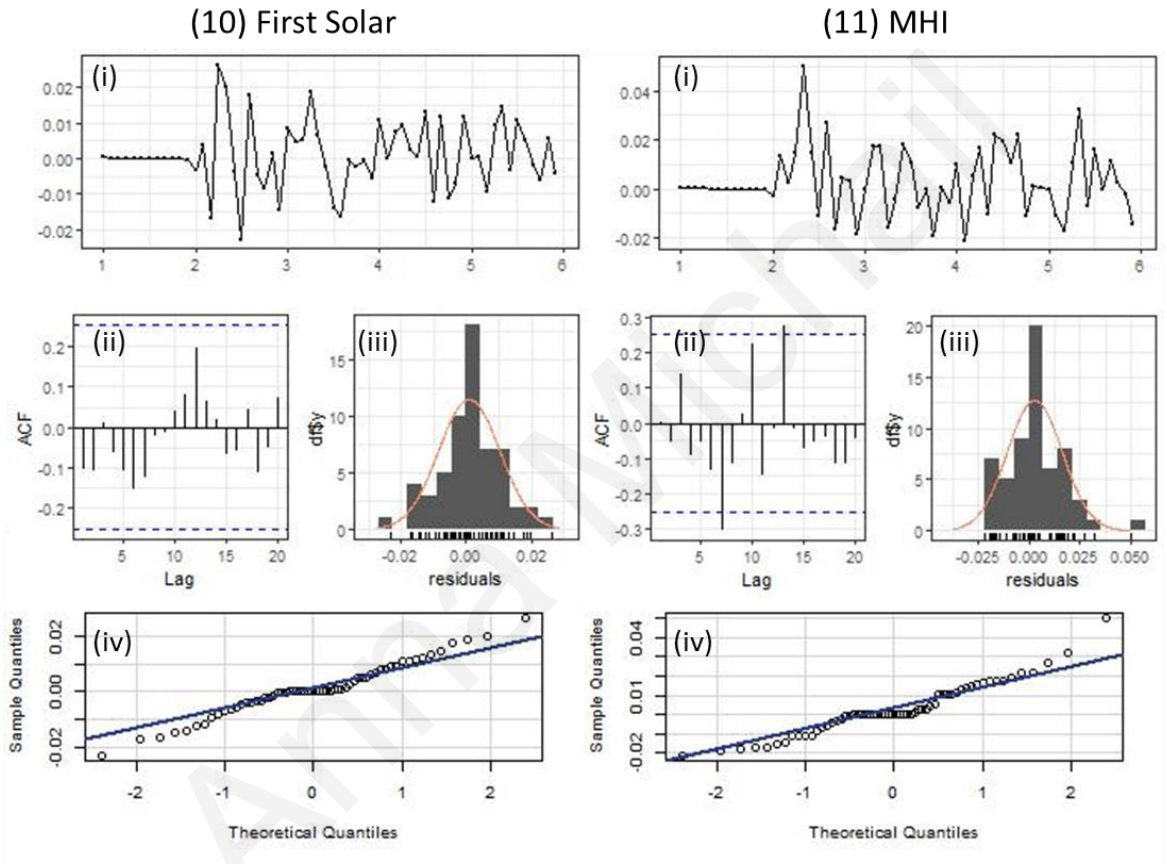


**Figure A.10. (i) Time plot, (ii) ACF, (iii) histogram of the residuals and (iv) Q-Q plot that displays the sample residuals with "o" and the theoretical-normal normal quantiles with a solid line obtained using the AICc information criterion for identifying the SARIMA model for the 10$^{th}$ and 11$^{th}$ PV system.**

**Table A.8. Actual and forecast performance loss rate per PV system as calculated using the SARIMA models based on the AICc for each system and then the RPCA methodology. (*Partially shaded systems)**

| | | 6th year | | | 7th year | | | 8th year | | |
|---|---|---|---|---|---|---|---|---|---|---|
| a/a | Manufacturer | Actual (%/yr) | Forecast (%/yr) | Dif. (%/yr) | Actual (%/yr) | Forecast (%/yr) | Dif. (%/yr) | Actual (%/yr) | Forecast (%/yr) | Dif. (%/yr) |
| 01 | Solon * | 0.69 | 0.85 | 0.16 | 1.30 | 0.82 | -0.49 | 1.11 | 0.89 | -0.21 |
| 02 | Sanyo | 0.83 | 0.97 | 0.14 | 0.71 | 0.96 | 0.25 | 0.73 | 0.99 | 0.26 |
| 03 | Atersa | 0.90 | 0.39 | -0.51 | 0.88 | 0.40 | -0.49 | 0.47 | 0.46 | -0.02 |
| 04 | Suntechnics | 0.81 | 0.81 | 0.00 | 0.65 | 0.84 | 0.19 | 0.76 | 0.88 | 0.13 |
| 05 | Schott Solar (EGF) | 0.38 | 0.50 | 0.12 | 0.79 | 0.53 | -0.26 | 0.60 | 0.57 | -0.03 |
| 06 | BP Solar * | 0.73 | 0.62 | -0.11 | 1.09 | 0.26 | -0.83 | 1.03 | 0.12 | -0.91 |
| 07 | SolarWorld | 0.87 | 1.18 | 0.32 | 1.13 | 1.18 | 0.05 | 0.97 | 1.25 | 0.28 |
| 08 | Schott Solar (MAIN) | 0.68 | 0.63 | -0.04 | 0.97 | 0.67 | -0.29 | 0.78 | 0.71 | -0.07 |
| 09 | Würth Solar | 1.97 | 1.88 | -0.10 | 2.26 | 1.94 | -0.33 | 2.39 | 2.00 | -0.38 |
| 10 | First Solar | 2.11 | 1.86 | -0.25 | 2.14 | 1.85 | -0.30 | 1.89 | 1.82 | -0.07 |
| 11 | MHI | 1.39 | 1.15 | -0.24 | 1.58 | 1.11 | -0.47 | 1.37 | 1.09 | -0.28 |

**Table A.9. Maximum, minimum, mean and median value of the forecast performance loss rate per PV technology and separated for the two partially shaded PV system using the SARIMA models based on the AICc.**

| | | Forecast PLR (%/yr) | | | |
|---|---|---|---|---|---|
| a/a | PV technology | Max | Min | Mean | Median |
| 01 | Mono-c-Si | 0.99 | 0.39 | 0.75 | 0.84 |
| 02 | Multi-c-Si | 1.25 | 0.50 | 0.80 | 0.67 |
| 03 | Thin-Film | 2.00 | 1.09 | 1.63 | 1.85 |
| 04 | Partially shaded PV systems | 0.89 | 0.12 | 0.59 | 0.72 |
| 05 | For all the PV systems | 2.00 | 0.12 | 0.97 | 0.88 |

**Table A.10. Maximum, minimum, mean and median value of the absolute difference between the actual and forecast performance loss rate per PV technology and separated for the two partially shaded PV system using the SARIMA models based on the AICc.**

| | | Absolute Difference (%/yr) | | | |
|---|---|---|---|---|---|
| a/a | PV technology | Max | Min | Mean | Median |
| 01 | Mono-c-Si | 0.51 | 0.00 | 0.22 | 0.19 |
| 02 | Multi-c-Si | 0.32 | 0.03 | 0.16 | 0.12 |
| 03 | Thin-Film | 0.47 | 0.07 | 0.27 | 0.28 |
| 04 | Partially shaded PV systems | 0.91 | 0.11 | 0.45 | 0.35 |
| 05 | For all the PV systems | 0.91 | 0.00 | 0.26 | 0.25 |

## A.3 Results of SARIMA models based on the BIC information criterion

**Table A.11. Identified SARIMA models for the PV grid connected systems based on the BIC value. (*Partially shaded systems)**

| a/a | Manufacturer | $(\boldsymbol{p},\boldsymbol{d},\boldsymbol{q}),(\boldsymbol{P},\boldsymbol{D},\boldsymbol{Q})_{\boldsymbol{s}}$ | RMSE (%) | MAE (%) | AIC | AICc | BIC |
|---|---|---|---|---|---|---|---|
| 01 | Solon * | $(1,0,0),(1,1,0)_{12}$ | 4.07 | 3.23 | -222.33 | -221.40 | -214.85 |
| 02 | Sanyo | $(0,1,1),(1,1,1)_{12}$ | 2.52 | 2.11 | -251.09 | -250.13 | -243.69 |
| 03 | Atersa | $(0,0,0),(1,1,0)_{12}$ | 2.69 | 2.11 | -246.87 | -246.32 | -241.25 |
| 04 | Suntechnics | $(0,0,0),(1,1,0)_{12}$ | 2.47 | 2.00 | -238.11 | -237.57 | -232.50 |
| 05 | Schott Solar (EGF) | $(0,0,0),(1,1,1)_{12}$ | 1.47 | 1.25 | -262.09 | -261.16 | -254.61 |
| 06 | BP Solar * | $(0,1,0),(1,1,0)_{12}$ | 3.37 | 2.75 | -188.62 | -188.35 | -184.92 |
| 07 | SolarWorld | $(0,0,0),(1,1,0)_{12}$ | 2.31 | 1.92 | -255.01 | -254.46 | -249.40 |
| 08 | Schott Solar (MAIN) | $(0,0,0),(1,1,0)_{12}$ | 1.70 | 1.35 | -257.62 | -257.08 | -252.01 |
| 09 | Würth Solar | $(0,0,0),(1,1,0)_{12}$ | 2.33 | 1.99 | -247.63 | -247.08 | -242.02 |
| 10 | First Solar | $(1,1,0),(1,1,0)_{12}$ | 1.63 | 1.26 | -275.45 | -274.90 | -269.90 |
| 11 | MHI | $(0,1,1),(1,1,0)_{12}$ | 1.93 | 1.62 | -246.89 | -246.33 | -241.34 |

**Table A.12. RMSE and MAE between the actual and forecasted (using the SARIMA models based on the BIC) PR values after the application of the RPCA methodology for the 6th, 7th and 8th year. (*Partially shaded systems).**

| | | 6th year | | 7th year | | 8th year | |
|---|---|---|---|---|---|---|---|
| a/a | Manufacturer | RMSE (%) | MAE (%) | RMSE (%) | MAE (%) | RMSE (%) | MAE (%) |
| 01 | Solon * | 1.86 | 1.70 | 3.06 | 2.35 | 2.13 | 1.66 |
| 02 | Sanyo | 1.10 | 0.94 | 1.78 | 1.73 | 2.02 | 1.98 |
| 03 | Atersa | 2.51 | 2.35 | 2.86 | 2.70 | 1.13 | 0.90 |
| 04 | Suntechnics | 1.12 | 0.97 | 1.83 | 1.62 | 1.77 | 1.51 |
| 05 | Schott Solar (EGF) | 1.00 | 0.83 | 1.59 | 1.50 | 0.56 | 0.45 |
| 06 | BP Solar * | 2.65 | 2.31 | 3.34 | 2.88 | 3.72 | 3.09 |
| 07 | SolarWorld | 1.69 | 1.42 | 1.08 | 0.81 | 1.96 | 1.80 |
| 08 | Schott Solar (MAIN) | 0.67 | 0.60 | 1.88 | 1.74 | 0.69 | 0.56 |
| 09 | Würth Solar | 1.06 | 0.88 | 2.11 | 1.90 | 2.80 | 2.59 |
| 10 | First Solar | 1.40 | 1.27 | 1.90 | 1.83 | 0.69 | 0.54 |
| 11 | MHI | 1.13 | 0.96 | 2.31 | 2.19 | 1.43 | 1.31 |

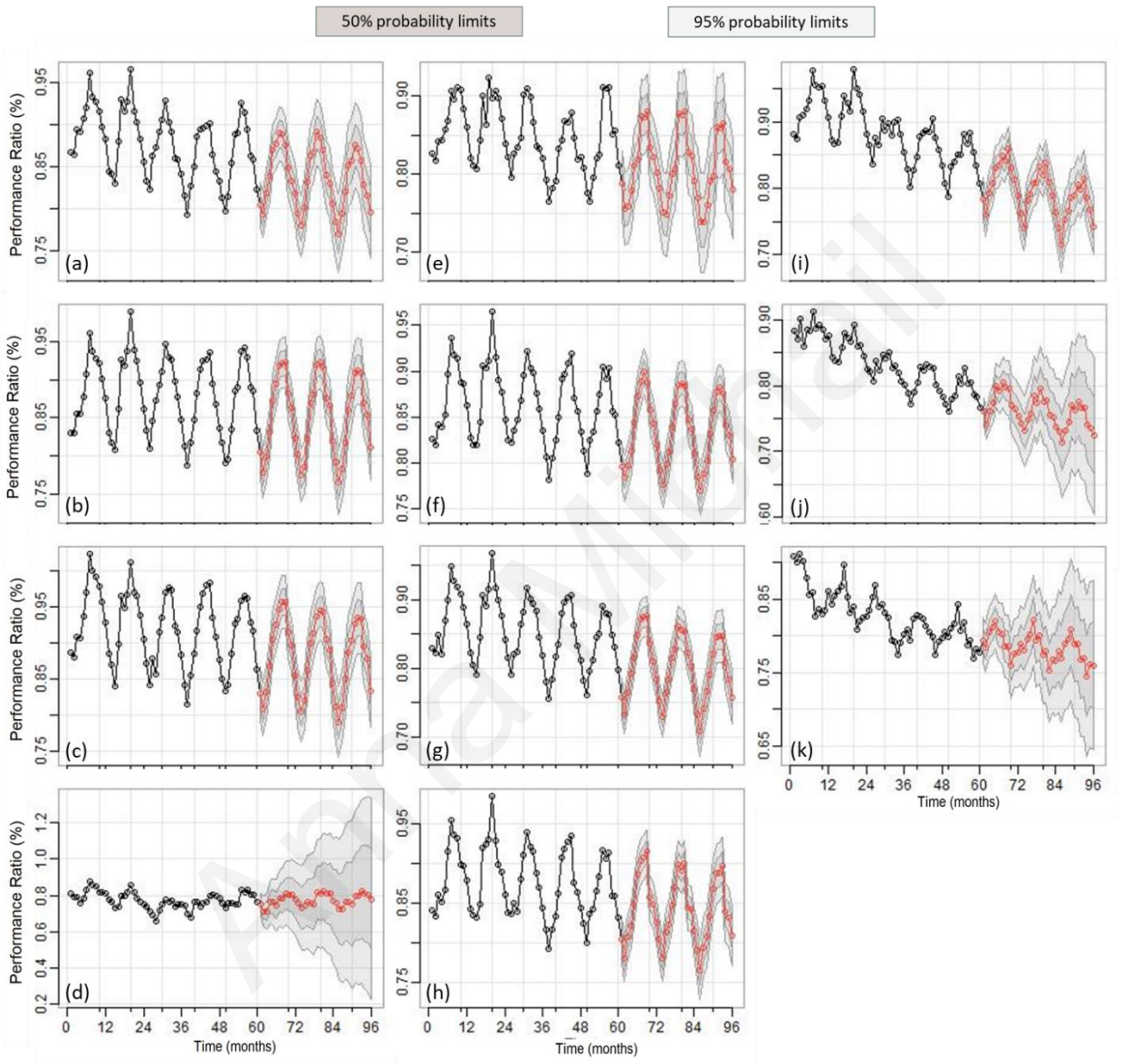


**Figure A.11. Upper and lower 50% and 95% probability limits using the SARIMA models based on the BIC. Mono-c-Si systems: (a) Sanyo (b) Atersa (c) Suntechnics and (d) BP Solar*. Multi-c-Si systems: (e) Solon*(f) Schott Solar (EGF) (g) Solar World and (h) Schott Solar (MAIN). Thin film systems: (i) Würth Solar (j) First Solar (EGF) and (k) MHI. (* Partially shaded systems)**

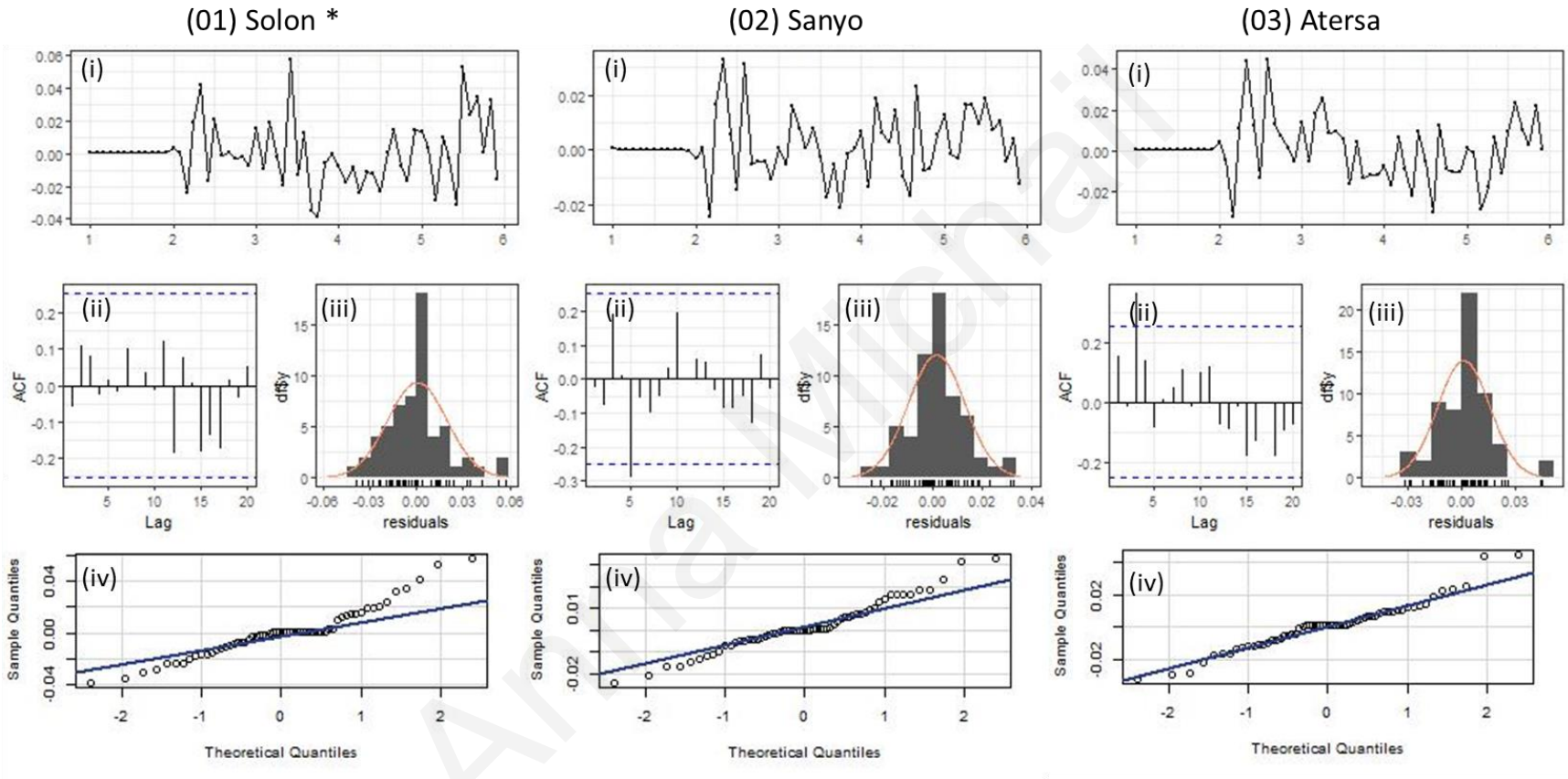


**Figure A.12. (i) Time plot, (ii) ACF, (iii) histogram of the residuals and (iv) Q-Q plot that displays the sample residuals with "o" and the theoretical-normal normal quantiles with a solid line obtained using the BIC information criterion for identifying the SARIMA model for the 1st, 2nd and 3rd PV system. (*Partially shaded systems)**

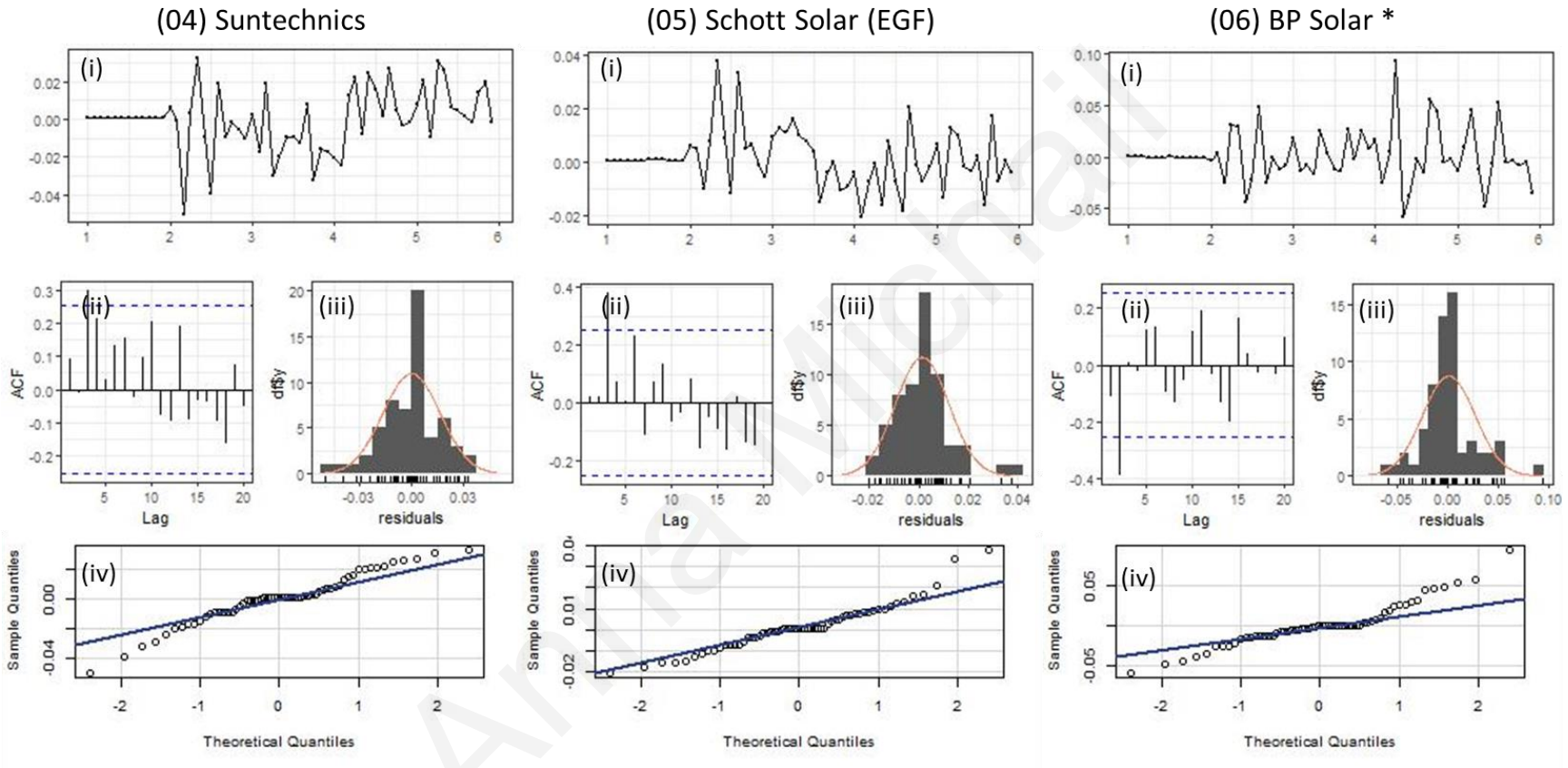


**Figure A.13. (i) Time plot, (ii) ACF, (iii) histogram of the residuals and (iv) Q-Q plot that displays the sample residuals with "o" and the theoretical-normal normal quantiles with a solid line obtained using the BIC information criterion for identifying the SARIMA model for the 4th, 5th and 6th PV system. (*Partially shaded systems)**

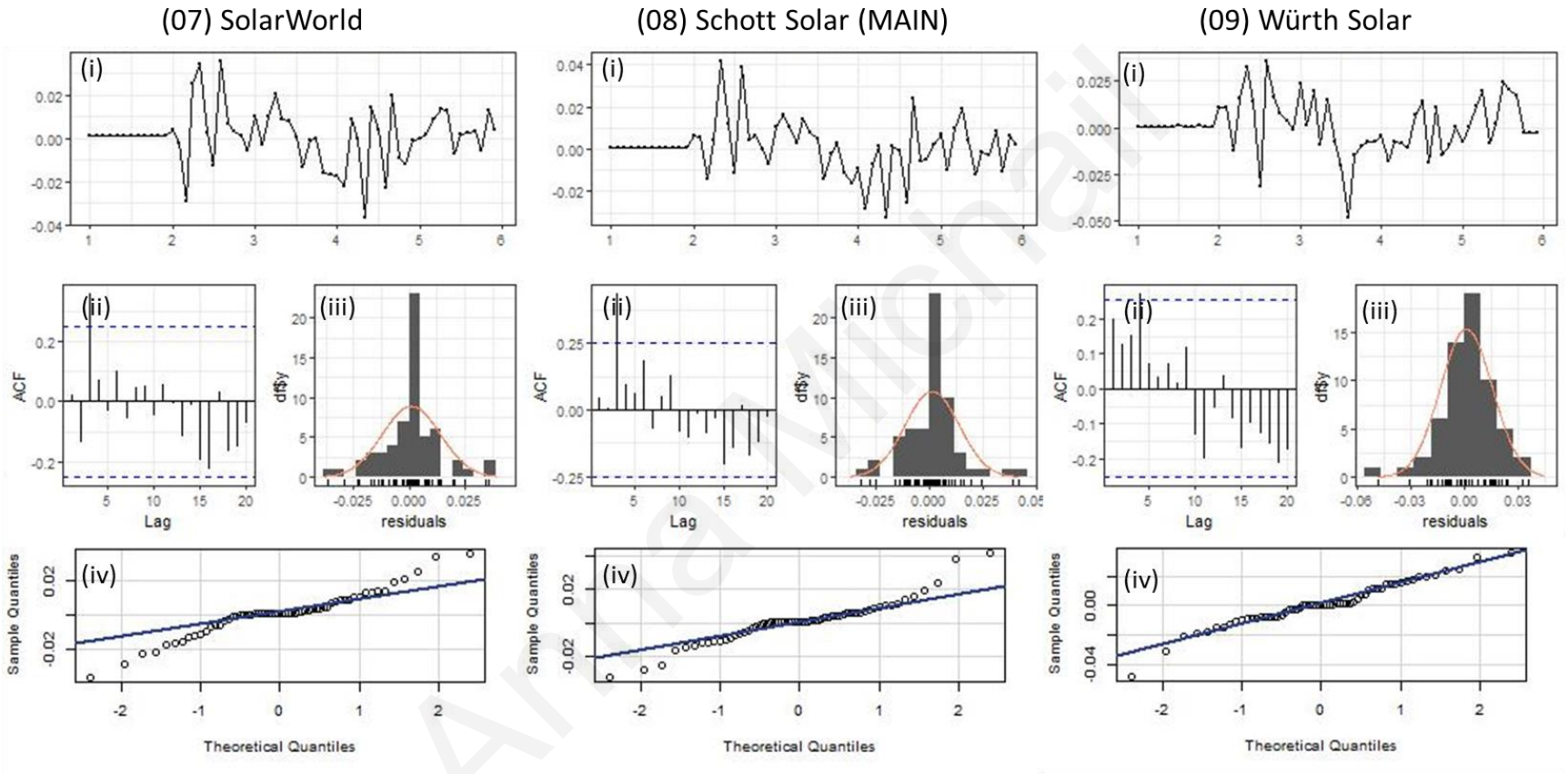


**Figure A.14. (i) Time plot, (ii) ACF, (iii) histogram of the residuals and (iv) Q-Q plot that displays the sample residuals with "o" and the theoretical-normal normal quantiles with a solid line obtained using the BIC information criterion for identifying the SARIMA model for the 7th, 8th and 9th PV system.**

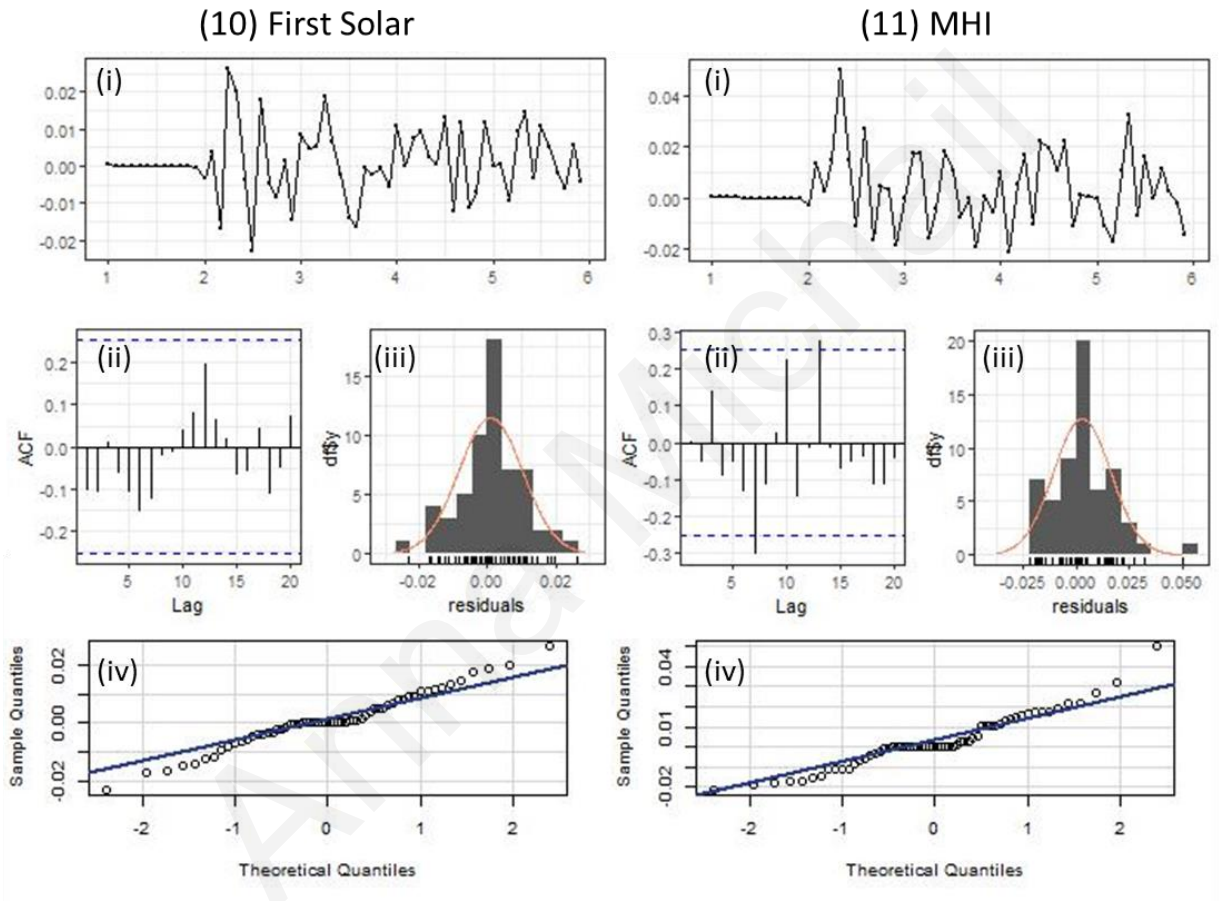


**Figure A.15. (i) Time plot, (ii) ACF, (iii) histogram of the residuals and (iv) Q-Q plot that displays the sample residuals with "o" and the theoretical-normal normal quantiles with a solid line obtained using the BIC information criterion for identifying the SARIMA model for the 10$^{th}$ and 11$^{th}$ PV system.**

**Table A.13. Actual and forecast performance loss rate per PV system as calculated using the SARIMA models based on the BIC for each system and then the RPCA methodology. (*Partially shaded systems)**

| | | 6th year | | | 7th year | | | 8th year | | |
|---|---|---|---|---|---|---|---|---|---|---|
| a/a | Manufacturer | Actual (%/yr) | Forecast (%/yr) | Dif. (%/yr) | Actual (%/yr) | Forecast (%/yr) | Dif. (%/yr) | Actual (%/yr) | Forecast (%/yr) | Dif. (%/yr) |
| 01 | Solon * | 0.69 | 0.85 | 0.16 | 1.30 | 0.82 | -0.49 | 1.11 | 0.89 | -0.21 |
| 02 | Sanyo | 0.83 | 0.97 | 0.14 | 0.71 | 0.96 | 0.25 | 0.73 | 0.99 | 0.26 |
| 03 | Atersa | 0.90 | 0.39 | -0.51 | 0.88 | 0.40 | -0.49 | 0.47 | 0.46 | -0.02 |
| 04 | Suntechnics | 0.81 | 0.81 | 0.00 | 0.65 | 0.84 | 0.19 | 0.76 | 0.88 | 0.13 |
| 05 | Schott Solar (EGF) | 0.38 | 0.50 | 0.12 | 0.79 | 0.53 | -0.26 | 0.60 | 0.57 | -0.03 |
| 06 | BP Solar * | 0.73 | 0.90 | 0.17 | 1.09 | 0.58 | -0.51 | 1.03 | 0.56 | -0.47 |
| 07 | SolarWorld | 0.87 | 1.20 | 0.33 | 1.13 | 1.18 | 0.05 | 0.97 | 1.26 | 0.29 |
| 08 | Schott Solar (MAIN) | 0.68 | 0.62 | -0.06 | 0.97 | 0.63 | -0.34 | 0.78 | 0.69 | -0.09 |
| 09 | Würth Solar | 1.97 | 1.88 | -0.10 | 2.26 | 1.94 | -0.33 | 2.39 | 2.00 | -0.38 |
| 10 | First Solar | 2.11 | 1.86 | -0.25 | 2.14 | 1.85 | -0.30 | 1.89 | 1.82 | -0.07 |
| 11 | MHI | 1.39 | 1.15 | -0.24 | 1.58 | 1.11 | -0.47 | 1.37 | 1.09 | -0.28 |

**Table A.14. Maximum, minimum, mean and median value of the forecast performance loss rate per PV technology and separated for the two partially shaded PV system using the SARIMA models based on the BIC.**

| | | Forecast PLR (%/yr) | | | |
|---|---|---|---|---|---|
| a/a | PV technology | Max | Min | Mean | Median |
| 01 | Mono-c-Si | 0.99 | 0.39 | 0.75 | 0.84 |
| 02 | Multi-c-Si | 1.26 | 0.50 | 0.80 | 0.63 |
| 03 | Thin-Film | 2.00 | 1.09 | 1.63 | 1.85 |
| 04 | Partially shaded PV systems | 0.90 | 0.56 | 0.77 | 0.83 |
| 05 | For all the PV systems | 2.00 | 0.39 | 1.00 | 0.89 |

**Table A.15. Maximum, minimum, mean and median value of the absolute difference between the actual and forecast performance loss rate per PV technology and separated for the two partially shaded PV system using the SARIMA models based on the BIC.**

| | | Absolute Difference (%/yr) | | | |
|---|---|---|---|---|---|
| a/a | PV technology | Max | Min | Mean | Median |
| 01 | Mono-c-Si | 0.51 | 0.00 | 0.22 | 0.19 |
| 02 | Multi-c-Si | 0.34 | 0.03 | 0.17 | 0.12 |
| 03 | Thin-Film | 0.47 | 0.07 | 0.27 | 0.28 |
| 04 | Partially shaded PV systems | 0.51 | 0.16 | 0.33 | 0.34 |
| 05 | For all the PV systems | 0.51 | 0.00 | 0.24 | 0.25 |

## A.4 Results of SARIMA models as proposed in [6]

**Table A.16. Identified SARIMA models for the PV grid connected systems as proposed in [6] . (*Partially shaded systems)**

| a/a | Manufacturer | $(p,d,q),(P,D,Q)_S$ | RMSE (%) | MAE (%) | AIC | AICc | BIC |
|---|---|---|---|---|---|---|---|
| 01 | Solon * | $(1,1,1),(1,1,1)_{12}$ | 3.80 | 2.95 | -217.97 | -216.51 | -208.72 |
| 02 | Sanyo | $(2,1,1),(1,1,1)_{12}$ | 2.47 | 2.06 | -248.19 | -246.09 | -237.09 |
| 03 | Atersa | $(2,1,1),(1,1,1)_{12}$ | 2.83 | 2.35 | -241.70 | -239.60 | -230.59 |
| 04 | Suntechnics | $(3,1,1),(1,1,1)_{12}$ | 2.37 | 1.92 | -231.98 | -229.11 | -219.03 |
| 05 | Schott Solar (EGF) | $(3,1,1),(1,1,1)_{12}$ | 1.63 | 1.35 | -255.36 | -252.49 | -242.41 |
| 06 | BP Solar * | $(3,1,2),(1,1,0)_{12}$ | 5.67 | 4.88 | -190.84 | -187.97 | -177.89 |
| 07 | SolarWorld | $(3,1,1),(1,1,1)_{12}$ | 2.33 | 1.94 | -244.57 | -241.70 | -231.62 |
| 08 | Schott Solar (MAIN) | $(3,1,1),(1,1,1)_{12}$ | 1.66 | 1.31 | -250.18 | -247.31 | -237.23 |
| 09 | Würth Solar | $(3,1,1),(1,1,1)_{12}$ | 2.29 | 1.95 | -236.71 | -233.84 | -223.76 |
| 10 | First Solar | $(3,1,1),(1,1,1)_{12}$ | 1.52 | 1.18 | -270.33 | -267.46 | -257.38 |
| 11 | MHI | $(3,1,1),(1,1,1)_{12}$ | 1.68 | 1.27 | -240.71 | -237.84 | -227.76 |

**Table A.17. RMSE and MAE between the actual and forecasted (using SARIMA models as proposed in [6]) PR values after the application of the RPCA methodology for the 6th, 7th and 8th year. (*Partially shaded systems).**

| | | 6th year | | 7th year | | 8th year | |
|---|---|---|---|---|---|---|---|
| a/a | Manufacturer | RMSE (%) | MAE (%) | RMSE (%) | MAE (%) | RMSE (%) | MAE (%) |
| 01 | Solon * | 0.70 | 0.57 | 3.52 | 3.24 | 2.57 | 2.19 |
| 02 | Sanyo | 1.10 | 0.93 | 1.73 | 1.67 | 1.91 | 1.86 |
| 03 | Atersa | 3.05 | 2.93 | 3.22 | 3.07 | 1.30 | 1.14 |
| 04 | Suntechnics | 0.67 | 0.59 | 1.84 | 1.69 | 1.62 | 1.47 |
| 05 | Schott Solar (EGF) | 1.16 | 0.99 | 1.71 | 1.55 | 0.59 | 0.49 |
| 06 | BP Solar * | 2.68 | 2.14 | 5.85 | 5.51 | 7.63 | 7.27 |
| 07 | SolarWorld | 1.70 | 1.44 | 1.04 | 0.78 | 1.98 | 1.83 |
| 08 | Schott Solar (MAIN) | 0.53 | 0.48 | 1.77 | 1.66 | 0.55 | 0.46 |
| 09 | Würth Solar | 1.02 | 0.94 | 2.01 | 1.81 | 2.72 | 2.51 |
| 10 | First Solar | 1.19 | 1.04 | 1.59 | 1.51 | 0.51 | 0.35 |
| 11 | MHI | 0.61 | 0.49 | 0.93 | 0.82 | 0.77 | 0.72 |

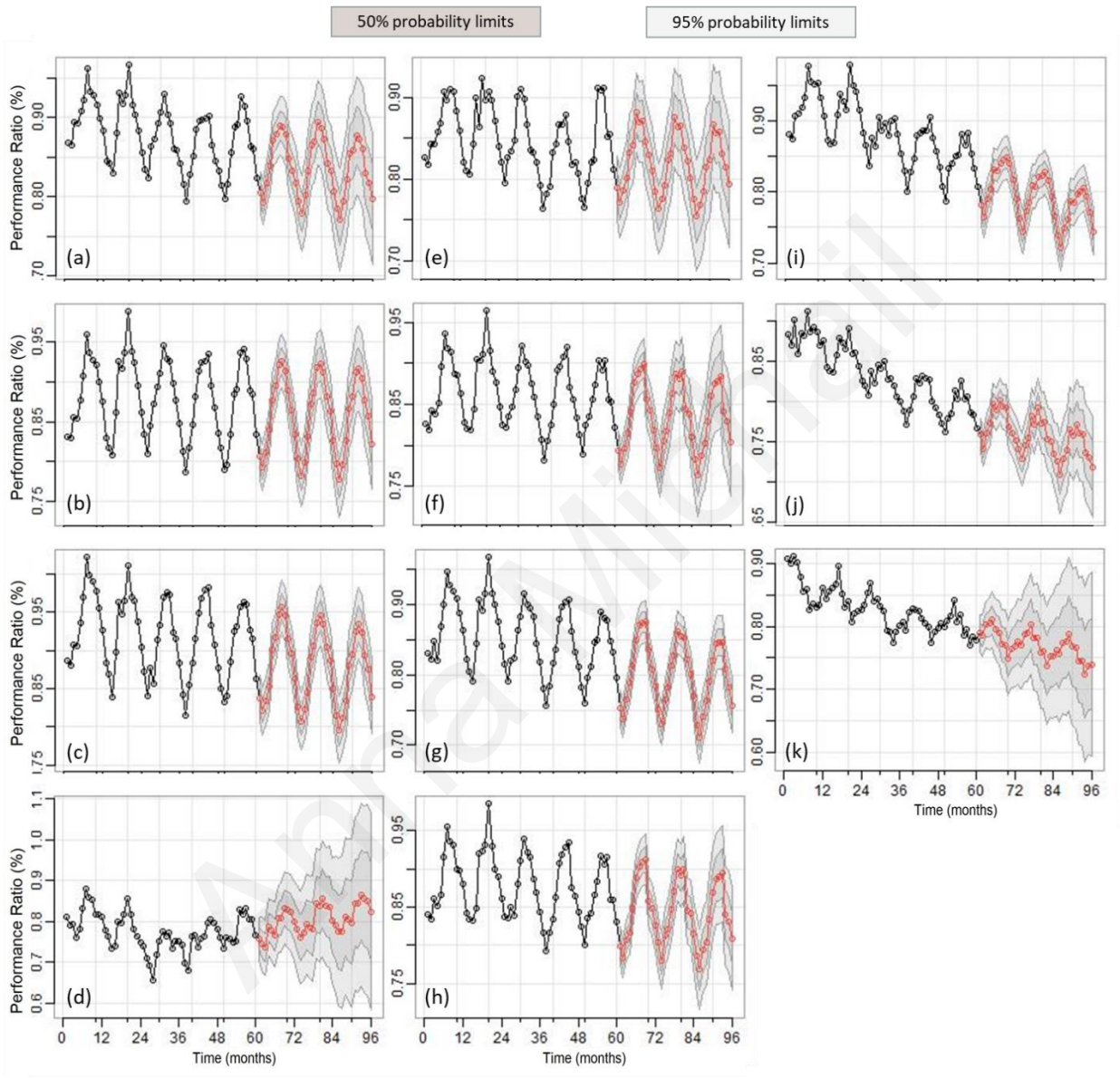


**Figure A.16. Upper and lower 50% and 95% probability limits using the SARIMA models as proposed in** [6]**. Mono-c-Si systems: (a) Sanyo (b) Atersa (c) Suntechnics and (d) BP Solar*. Multi-c-Si systems: (e) Solon*(f) Schott Solar (EGF) (g) Solar World and (h) Schott Solar (MAIN). Thin film systems: (i) Würth Solar (j) First Solar (EGF) and (k) MHI. (* Partially shaded systems)**

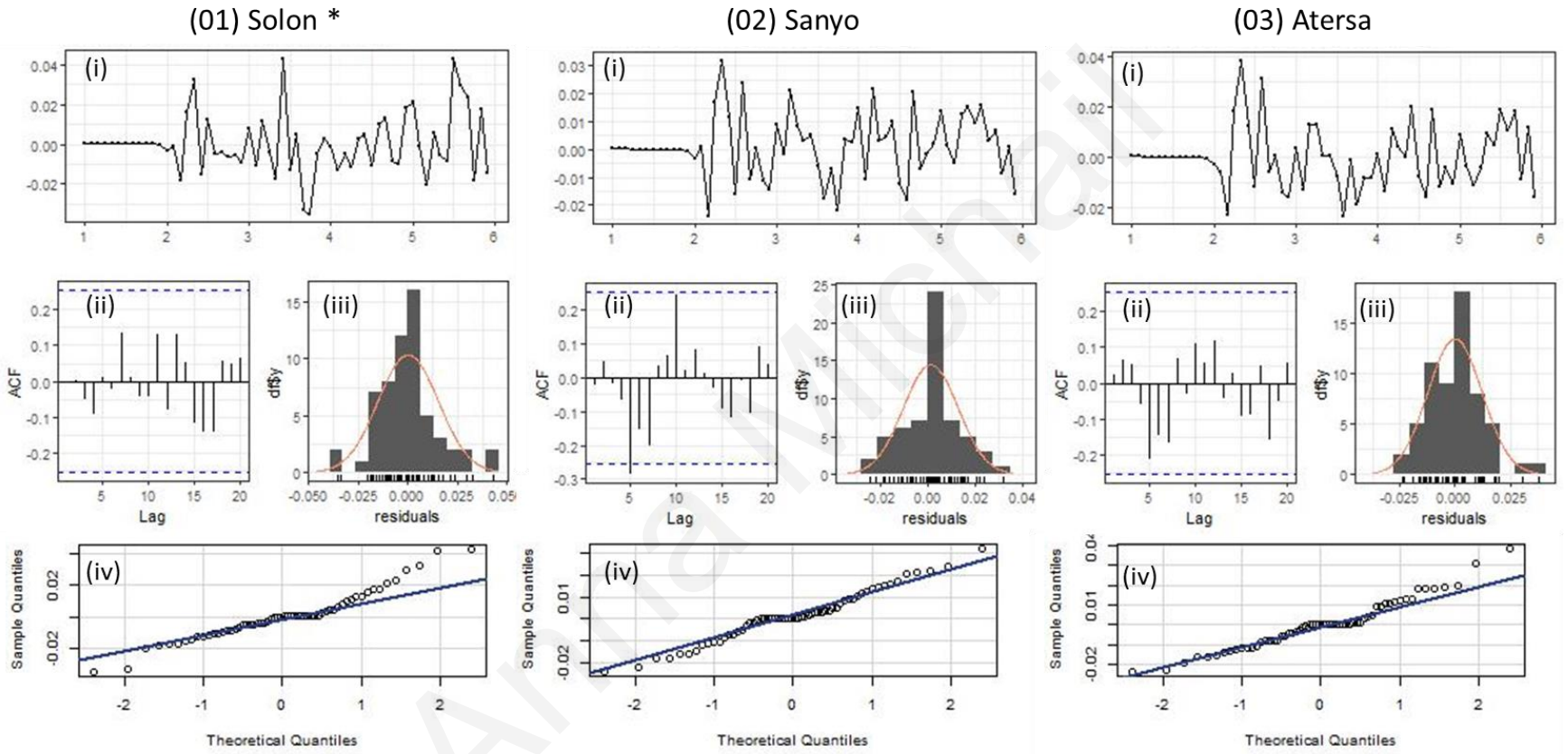


**Figure A.17. (i) Time plot, (ii) ACF, (iii) histogram of the residuals and (iv) Q-Q plot that displays the sample residuals with "o" and the theoretical-normal normal quantiles with a solid line obtained using the SARIMA models proposed in** [6] **for the 1st, 2nd and 3rd PV system. (*Partially shaded systems)**

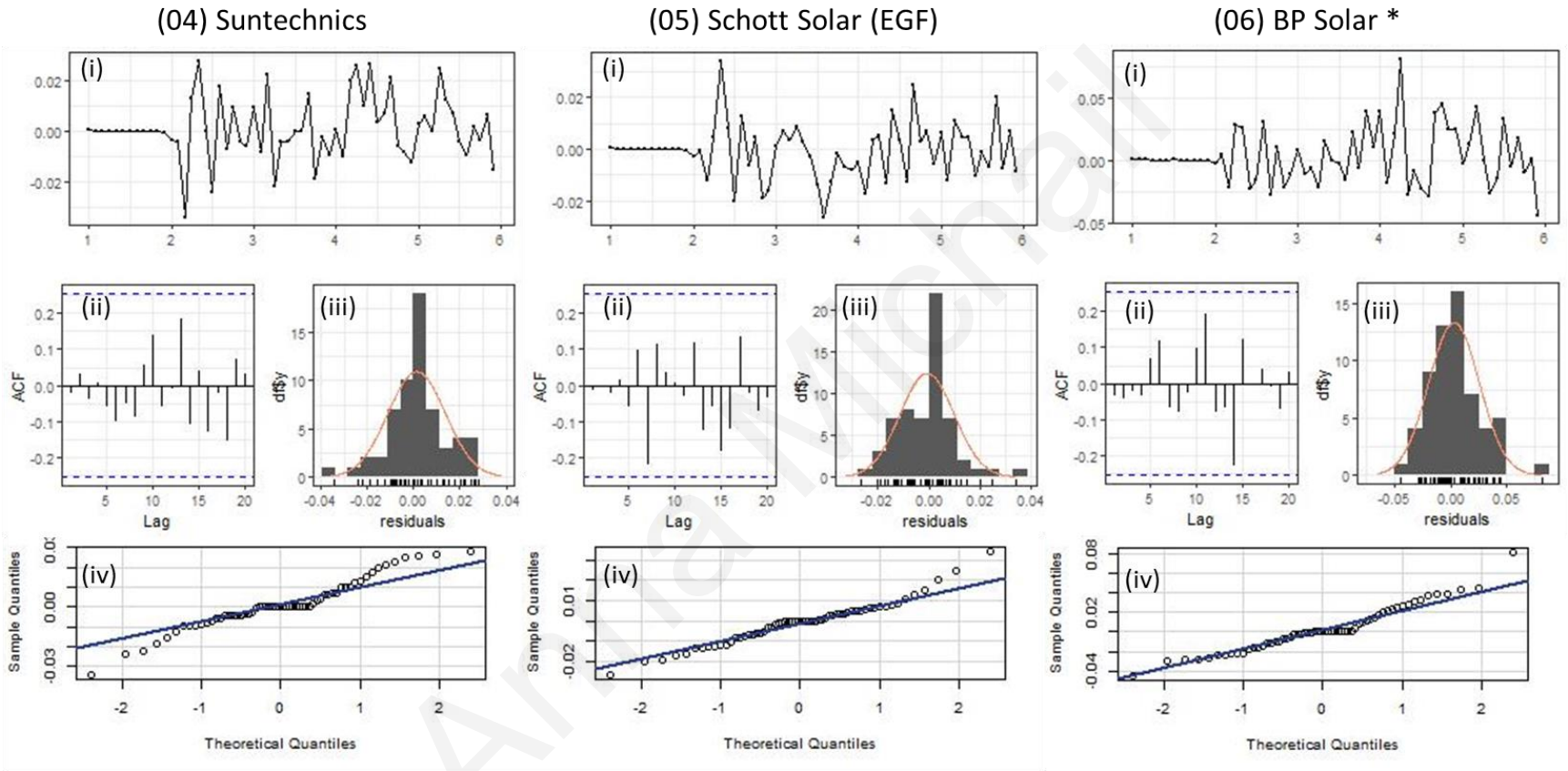


**Figure A.18. (i) Time plot, (ii) ACF, (iii) histogram of the residuals and (iv) Q-Q plot that displays the sample residuals with "o" and the theoretical-normal normal quantiles with a solid line obtained using the SARIMA models proposed in [6] for the 4th, 5th and 6th PV system. (*Partially shaded systems)**

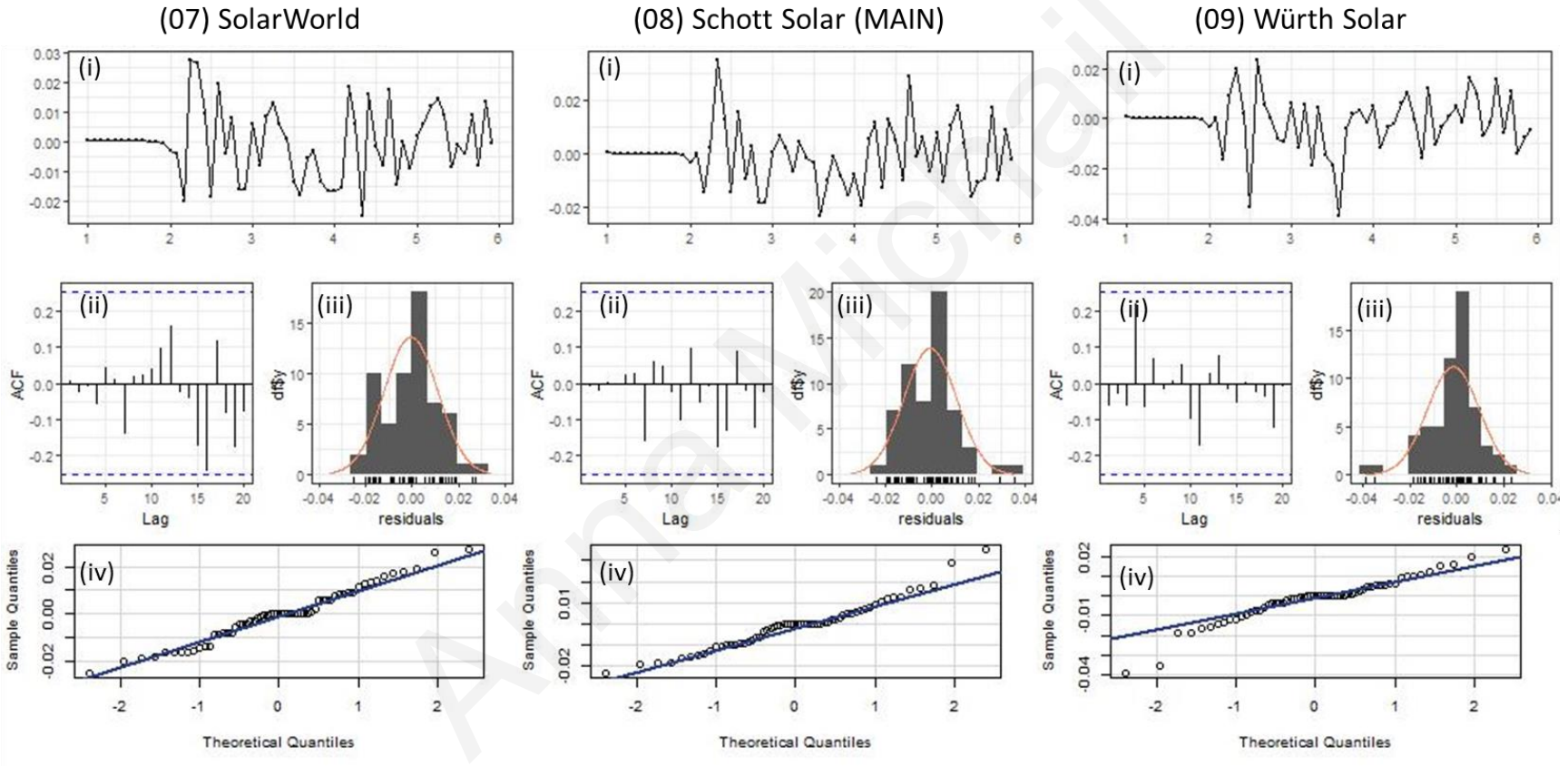


**Figure A.19. (i) Time plot, (ii) ACF, (iii) histogram of the residuals and (iv) Q-Q plot that displays the sample residuals with "o" and the theoretical-normal normal quantiles with a solid line obtained using the SARIMA models proposed in** [6] **for the 7th, 8th and 9th PV system.**

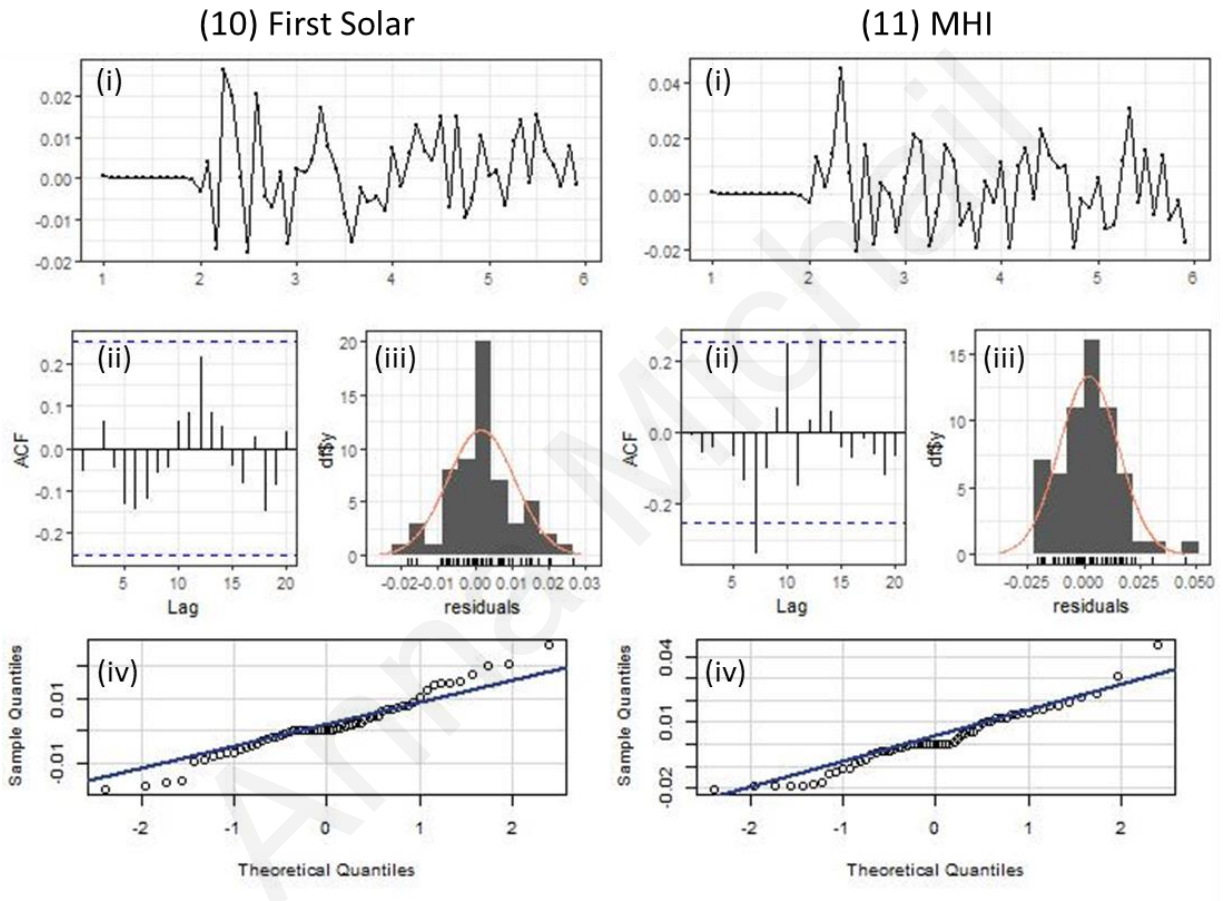


**Figure A.20. (i) Time plot, (ii) ACF, (iii) histogram of the residuals and (iv) Q-Q plot that displays the sample residuals with "o" and the theoretical-normal normal quantiles with a solid line obtained using the SARIMA models proposed in** [6] **for the 10th and 11th PV system.**

**Table A.18. Actual and forecast performance loss rate per PV system using SARIMA models as proposed in** [6]**.**

| | | 6th year | | | 7th year | | | 8th year | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **a/a** | **Manufacturer** | **Actual (%/yr)** | **Forecast (%/yr)** | **Dif. (%/yr)** | **Actual (%/yr)** | **Forecast (%/yr)** | **Dif. (%/yr)** | **Actual (%/yr)** | **Forecast (%/yr)** | **Dif. (%/yr)** |
| 01 | Solon * | 0.69 | 0.64 | -0.04 | 1.30 | 0.68 | -0.63 | 1.11 | 0.71 | -0.39 |
| 02 | Sanyo | 0.83 | 0.96 | 0.13 | 0.71 | 0.95 | 0.24 | 0.73 | 0.97 | 0.24 |
| 03 | Atersa | 0.90 | 0.28 | -0.61 | 0.88 | 0.34 | -0.55 | 0.47 | 0.37 | -0.10 |
| 04 | Suntechnics | 0.81 | 0.89 | 0.08 | 0.65 | 0.93 | 0.27 | 0.76 | 0.96 | 0.20 |
| 05 | Schott Solar (EGF) | 0.38 | 0.52 | 0.14 | 0.79 | 0.51 | -0.28 | 0.60 | 0.57 | -0.03 |
| 06 | BP Solar * | 0.73 | 0.48 | -0.24 | 1.09 | 0.10 | -0.99 | 1.03 | 0.11 | -0.92 |
| 07 | SolarWorld | 0.87 | 1.21 | 0.34 | 1.13 | 1.18 | 0.06 | 0.97 | 1.27 | 0.30 |
| 08 | Schott Solar (MAIN) | 0.68 | 0.65 | -0.03 | 0.97 | 0.66 | -0.31 | 0.78 | 0.71 | -0.07 |
| 09 | Würth Solar | 1.97 | 1.91 | -0.07 | 2.26 | 1.97 | -0.30 | 2.39 | 2.03 | -0.36 |
| 10 | First Solar | 2.11 | 1.91 | -0.21 | 2.14 | 1.90 | -0.25 | 1.89 | 1.89 | 0.00 |
| 11 | MHI | 1.39 | 1.35 | -0.04 | 1.58 | 1.37 | -0.21 | 1.37 | 1.40 | 0.03 |

**Table A.19. Maximum, minimum, mean and median value of the forecast performance loss rate per PV technology and separated for the two partially shaded PV system using the SARIMA models as proposed in** [6]**.**

| | | **Forecast PLR (%/yr)** | | | |
|---|---|---|---|---|---|
| **a/a** | **PV technology** | **Max** | **Min** | **Mean** | **Median** |
| 01 | Mono-c-Si | 0.97 | 0.28 | 0.74 | 0.93 |
| 02 | Multi-c-Si | 1.27 | 0.51 | 0.81 | 0.66 |
| 03 | Thin-Film | 2.03 | 1.35 | 1.75 | 1.90 |
| 04 | Partially shaded PV systems | 0.71 | 0.10 | 0.45 | 0.56 |
| 05 | For all the PV systems | 2.03 | 0.10 | 0.98 | 0.93 |

**Table A.20. Maximum, minimum, mean and median value of the absolute difference between the actual and forecast performance loss rate per PV technology and separated for the two partially shaded PV system using the SARIMA models as proposed in** [6]**.**

| | | **Absolute Difference (%/yr)** | | | |
|---|---|---|---|---|---|
| **a/a** | **PV technology** | **Max** | **Min** | **Mean** | **Median** |
| 01 | Mono-c-Si | 0.61 | 0.08 | 0.27 | 0.24 |
| 02 | Multi-c-Si | 0.34 | 0.03 | 0.17 | 0.14 |
| 03 | Thin-Film | 0.36 | 0.00 | 0.16 | 0.21 |
| 04 | Partially shaded PV systems | 0.99 | 0.04 | 0.54 | 0.51 |
| 05 | For all the PV systems | 0.99 | 0.00 | 0.26 | 0.24 |