DETECTION OF DIFFERENT DEGRADATION MECHANISMS FOR PHOTOVOLTAIC SYSTEMS USING DATA-DRIVEN ALGORITHMS

by

Panagiotis Goumenos

Submitted to the University of Cyprus in partial fulfilment of the requirements for the degree of Master of Science in Energy Technologies and Sustainable Design

Department of Electrical and Computer Engineering

December 2021

DETECTION OF DIFFERENT DEGRADATION MECHANISMS FOR PHOTOVOLTAIC SYSTEMS USING DATA-DRIVEN ALGORITHMS

by

Panagiotis Goumenos

Examination committee:

Georghiou E. George

Professor, Department of Electrical and Computer Engineering, Advisor

Charalambos A. Charalambous

Associate Professor, Department of Electrical and Computer Engineering

Kyprianou Andreas

Associate Professor, Department of Mechanical and Manufacturing Engineering

Abstract

A main requirement for the photovoltaic (PV) uptake is the assurance of the lifetime energy yield and formulation of strict warranties to reduce investment risk. This requires reliable detection of PV underperformance issues (e.g., failures and degradation modes) and accurate evaluation of the degradation rate of fielded PV systems.

Degradation is the decrease in PV system performance over time due to mechanisms not reversed in the field. It is well known that all PV systems suffer from degradation and it is observed at all levels (i.e., cell, module, array, and system) with different factors and degradation mechanisms apparent at each level. The key extrinsic variables that contribute to PV degradation are temperature, humidity and radiation.

A major degradation mechanism for silicon wafer-based PV modules is potential induced degradation (PID), which results from the use of increased system voltage of up to 1000 V. The factors that can cause PID include difference in potential between earth and module, high temperatures and humidity, creating leakage current through the ground from which loss in power is caused. PID does not occur on all PV modules/cells. Other degradation mechanisms include the light induced degradation (LID) and the light elevated temperature (LeTID). LID occurs during the first year of module operation due to sun exposure. It affects the performance of the installed modules with respect to name plate data delivered by some PV module manufactures. LeTID is a form of solar cell degradation seen in the field and is accelerated by high irradiation and temperatures.

The purpose of this master's thesis is to analyse data from grid-connected PV systems and diagnose different degradation modes (e.g., PID and degradation) at early stages in attempt to safeguard the optimal performance of PV systems. This will be achieved by performing statistical analysis on the acquired data. In parallel, predictive algorithms will be implemented for forecasting PV degradation rates. The proposed methodology will be benchmarked against field data from test PV systems installed at the University of Cyprus (UCY) in Nicosia, Cyprus.

Acknowledgments

First and foremost, I would like to express my gratitude towards my supervisor, Prof. George E. Georghiou for the valuable advice he has offered me, for the trust he has shown in me and for giving me the opportunity to undertake such an exciting and challenging project.

I am also deeply grateful to Mr. Andreas Livera for his useful guidance and assistance in 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 PV Technology Lab staff for their continued support throughout my studies at the University of Cyprus.

Table of Contents

Table of Figures

List of Tables

Nomenclature

AC	Alternating current
ACF	Autocorrelation Function
AIC	Akaike's Information Criterion
c-Si	Crystalline silicon
CSD	Classical series decomposition
DC	Direct current
E_{DC}	Measured DC energy
HW	Holt-Winters exponential smoothing
I_A	Array DC Current
G_I	In-plane irradiance
KPI	Key performance indicator
LeTID	Light and Elevated Temperature Induced Degradation
LID	Light Induced Degradation
LOESS	Locally estimated scatterplot smoothing
LR	Linear Regression
MAE	Mean Absolute Error
NA	Not Available
OTF	Outdoor Test Facility
P_A	Array DC Power
PACF	Partial Autocorrelation Function
PID	Potential Induced Degradation

Poly c-Si	Polycrystalline Silicon
PR	Performance Ratio
PR_{corr}	Temperature corrected performance ratio
PV	Photovoltaic
Poly-c Si	Polycrystalline silicon
R_D	Degradation Rate
RMSE	Root Mean Square Error
RPCA	Robust Principal Component Analysis
SARIMA	Seasonal Autoregressive Integrated Moving Average
T_{amb}	Ambient Temperature
T_{mod}	Module Temperature
UCY	University of Cyprus
UV	Ultraviolet
V_A	Array DC Voltage
Y_A	Actual energy yield
Y_r	Reference energy yield

1. Introduction

1.1. Motivation

An important aspect in the field of photovoltaics (PV) that is yet to be addressed is the correct estimation of the degradation rate (R_D) through field measurements. Degradation occurs during the field operation of PV modules, either as an unavoidable degradation process or as a degradation mechanism that occur in some PV modules/cells for various causes. Performance degradation occurs in all levels such as cell, module, array and system, and the main factors that are related to degradation are temperature, humidity, soiling, snow, precipitation, and solar irradiation. At the cell level, the main mechanisms that underline the degradation rate are corrosion, contact stability and cracks. At the module level, degradation can be caused by mismatches, broken interconnects and hot spots, while in array level it is associated with soiling and module mismatches [1]. These factors cause different degradation mechanisms and place substantial stress on a PV system during its lifespan resulting in a loss in durability that must be evaluated using the degradation rate. Outdoor field testing has been critical in evaluating long-term behavior and longevity, since it constitutes the operating environment of PV systems [2]. In the first 10 years of PV module operation, most warranties indicate that the PV module's output power will not decrease more than 1% each year. This value is not carried out by testing the PV modules to the end of their lifetime in the field, as they taken under laboratory conditions at STC. Knowledge of the R_D is crucial for reducing uncertainties and financial risks [2]. It is important to note that up to date, there is no standardized method for accurate calculating the R_D of fielded PV systems.

In this thesis, comparison of different methods for degradation rate estimation is performed for not only modules affected by aging but also for modules affected from other degradation modes. The investigated methods were the Linear Regression (LR), Locally estimated scatterplot smoothing (LOESS) and Robust Principal Component Analysis (RPCA) [3]. The LR method was employed due to the clarity that describes the model and the run time complexity during execution. The LOESS method was selected because it provides robust estimates of the trend and seasonal components that are not distorted by outliers and missing values [4]. The RPCA method was used because it does not produce

any uncertainty during the calculation and because it extracts the outliers from the timeseries.

The work presented in this dissertation demonstrates how current arbitrary practices can be replaced by generalized methodologies that can provide inference, eliminate selection bias, and allow quantification of uncertainty while extracting as much information as possible from all useful measurement data in an unsupervised manner using proper statistical analyses.

1.2. Research Objectives

The objectives of this thesis were to:

- Accurately estimate the degradation rate of PV systems that either suffer from aging or PID with different data-driven methods (LR, LOESS and RPCA).
- Forecast the performance ratio and degradation rate of PV systems and evaluate the forecasting accuracy.
- Detect PID in PV systems with the use of data-driven algorithms.

1.3. Novelty

The novelty of the thesis is that PID will be detected using data-driven algorithms applied on the available field measurements without the need of extra equipment (i.e., I-V tracer, SunSniffer, etc.). This algorithm will be used for detecting the presence of PID, at early stages (from 4 to 10 months' time). Furthermore, a methodology will be developed capable of differentiating degradation mechanisms (such as PID, actual degradation due to aging). Even though the comparative methods for calculating the degradation rate were utilized before, this was never done for a PID affected system, in comparison to a system affected from aging. Through the research, how PID affects the performance rate of PV modules is introduced for the first time. The methods of detecting the PID were neither used in literature on any level.

1.4. Research Impact

Assuming that a *1 kW* system would generate *4 kWh* per day on an average, a *5 kW* PV system would generate about *20 kWh/day*. Considering an electricity price of *0.09 €/kWh* [5], the solar system will be able to generate approximately *54 €/month*. This number is assumed to be in perfect conditions in a PV system without the effect of degradation occurring on them. When considering degradation, the generation of a PV system is less as the power production decreases due to the occurred degradation mode.

This research will aim to assist the early detection of different degradation mechanisms. Degradation mechanisms, except degradation due to aging of the modules, does not occur in all modules in the span of their operation and when they are observed the performance of the modules is affected in a negative way. The performance ratio of a PV module signifies how well it operates and the degradation rate how that performance decreases through the years. The accurate estimation of the degradation rate in the range of *0.5%/yr - 0.7%/yr* (the average degradation rate is currently in the range of *0.8%/yr*), could result in an annual gain of *50k €* for a *100 MW* site. Thus, it is important to detect the occurrence of PID which leads to larger degradation rates and thus could damage more the production of the PV modules and accurate estimation could provide even more gain. This detection should cost less with no extra equipment necessary for detection, fast so power loss could be prevented from happening and accurate such that correct detection is conducted.

2. Literature Review

Degradation is due to different mechanisms that can cause decrease in the PV module's performance. These mechanisms differ depending on the reason of occurrence. Some degradation mechanisms studied through the years were due to aging, PID, LeTID and LID. Based on the available data, degradation due to aging and PID will be studied in this thesis.

Degradation due to aging in module level is the degradation mode that occurs after use for several years. It is a mechanism that is evidenced at all levels, i.e., cell, module, array and system with different factors and degradation mechanisms apparent at each level. It is well known that all PV modules suffer from aging degradation. In all cases, the main extrinsic factors related to performance degradation in field operation include temperature, humidity, precipitation, dust, snow, and solar irradiation. For estimating the degradation rate due to aging, most reported research is based on the use of data-driven algorithms. Different statistical methods such as, LR method, Classical series decomposition (CSD), Holt-Winters exponential smoothing (HW) and LOESS were used to estimate the degradation rate for different PV technologies [6]. In [7], the effects that may occur on a PV module after *20* years were presented with the use of optical and IR images. Furthermore, I-V measurements were presented observing the decrease in power, short circuit current, open circuit voltage and field factor through those years obtained a degradation in the range of *0.5 - 0.6 %/yr*. The negative aspect of recording I-V characteristics is that their use is made through additional equipment, while on the contrary, yield data are widely available.

In [8], analytical models for computing the degradation rates were reported, such as degradation models for corrosion, degradation models of PID, models for UV degradation and models based on multiple stresses. These models commonly use electrical parameters such as power at maximum power point (P_{mp}), short circuit current (I_{sc}), shunt resistance (R_{sh}) and series resistance (R_s) as degradation indicators. Analytical models were developed based on the physical or chemical theories of the degradation mode in investigation. Manufacturers of PV modules supply warranties for the performance of the

modules for at least 20 years, with a maximum loss of less than 20% of the rated power, however, people must be aware that the PV technologies and the location where they were installed affected the degradation [3].

The comparison of the degradation of PV technologies installed outdoors were based on 2 different selections of datasets. The first one was the use of the performance ratio (PR) which specified the overall effect of losses on the module's rated output due to the module's temperature and irradiance [9]. The second was the use of the maximum power normalized to photovoltaics for utility scale applications test conditions of solar irradiance 1000 W/m^2, air temperature 20 ℃ and wind speed of 1 m/s. Afterwards, statistical methods were used on the time series to compute the degradation rate [10][11]. Methods were also used for forecasting the effects of degradation that occur in modules. The methods used for forecasting in [12] where the Persistent Model, Auto-Regressive Integrated Moving Average (ARIMA), k-Nearest Neighbors (kNNs), Artificial Neural Networks (ANNs), and ANNs optimized by Genetic Algorithms. Due to the seasonality of the data, the method used for forecasting in this thesis was the Seasonal ARIMA (SARIMA), as proposed in [13].

A major degradation mechanism for silicon wafer-based PV modules is PID, which results from the use of increased voltage of up to *1000* V. The factors that can cause PID are related to the difference in potential of the module in respect to earth combined with high temperatures and humidity, thus causing loss in output power due to undesirable leakage current to the ground [14]. In recent years, PID has received a considerable amount of attention, as it could potentially lead to catastrophic failure of the PV modules under outdoor conditions [15][16]. The term PID was first introduced from Pingel et al. in 2010 [17], where the research focused on PID of wafer based standard p-type silicon technology with the aim of increasing the lifetime of PV modules. The PID effect, in contrast to the degradation from aging, does not occur on all PV modules. PID usually has no visual effect on the module and different analysis techniques (such as the I-V curve method and image analysis techniques) are available in the literature for detection of this loss. In [18], testing and analysis was performed on a system with voltage bias of -600 V in Florida with the

use of an I-V curve tracer to obtain dark I-V curves. The results of the research performed indicated a decrease of 11% on the power for the c-Si modules under investigation.

Based on [19] because most of the companies perform indoor PID tests, outdoor experimental tests and different methods of detecting PID are necessary for better evaluation. In [20] calculation of the degradation rate was performed for PV modules under high system voltage in the field. The experimental setup comprised of outdoor monitoring of thin film modules in the hot and humid climate of Florida for a period of 2.5 years. The method used was the PV for Utility Scale Applications type regression analysis and the results revealed degradation rate of -5.13 ± 1.53 %/yr and -4.5 ± 1.46 %/yr for positive and negative strings, respectively. In research contacted in Germany for several months, [21] a 35% decrease in power was observed for a back-contact crystalline silicon PV module in which PID occurred.

Although important progress has been produced regarding the understanding of PID in PV modules, there is still research to be made and the complexity of PID introduces a challenge for the researchers. Based on [22] the main causes of PID vary on the different types of module technologies. Although as mentioned, different PID modes may be presented when the same type of modules is stressed under dissimilar conditions. In literature, PID detection was made with several ways, with thermographic imaging [23], electroluminescence imaging [24], open circuit voltage measurement, plotting the I-V curves of modules [25], shunt resistance measurement [24], and with the use of the mass spectrometer to study the sodium migration inside the module [25]. Based on [26] the effects were reversible, climatic factors such as temperature can help overcome PID, it was observed that panels stored at temperature around *100* ℃ for *10* hours lead to a recovery of close to *100%*, and based on [23] reverse voltage system techniques were also studied and had recovery rates of *80* - *96%*. In [27], a prevention method was proposed for PID, the researchers method suggests that for preventing PID, all the components of the module should be resistant to PID, and for the succession of this, a possible solution was to optimize the anti-reflective coating. Based on [28] PID-resistant cells were developed after the encapsulation of the standard glass (EVA), in which the cells degraded by less than 3%

within 25.5 weeks outdoor operation, in contrast to standard multi c-Si modules, which degraded by 51% in the same period.

In this thesis, outdoor measurements are investigated for a 5-year period for modules that suffer from degradation due to aging and for modules that PID was detected. The degradation rate is computed for both scenarios to observe how different degradation mechanisms affect modules of same technologies under the same conditions. The computation of degradation rate is performed under 3 different statistical methods to have a better perspective and verify the results. The last stage of the thesis focusses on the deployment of a method of detection of PID with the use of statistical method taking into account observations made during the research.

3. Methodology

3.1. Flowchart Diagram

The methodology that was followed in this work is depicted in Figure 1**Error! Reference source not found.**. The methodology includes the recording of operational and meteorological data from modules that were affected by aging and PID.

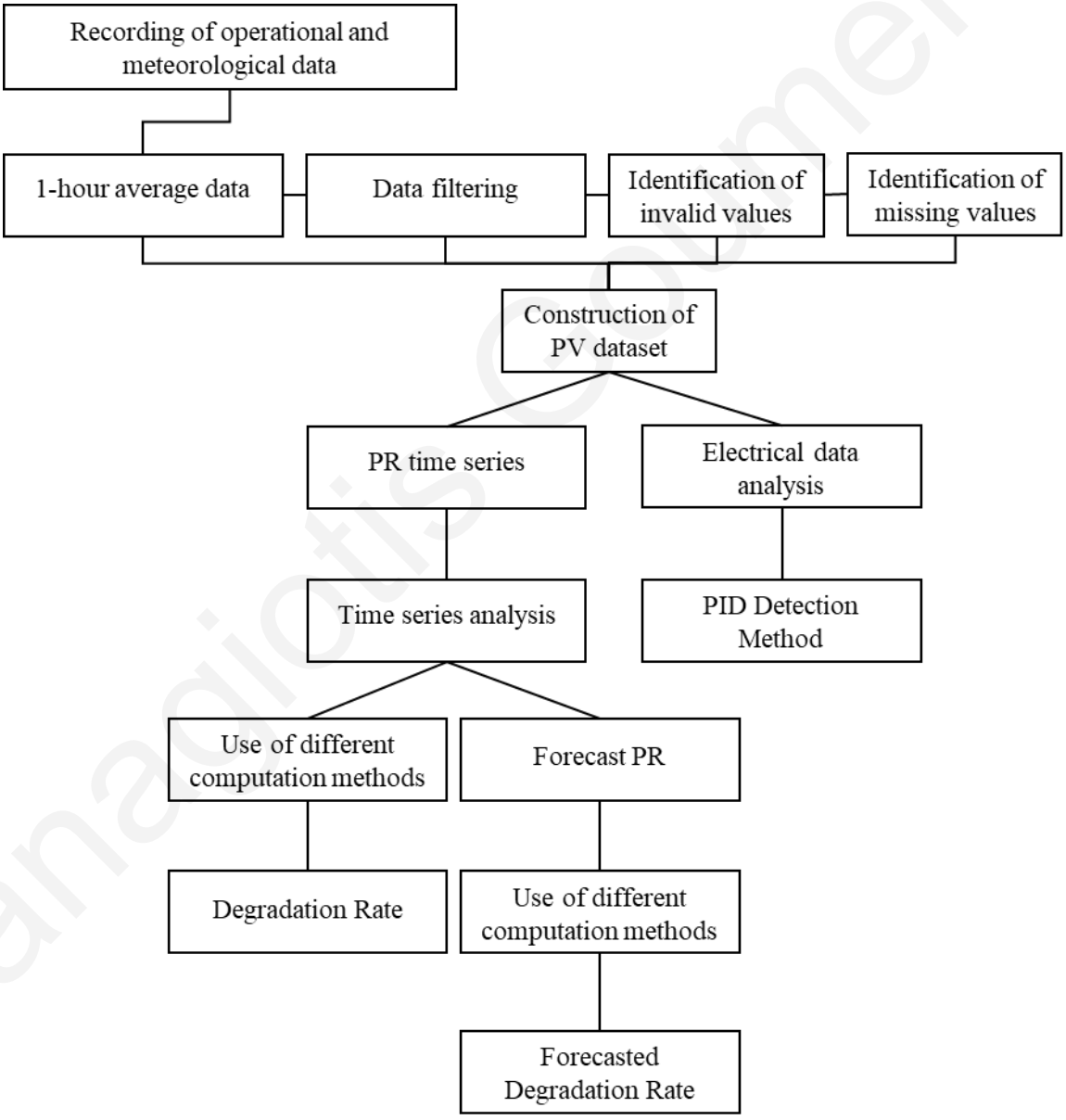

Figure 1. Flowchart of the proposed methodology.

The data processing (data filtering, identification of invalid values, identification of missing values) is first applied to the acquired data to construct the PV dataset. The PV dataset is used for the construction of PR time series and for further analysis for PID detection. The PR time series is then analysed with different methods (LR, LOESS, RPCA) to calculate the degradation rate of the systems. Furthermore, the PR time series is also used for forecasting the PR and with the use of the same methods, compute the forecasted degradation rate that will be compared to the actual.

3.2. Experimental Setup

The analysis for PID detection was performed on poly-crystalline silicon (poly-c Si) PV modules. The PV modules are installed in an open-field mounting arrangement and at an inclination angle of 27.5° due South, at the outdoor test facility (OTF) of the University of Cyprus (UCY) in Nicosia, Cyprus (see Figure 2). Each small-scale system comprises of 2 PV modules connected in series, with the nominal power of one module being 240 W_p. 2 out of the 4 PV systems, were affected by PID (verified and tested indoors) and measurements were taken during the PV normal and faulty operation for a reporting period of 5 years (April 2015 – April 2020).

Figure 2. Poly c-Si grid-connected PV system at the OTF of the UCY.

The meteorological measurements include the in-plane irradiance (G_I), ambient (T_{amb}) and module (T_{mod}) temperature, while the electrical measurements include the array DC current (I_A), voltage (V_A) and power (P_A) for both analyses.

3.3. Data collection and construction of PV datasets

The data were collected every 1 hour and were processed to fill missing data to complete the time stamp so that the analysis of the data could be made. For example, a correct sample of the dataset is presented in Table 1.

Table 1. Sample of PV measurements.

Timestamp	**P_A (W)**	**V_A (V)**	**T_{mod} (°C)**	**G_I (W/m^2)**
07/06/2015 11:00	359.81	49.67	50.61	923.13
07/06/2015 12:00	373.28	48.27	53.18	1002.66
07/06/2015 13:00	285.66	51.32	42.11	723.55

The daily electrical data were not uniform as the end time differ per day and thus, every day consisted of different number of data points. The electrical data acquired were corrected with the insertion of NAs such as the daily data become uniform. This correction was necessary as the electrical data were matched with the meteorological data that were also acquired to create the complete datasets containing both electrical and meteorological measurements of the modules. By obtaining both meteorological and electrical data, PR was computed.

The outliers of the measurements, were removed from the datasets, replacing those values with NAs. For removing the outliers, a statistical method was deployed, the boxplot outlier rule. In this method, the lower quantile, Q1 (25th percentile), the median, the upper quantile, Q3 (75th percentile) and the interquartile range, IQR, which is the difference between Q1 and Q3, are used to describe the variation of the data. This method avoids the mean and standard deviation because they are affected from the outliers.

The boxplot method can be expressed as:

$$x_1 > Q3 + 1.5IQR \bigcup x_1 < Q1 - 1.5IQR \quad (1)$$

The boxplot outlier rule is used to detect values that are outside an estimated interval. By definition, 50% of all measurements are within the ±0.5 IQR of the median, thus providing a reliable measure of scale.

For better observations, the electrical parameters were aggregated to monthly level, so that better analysis could be made through the years of operation.

3.4. Construction of performance index

For this thesis, the PR was selected as the performance metric for calculating the R_D. The PR is a normalized key performance indicator (KPI), typically used to characterize PV plant performance for acceptance and operations testing [29]. Daily aggregation was not preferred due to larger fluctuations and so monthly aggregation was chosen. Further, AC PR was not chosen as it would represent a complete PV system degradation (cables, inverter, modules etc.) which is not the objective of this analysis, since the analysis focuses exclusive in PV modules.

To calculate the DC PR the following equations were utilized:

$$Y_A = \frac{E_{DC}}{P_{STC}} \quad (2)$$

$$Y_r = \frac{G_I \times \text{timestep}}{G_{STC}} \quad (3)$$

$$PR = \frac{Y_A}{Y_r} \quad (4)$$

where E_{DC} is the measured DC energy, computed by multiplying the P_A and time interval, P_{STC} is the nominal power of the system, G_{STC} is the irradiance in standard test conditions (1000 W/m^2), Y_A and Y_r represent the actual and reference energy yields of the PV system. Furthermore, irradiance values lower than 200 W/m^2 were deleted from the test dataset.

After computing the PR for each hour, monthly aggregation was performed by taking the mean value per month.

The PR constructed from all four systems is presented in Figure 3. Systems A and B are systems that no PID occurred during their five-year operation period and were affected by aging degradation only. In contrary, Systems C and D are PID affected systems as observed and verified through indoor and outdoor testing.

Temperature corrected PR (PR_{corr}) is utilized only in section 4.6 for the detection of PID. PR_{corr} was used due to the high temperatures exhibited in Cyprus during the summer months and contains a term to translate modelled power to the average operating cell temperature. The PR formula is corrected with temperature factor to neutralise short-term PR fluctuation due to temperature variations from STC (25°C). Thus, to exclude the temperature interferences, PR_{corr} was preferred over normal PR for the purposes of detecting PID. For calculating the PR_{corr}, temperature correction is performed on the measured DC energy, and then this temperature corrected energy is used to calculate the PR_{corr}. The equations for implementing the temperature corrected PR_{corr} are the following [30]:

$$E_{DC} = P_{STC} \times \frac{G_I}{G_{STC}} \times [1 - \gamma(T_{mod} - T_{STC})] \times timestep \quad (5)$$

Where E_{DC} is the temperature corrected DC energy, γ is the temperature coefficient for power that correspond to the installed modules equal to *-0.0045*%/℃, and T_{STC} is the temperature on standard test conditions equal to *25* ℃. With the use of equations 1,2 and 3, and the temperature corrected E_{DC} the PR_{corr} is computed.

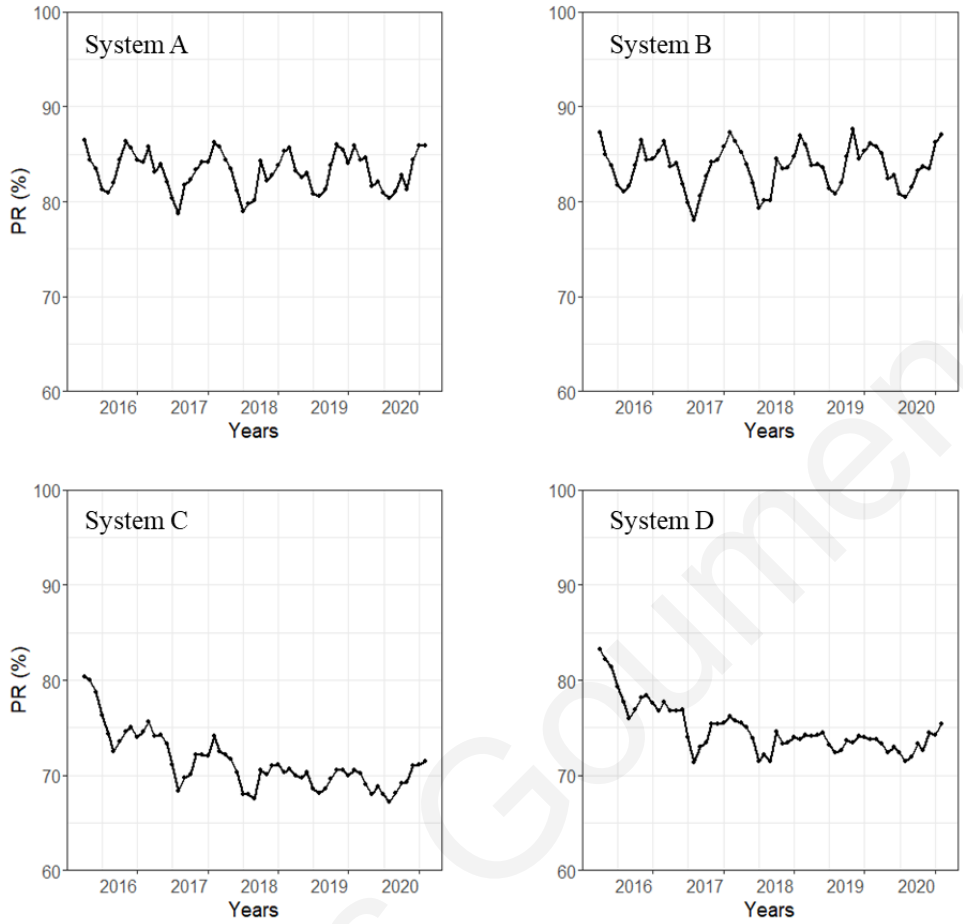

Figure 3. Computed PR for 4 PV systems. Systems A and B are affected by aging, while systems C and D are affected by PID.

3.5. Time series analysis

The annual degradation rate of the PV systems is calculated by applying statistical, comparative or change-point techniques on the PV performance metric [19]. The goal of the statistical analysis is to extract the trend of the PV performance time series and translate the slope of the trend to the annual degradation rate (%/yr) [19]. Similarly, comparative techniques perform linear comparisons on subsequent PV performance time series in order to extract the annual degradation rate (%/yr) [31]. The conventional statistical/comparative methods assume a constant rate of power decrease over time (linear power loss) and the relative degradation rate is calculated by multiplying the ratio of the slope to intercept by 12 or 365 for monthly or daily aggregation, respectively.

For all the techniques, the following issues should be considered:

- Length of available data: Minimum 5-year time series should be available to perform a reliable evaluation [32]. The accuracy of the several methodologies available for the calculation of the PV degradation rate increases with the duration of available time series.
- Seasonality and outliers: The accuracy of the several methodologies is strongly affected by the yearly oscillation of the performance of PV modules/systems. For this reason, special care should be taken in adopting appropriate filtering techniques to eliminate outliers and diminish seasonality.
- Missing data: Missing measurements - outages can greatly affect the value of the annual degradation rate. Appropriate dealing with missing data (using deletion and data imputation techniques) should be considered in order to accurately calculate the degradation rate of PV arrays.
- Construction of the performance time series: The performance metric used to assess the degradation rate can influence the end results. The most commonly used metric is the PR, at a monthly resolution.

For the purposes of this thesis, three different statistical methods were employed for the computation of degradation rate. The evaluation was made with the use of LR, LOESS and RPCA. The LR method was employed due to the clarity that describes the model and the run time complexity during execution, while the LOESS method was selected because it provides robust estimates of the trend and seasonal components that are not distorted by outliers and missing values [4] and the RPCA method was used because it does not produce any uncertainty during the calculation and because it extracts the outliers from the timeseries.

For forecasting the PR time series, SARIMA was used. SARIMA was employed to observe the accuracy of predicting the degradation rate and the PR time series in later years for different degradation mechanisms and distinguish which can be predicted with more accuracy. To compute the SARIMA model's parameters, the 'auto.arima' function [33] was used, while to execute the forecasting operation with those parameters, 'sarima.for' was utilized.

3.5.1. Linear Regression

LR in statistics is defined as a linear approach for modelling the relationship between a scalar response and independent variables. LR was firstly constructed for use in the field of statistics to understand the relation between an input and an output but is also used for machine learning purposes. It is one of the most known and understandable algorithms in machine learning and statistics.

For the formation of the linear equation, specific set of input values are combined to predict the output of those input values. The following equation is used for demonstrating LR:

$$y = A + Bx \tag{6}$$

Where, *x* is the explanatory variable and *y* is the dependent variable. The slope of the formed line created by LR is *B* while the intercept is *A*, which is the value of *y* when *x* is equal to zero. The annual performance loss rates were calculated by multiplying the slope estimator B, by 12 due to the monthly time series [6]. With the use of 'lm' function in R, the uncertainty of the model can also be found. The uncertainty, u_{LR} was computed with the use of the standard error of the slope estimator of the LR as presented in the following equation:

$$u_{LR} = \sqrt{\frac{SS_{residuals}}{df_{residuals}}} \tag{7}$$

Where $SS_{residuals}$ is the residual sum of squares and $df_{residuals}$ is the residual degrees of freedom, calculated as $n-(k+1)$ where, n is the total observations and k is the total model parameters.

The expanded uncertainty, U_{LR}, which is used to express the uncertainty of a measurement result [34], is calculated as follows:

$$U_{LR} = k \times u_{LR} \tag{8}$$

Where k is the coverage factor. This factor was selected as $k = 2$ which produces an interval having a level of confidence of approximately *95%* based on [34].

3.5.2. Seasonal trend decomposition using LOESS

LOESS refers to a nonparametric method that is used to smooth a series of data, for which no assumptions regarding the data's underlying structure are made. In this thesis, LR was used to fit a smooth curve through a scatterplot of data of the created trend. LOESS can provide robust estimations of the trend and seasonal components that are not distorted by outliers or missing values.

LOESS integrates the simplicity of linear least squares regression with the flexibility of nonlinear regression. This is done by fitting models to localized subsets of the data to build a function that describes the deterministic part of the variation in the data [35]. To define a LOESS model at each point in the data set a low-degree polynomial is fitted to a subset of the data, with values of explanatory variables close to the point whose response is being estimated. The polynomial is adjusted using weighted least squares, giving more weight to the points close to the point whose response is being estimated and less weight to the points further away. Furthermore, with the evaluation of the local polynomial using the explanatory variable values of a data point, the value of the regression function of that point can be obtained. Finally, to complete the LOESS fit regression function values are computed for each data point.

A nearest neighbors algorithm is used to determine the subsets of data utilized for each weighted least square fit. In this thesis, a traditional weight function is used for LOESS, the tricube weight function [35]:

$$w(x) = \begin{cases} (1 - |x|^3)^3 & for\ |x| < 1 \\ 0 & for\ |x| \geq 1 \end{cases} \quad (9)$$

The weight for a particular point in any localized subset of data is obtained by evaluating the weight function for the distance between that point and the estimated point, after scaling the distance so that the maximum absolute distance over all the points in the subset is exactly one.

For the computation of degradation rate, LR was applied on the extracted trend of the LOESS method and then a multiplication of the slope estimator by 12 due to the seasonality

that the 12 months of the year provide [6]. The uncertainty of the model is computed with the same method as described in subsection 3.5.1. for the LR method.

3.5.3. Robust Principal Component Analysis

RPCA was used for the computation of the yearly degradation rate [36]. RPCA presumes that the data matrix A is demonstrated with the following equation:

$$A = D + P \tag{10}$$

where D is the low rank matrix, which in the context of data analysis designates that the number of characteristic features dominating the original data in A must be smaller than its size, and P is the unknown sparse perturbation matrix, which is the matrix supposed to cause the outliers. Furthermore, the sparseness of matrix P, indicates that perturbations should not be extended over long periods of time throughout the data. In this research, the matrix A was formed by the PR time series. RPCA extracts the matrix D from the original matrix A by solving the following optimization:

$$minimize\ ||D||_s + \mu||P||_1 \tag{11}$$

where $||D||_s$ is the norm defined by the sum of the singular values of D, $||P||_1$ is the norm defined by the summation of the absolute values of P and μ is a user-defined parameter that trades off between the two norms. In this work, the RPCA was solved with the use of augmented Lagrange multiplier method. In Figure 4, the robust PR data obtained from the RPCA method are compared to the actual data of the time series for one of the systems that suffers from PID.

In order to implement the $RPCA$ methodology, the 60-month DC PR time series was divided to five 12-month time series sections. Thus, creating a data matrix A of dimension 5x12. In Figure 5, the five 12-month time series is plotted before and after the implementation of RPCA for one of the systems under evaluation. The data matrix D was used for the computation of the yearly degradation rate at the end of each test year, namely years 4 and 5. The yearly degradation rate was estimated as follows:

- The area of the region between the line of the 1st year (i.e., Year 1) and the line of the year in question, was divided by the area beneath the curve of the robust data of the first year defined as the yearly degradation rate (see Figure 5 (b)).
- The average annual percentage degradation rate was then defined as the percentage of the degradation rate divided by the number of the year inspected (i.e., by 5).

With the use of a more mathematical representation, to calculate the degradation rate after 5 years of operation, the following equations were utilized:

$$A_{1-5} = \int_{1}^{12} PR_1 - PR_5 \tag{12}$$

$$Degradation\ rate = \frac{A_{1-5}/A_1}{5} \tag{13}$$

Where $A_{1\text{-}5}$ is the area formed between the curve of the first year and the fifth year and A_1 is the area formed below the first curve of the first year.

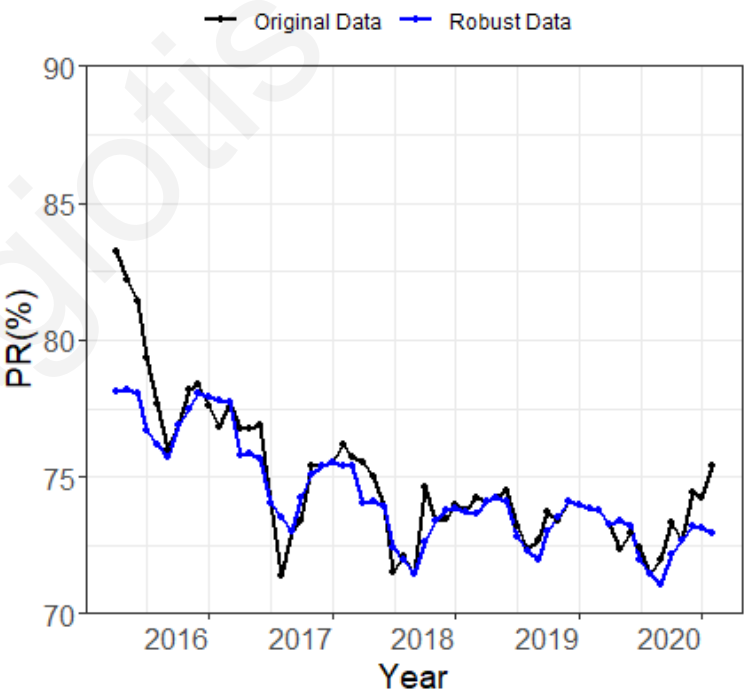

Figure 4. Comparison between the original data and the data after RPCA.

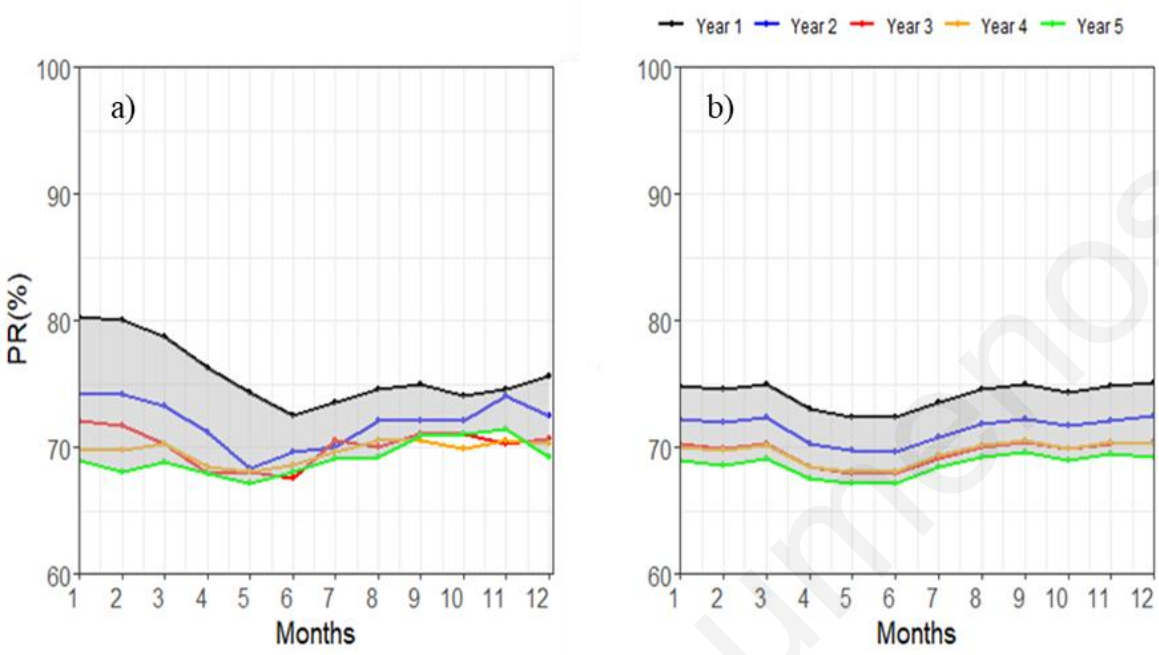

Figure 5. The yearly robust data of the system before (a) and after (b) the implementation of RPCA.

3.5.4. Seasonal Autoregressive Integrated Moving Average

For forecasting the DC PR time series, seasonal autoregressive integrated moving average ($SARIMA$) was used. The $SARIMA$ model was fitted to the time series data either to better understand the data or to predict future points in the series. Furthermore, $SARIMA$ models were useful for modelling a seasonal time series in which the mean and other statistics were not stationary across the years [13].

The selection of the best suited SARIMA model that was made with the use of the 'auto.arima' function was made for each system separately. The general model is denoted as follows:

$$SARIMA(p,d,q)(P,D,Q)_s \tag{14}$$

Where, p and seasonal P indicate number of autoregressive terms (lags of the stationarized series), d and seasonal D indicate differencing that must be done to stationarize series, q

and seasonal Q indicate number of moving average terms (lags of the forecast errors) and s indicates seasonal length in the data. The multiplicative SARIMA model is given by [37]:

$$\Phi_P(B^s)\varphi(B)\nabla_s^D\nabla^d x_t = a + \Theta_Q(B^s)\theta(B)\omega_t \quad (15)$$

Where ω_t is the usual Gaussian white noise process, polynomials $\varphi(B)$ and $\theta(B)$ of orders p and q represent the ordinary autoregressive and moving average components, seasonal autoregressive and moving average components are represented by $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ of orders *P* and *Q*, respectively and ordinary and seasonal difference components represented by $\nabla^d = (1-B)^d$ and $\nabla_s^D = (1-B^s)^D$, respectively. If $\alpha \neq 0$, there is an implied polynomial of order $d + D$ in the forecast function.

For the estimation of SARIMA parameters, several different methods can be employed [38]. Values *D* and *d* are fixed, and the next step is to determine the values of *P*, *Q*, *p* and *q*. One method is to determine those values with the use of Autocorrelation function (ACF) and Partial Autocorrelation function (PACF). For the purposes of this thesis, a different method was deployed, and the SARIMA parameters were selected with the use of Akaike's Information Criterion (AIC). AIC is a measure of prediction error and hence of statistical model quality for a given set of data. The mathematical representation of this model is presented as follows [37]:

$$AIC = \ln\hat{\sigma}_\kappa^2 + \frac{n+2k}{n} \quad (16)$$

Where $\hat{\sigma}_\kappa^2$ is the maximum likelihood estimator for the variance of a regression model with *k* parameters and *n* is the number of observations. For the 'auto.arima' function used from R, the main task is to select the appropriate model order.

The equation used in this function for the succession of this is [33]:

$$AIC = -2\log(L) + 2(p+q+P+Q+k) \quad (17)$$

Where *k* is equal to 1 if $\alpha \neq 0$ (eq. 15) and if not then is equal to *0*, *L* is the maximized likelihood of the model that is fitted to the differenced data and p, q, P, Q are the SARIMA model parameters estimated.

For the implementation of this method in this thesis, firstly the monthly DC PR time series was divided to training years and test years. The number of years that were to be trained was estimated with the use of the $RMSE$ and MAE metrics. Different number of training and testing years were examined to determine the correct number. After estimating the SARIMA parameters, the following step in the procedure was the comparison between the actual and the forecasted DC PR. The forecasted years DC PR time series obtained by the $SARIMA$ model was plotted against the actual DC PR.

3.6. Performance assessment

In order to compare the forecasted and actual PR and degradation rate estimations, the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) metrics were used. The MAE measures the difference between the actual and forecasted data, while the RMSE describes the standard deviation of the forecasting errors. These performance metrics are calculated as follows [39]:

$$MAE = \frac{1}{n} \times \sum_{i=1}^{n} \left| y_{measured,i} - y_{forecasted,i} \right| \tag{18}$$

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{i=1}^{n} (y_{measured,i} - y_{forecasted,i})^2} \tag{19}$$

Where $y_{measured,i}$ and $y_{forecasted,i}$ are the measured and forecasted data respectively.

3.7. PID Detection Methodology

For the detection of PID, analysis on the electrical parameters of the four PV systems was performed to obtain knowledge and observe relations through the parameters that can assist in identifying trends that differ in systems which suffer from PID. The detection of PID was performed by observing 2 parameters through the years of operation. The parameters were the voltage and the power produced by the PV system. The considerations were made for the voltage and power over time and through comparing voltage over power. To

implement those relations, LR was performed over the aforementioned parameters, so that the slope of each relation could be obtain for the extraction of correlations. From tests in PID affected modules and in not affected PID modules, an observation was made that if the first mentioned relation had a positive slope and the other two had a negative slope then the module was affected by PID.

The detection of PID is important to be made not only with the use of simple methods that do not need extra equipment, but also to be made in a small period of time. For quicker detection of PID, PR_{corr} was used.

Furthermore, boxplot representation was also used as another method that can assist in verifying the occurrence of PID. Comparison between each year's median value is conducted and by considering the decrease in voltage per year, detection is made. A boxplot representation is presented in Figure 6 [40]. Box plots divide the data into sections that each contain approximately 25% of the data in that set. The two extreme points of the boxplot represent the minimum and maximum values, the median is the middle value of the dataset which marks the mid-point of the data and is shown by the line that divides the box into two parts, the middle number between the smallest number and the median of the dataset represent the first quartile (Q1), while the middle value between the median and the highest value of the dataset represent the third quartile (Q3).

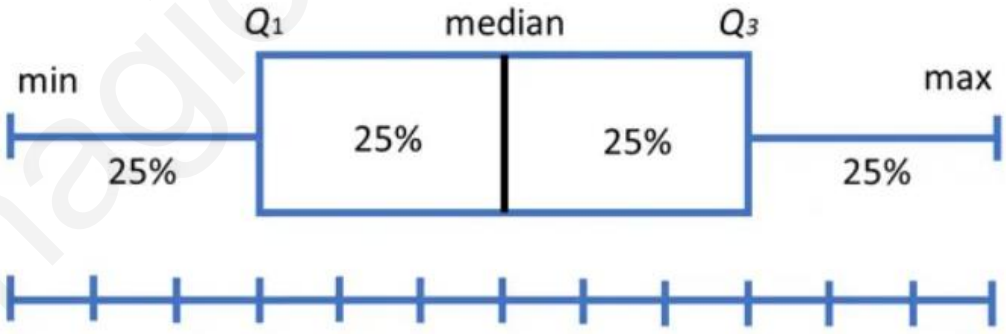

Figure 6. Boxplot representation [40].

4. Results

4.1. Degradation Rate with Linear Regression

With the use of LR, the results are presented in Figure 7 in which with black colour the PR time series is presented and with blue colour the line created by LR on the monthly PR time series. Systems A and B that exhibit normal degradation due to aging are observed to have lower decline that systems C and D that PID occurred. The degradation for the first system was calculated as $R_D = -0.081 \pm 0.37$ %/yr, for the second system $R_D = 0.058 \pm 0.42$ %/yr, for the third system $R_D = -1.54 \pm 0.37$ %/yr and for the fourth system $R_D = -1.28 \pm 0.33$ %/yr.

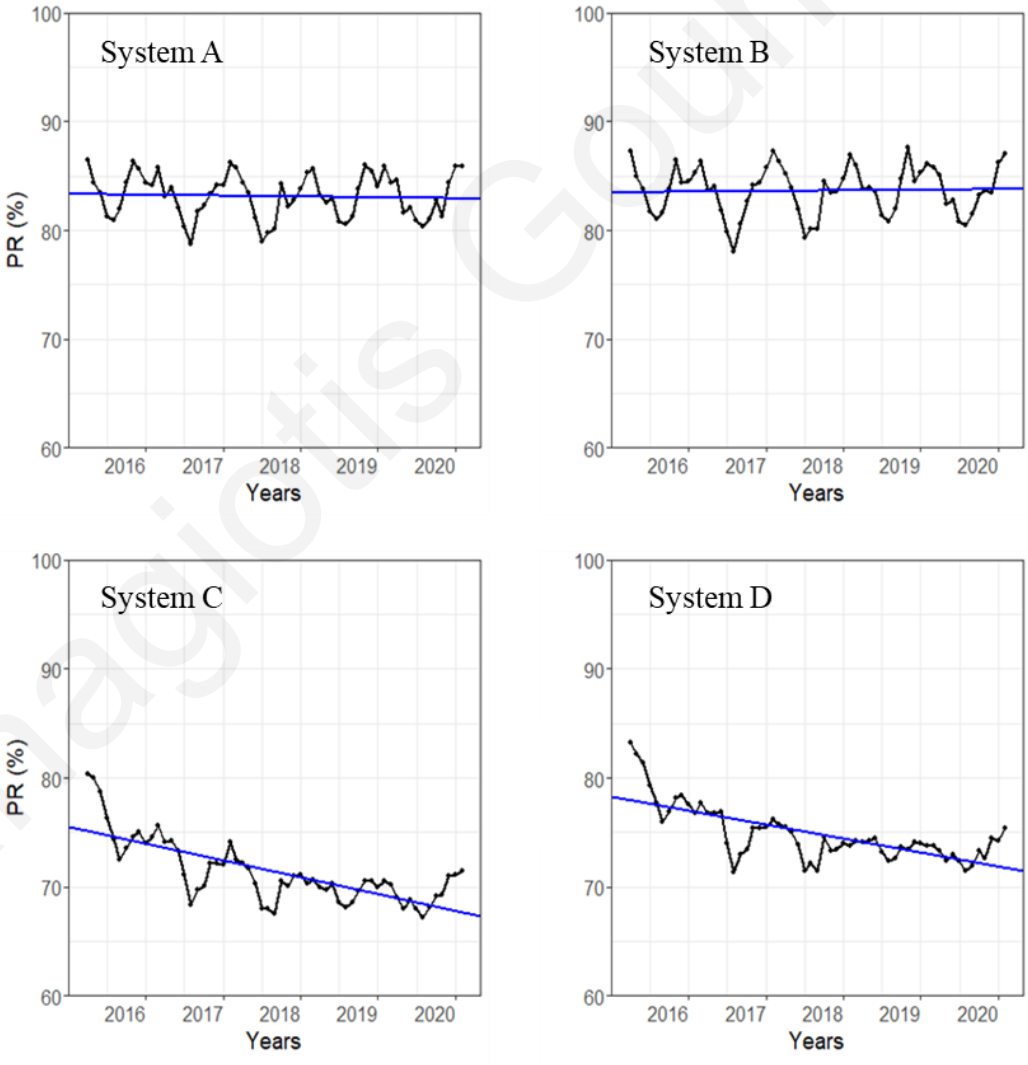

Figure 7. LR on the PR time series of the systems.

4.2. Degradation Rate with LOESS

By using LR on the extracted LOESS trend, the results are presented in Figure 8. With black colour the PR time series is observed, with red colour the extracted trend and with blue the LR fit on the trend. Similar to the previous results, systems C and D display higher rates of degradation than systems A and B. Degradation rate for systems A and B were R_D = *-0.069 ± 0.065 %/yr* and R_D = *0.063 ± 0.061 %/yr*, respectively and for systems C and D were R_D = *-1.55 ± 0.24 %/yr* and R_D = *-1.30 ± 0.23 %/yr,* respectively.

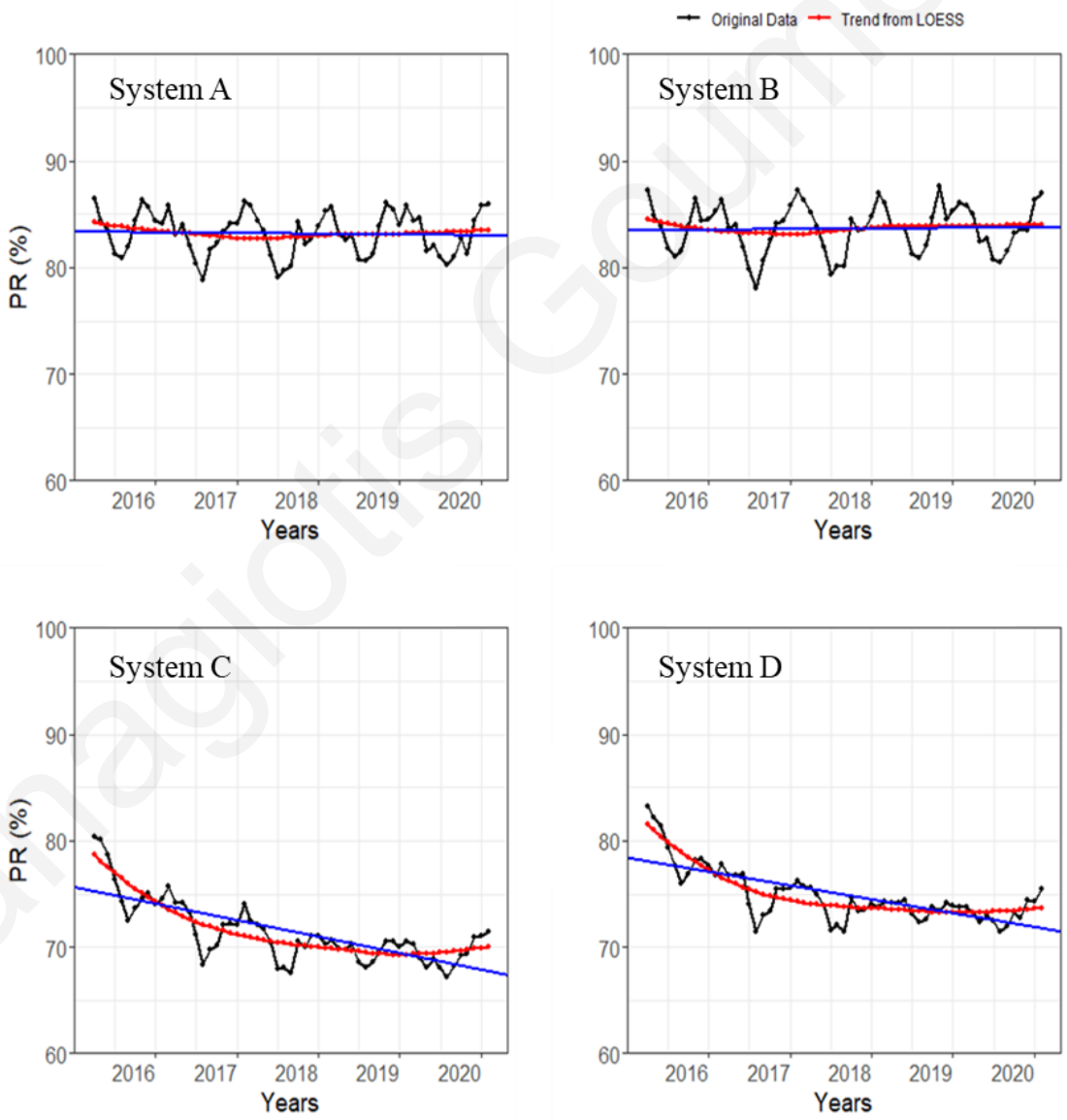

Figure 8. LOESS trend on original data with LR on the extracted trend.

4.3. Degradation Rate with RPCA

By splitting the data to five 12-months' time series sections and implementing the RPCA methodology, the results for the four systems are presented in Figure 9. As observed, the systems affected with PID have a bigger area between the first and fifth year of operation, and even between the first two years can be observed that there is a significant difference between the PR. With the use of the methodology explained in subsection 3.3.3. the degradation rate for each system was computed as, R_D = *-0.34 %/yr*, R_D = *-0.20 %/yr*, R_D = *-1.46 %/yr* and R_D = *-1.24 %/yr* for systems A, B, C and D, respectively.

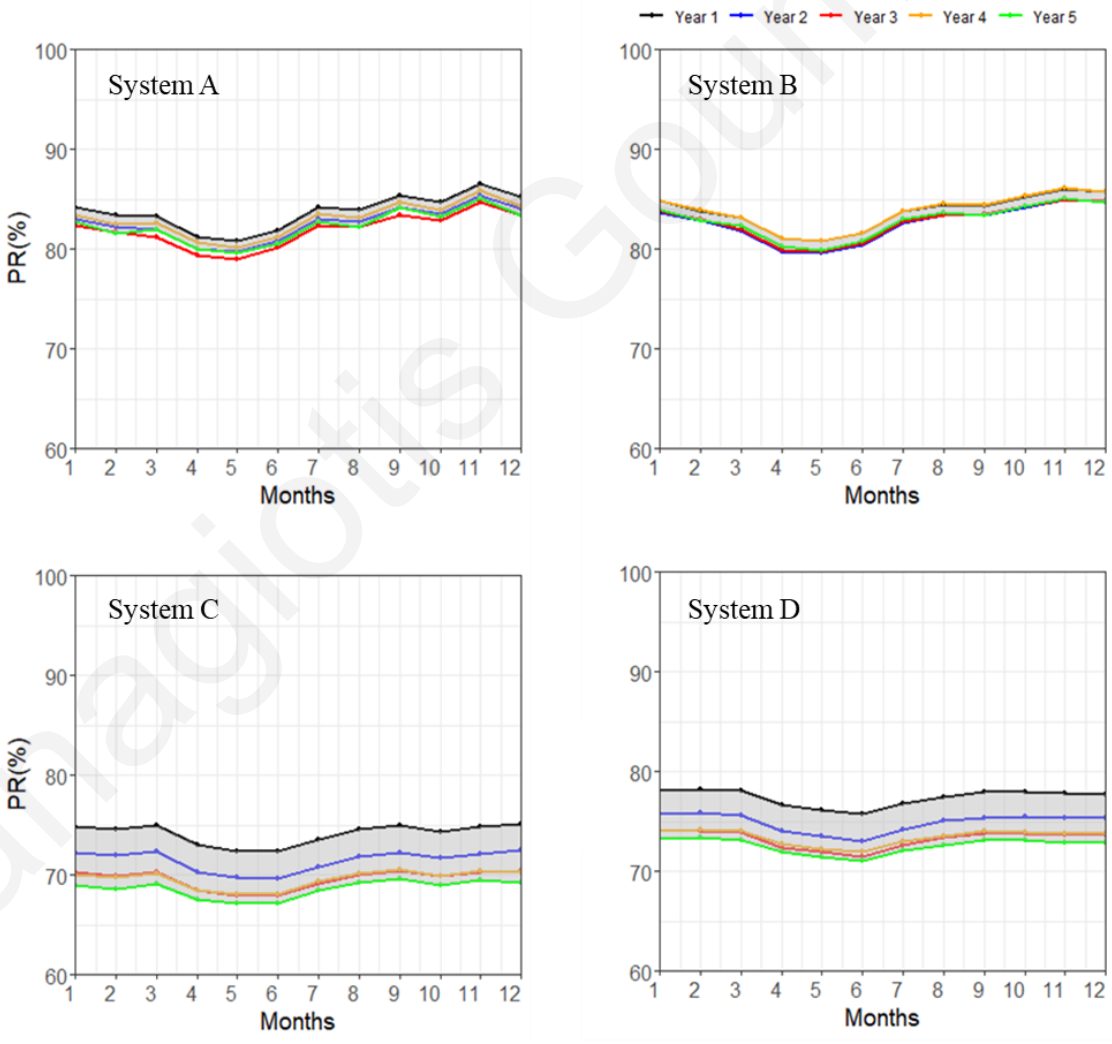

Figure 9. Differentiation on yearly data over the span of 12 months of PR time series.

4.4. Comparison of analysis methods

For purposes of comparing the three methods and the results obtained by them for the four studied systems Figure 10 was constructed. The representation is made with the use of boxplots, so that the uncertainty computed in methods LR, and LOESS could be also presented. The line in the centre of the boxplot demonstrates the median value of the system which is the calculated value from the method without the uncertainty. In the RPCA method no uncertainty was computed and thus it is represented with a straight line. Systems C and D which suffer from PID presented higher degradations rates than the systems that suffered from normal degradation due to aging. LR method expressed the bigger uncertainty values between the methods, although the median value obtained from this method was similar with the LOESS method.

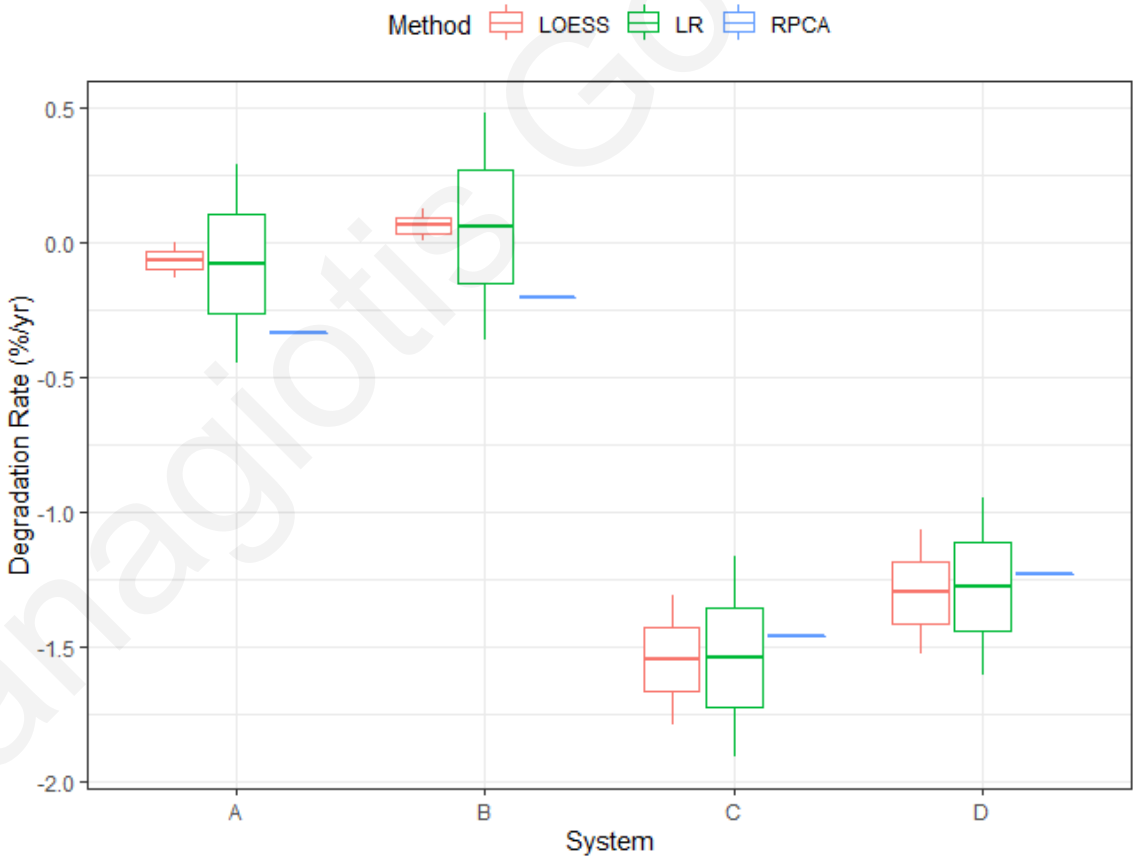

Figure 10. Analysis on different methods for computation of degradation rate in each system.

The results from the RPCA method were closer to the other methods when computing the R_D for systems C and D although for systems A and B the RPCA method computed bigger degradation rate than the other two methods. In general, the three methods computed similar values for each system, with systems affected with PID having more than *-1 %/yr* R_D and systems without having less than *-0.5 %/yr*.

From the beginning by computing the PR for the four systems it can be noted that the affected PID systems, had less PR from the start of their operation, in contrast from the non-affected systems, with values over 85% for systems A and B and less for the other two systems. Furthermore, through methods of LOESS and RPCA, another observation can be made for the systems affected from PID. Through the LOESS method it can be seen that the trend extracted (Figure 8), becomes smoother in the later years of operation and does not have the same decline as the starting years, implying smaller degradation in the later years and thus exhibits that PID affects the system more in the starting years. This assumption can be made with the use of RPCA also (Figure 9). In the PID affected systems, the area between the year and the following year begins to shrink as the years pass. For example, the first year (black line) and the second year (blue line) have a bigger area in between them than the area between years four (orange line) and five (green line). Thus, both methods imply that PID can be detected, and it is observed in early years of operation, even in the second year of operation. This observation can be validated through the research made by Florides et al. [41], were detection of PID is made from the first years of operation before even *1%* of power was lost.

For further investigation, Figure 11 was constructed. Degradation rate was computed for each system for every year of operation. As it can be observed, in every method systems A and B that are not affected from PID exhibit a stability through the years regarding the degradation rate with the degradation been close to *0 %/yr* to *-0.5 %/yr* in every year. In systems C and D, the degradation rate in the starting years of operation is bigger, having R_D over -3 %/yr in year 2 and decreasing through the years. The value of R_D in the following years start to convergence as years pass, showing that as years go by, the effects of PID are less, although the damage has already been made.

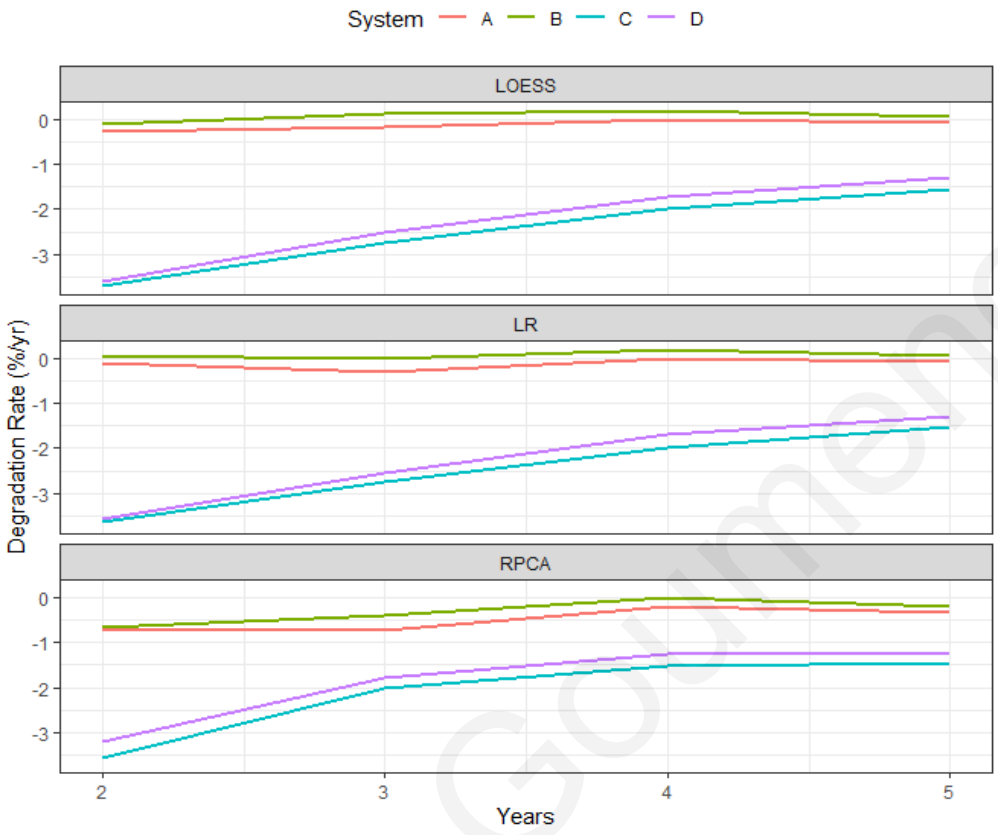

Figure 11. Degradation rate through the years using LOESS, LR and RPCA for each system.

4.5. Degradation Forecasting

For implementing the forecasting of the R_D of the systems, firstly the years of training and testing must be decided. As mentioned in [42], the minimum years of training a dataset for better results regarding the PR time series is 3 years, thus in this thesis results were obtained with the use of 3 years training and 2 years testing and 4 years training with 1 year testing. As aforementioned, SARIMA model was used for forecasting the PR time series and for the computation of the parameters *p*, *q*, *P* and *Q*, the AIC was used. Parameters of the seasonal differencing and differencing *D* and *d* were equal to 1, so that seasonal stationarity is obtained, and seasonality *s* is equal to *12* due to the seasonality of the *12* months of the year. In Table 2, the SARIMA model selected for each system for the cases of 3 training years and 4 training years are presented.

Table 2. SARIMA models for 3 and for 4 years used as training years.

System	SARIMA Model for 3 years training $(p,d,q)\ (P,D,Q)_s$	SARIMA Model for 4 years training $(p,d,q)\ (P,D,Q)_s$
A	$(0,1,1)\ (0,1,0)_{12}$	$(0,1,1)\ (1,1,0)_{12}$
B	$(1,1,0)\ (0,1,0)_{12}$	$(1,1,0)\ (1,1,0)_{12}$
C	$(1,1,0)\ (0,1,0)_{12}$	$(1,1,0)\ (1,1,0)_{12}$
D	$(0,1,1)\ (0,1,1)_{12}$	$(1,1,0)\ (0,1,1)_{12}$

After obtaining the results of both 3- and 4-years forecasting, the selection of one of those models was made for the continuation of the thesis. For selecting the most accurate model, performance metrics RMSE and MAE were utilized. In Figure 12, the RMSE and MAE of each system for 3 years training and 4 years training are presented. As observed, with the use of 4 years for training the model the percentage of error regarding MAE and RMSE was reduced in all systems. Specifically, in terms of systems affected with PID, the increase in training years the RMSE and MAE almost decreased by half of their value, from *3.62%* to *1.39%* RMSE and *3.25%* to *1.19%* MAE for system C, and from *4.24%* to *2.19%* RMSE and *3.83%* to *1.88%* MAE for system D. For systems A and B, there was also exhibited a decrease, although not in that scale, from *1.89%* to *1.25%* RMSE and *1.57%* to *0.96%* MAE for system A and from *1.57%* to *1.06%* RMSE and *1.25%* to *0.95%* MAE. Based on these results, in the next step, the SARIMA model obtained from 4 years training the model was utilized.

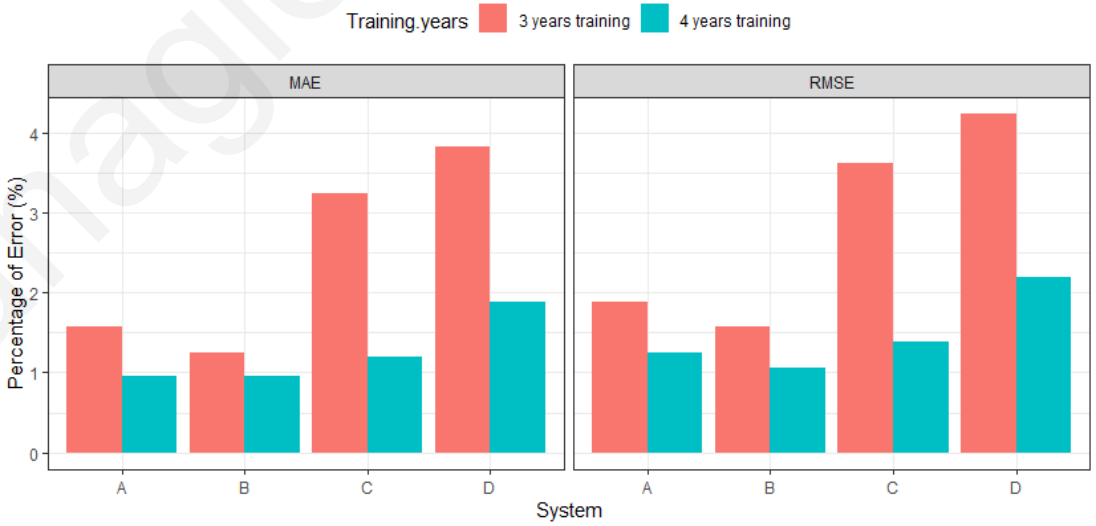

Figure 12. MAE and RMSE results for each system for 3 and 4 years of training data.

For the computation of the R_D obtain on the fifth year of the forecasted data, the methodologies introduced previously (LR, LOESS, RPCA) were used. The results of the computed degradation rates are presented in Table 3. The Actual R_D computed with each method is indicated, which is the R_D of the actual data as computed in subsection 4.1., 4.2. and 4.3., then the Forecasted R_D is demonstrated for each method as computed by forecasting the 5^{th} year of operation and the third column of each method indicates the difference between the Actual R_D and the Forecasted R_D in %/yr. As it can be observed, the difference between them on all occasions is relatively low, signifying the accurateness of the model. Specifically, the biggest difference is observed in the RPCA method on system D with *0.39 %/yr* and the smaller difference occurred in the LOESS method on system A with *0.001 %/yr*. As it can be noticed, the difference of the systems affected with PID is bigger than those with no affection, having an average difference of *0.17 %/yr*, and *0.37 %/yr* for systems C and D, respectively and *0.02 %/yr* and *0.06 %/yr* for systems A and B, respectively.

From these results, the assumption that was made is that the accuracy of the forecasting model depends significantly on the magnitude of the degradation that the system exhibits and thus, if the system suffers from PID or not. Although, system A has exhibited larger degradation than system B, the first one has a smaller difference than the second one in two out of the three methods, and similarly system C has exhibited smaller difference than system D in all methods. Thus, it is not only the occurrence of larger degradation rate that effects the system's accuracy but also the presence of PID that makes the forecasting more unpredictable. Better forecasting could have been made for a larger time series, as observed in Figure 11, the degradation rate through the years for the PID affected systems convergence as years passed, and, as it can be seen on Figure 12, 1 more year of training the model dropped the percentages of error to almost *50%* less. So, it can be noted that, for forecasting a PID affected system, the minimum amount of training years should be more than for a normal degradation through aging affected system, in such way that more accurate results could be obtained.

Table 3. Results of forecasted R_D regarding all methods and the difference that occurs between the forecasted and the actual R_D.

System	LR Actual R_D (%/yr)	LR Forecast R_D (%/yr)	LR Difference (%/yr)	LOESS Actual R_D (%/yr)	LOESS Forecast R_D (%/yr)	LOESS Difference (%/yr)	RPCA Actual R_D (%/yr)	RPCA Forecast R_D (%/yr)	RPCA Difference (%/yr)
A	-0.081	-0.077	0.004	-0.069	-0.068	0.001	-0.34	-0.29	0.05
B	0.058	-0.012	0.07	0.063	-0.01	0.073	-0.2	-0.23	0.03
C	-1.54	-1.72	0.18	-1.55	-1.75	0.2	-1.46	-1.6	0.14
D	-1.28	-1.62	0.34	-1.3	-1.66	0.36	-1.24	-1.63	0.39

4.6. PID Detection method

After the research made in the different degradation mechanisms, a method for detecting PID on small scale systems was developed. In Figure 13, the three relations that are used to determine whether a system is affected by PID are presented. Those three relations are, Voltage over Power, Power over Time, and Voltage over Time. In Figure 13 a system without (a) and a system with (b) PID are presented. Comparing the two systems and validating with the other two systems and several others that were labelled as PID affected and not, assumptions based on those relations were made with the use of the LR model.

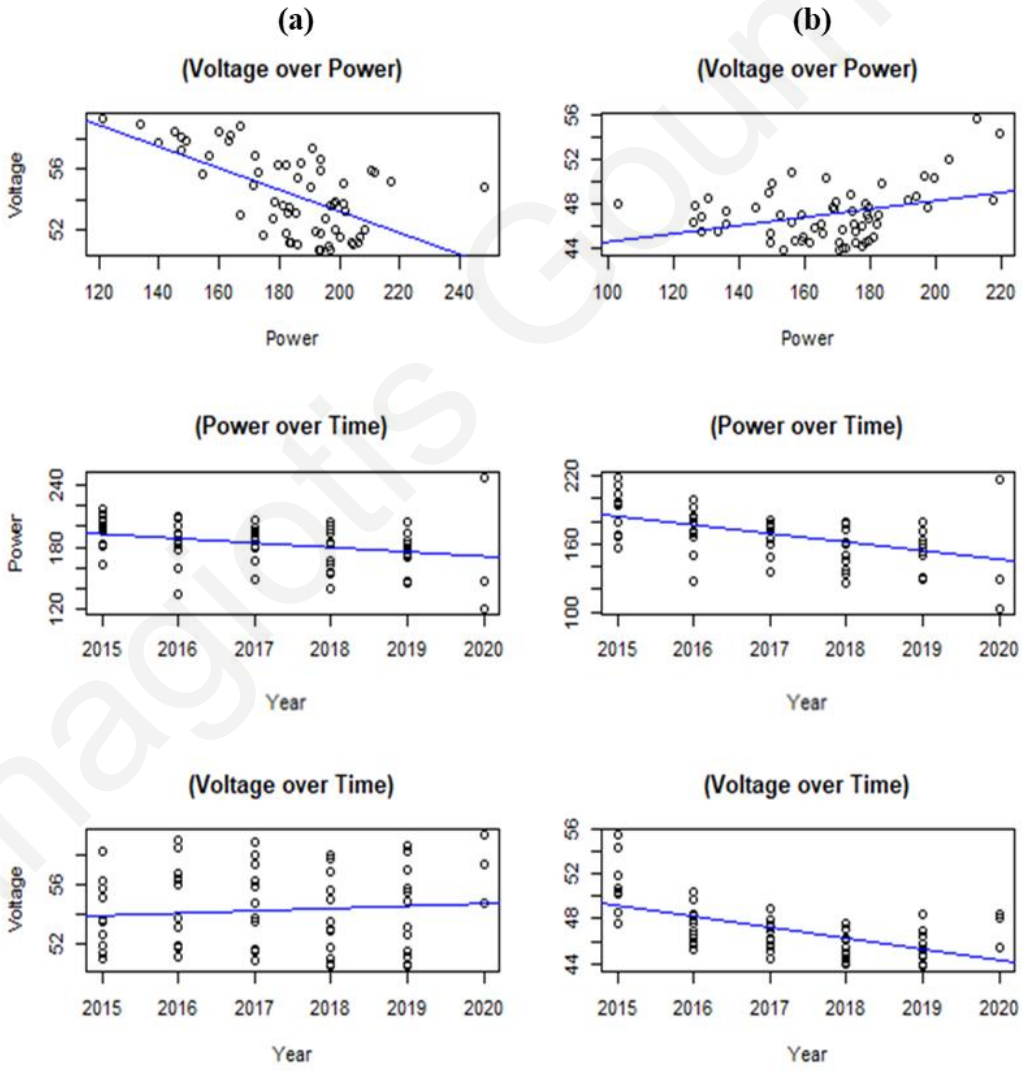

Figure 13. Detection of PID relation methods with the use of LR on a system (a) affected by aging (b) affected by PID.

Based on the first relation regarding the voltage and power, it was observed that if the system suffers from PID, then the slope that is created with the use of LR is positive, in contrast to when it only suffers from aging degradation the slope has a negative sign. Thus, signifying that the decrease in power of the system is mainly affected from the decrease in voltage production, in contrast to the degradation of a non-affected module. The second relation regards the power over time. Both systems exhibited decreased power over time, although as it can be observed, the system that suffers from PID has a bigger decrease through the years than the non-affected. The third and most important relation was the voltage over time. As year progress, in the affected PID system, the voltage shows to decrease through the years in contrast to the system that is not affected.

This method was also examined in smaller timeseries, to observe the accurate operation time needed from the modules so that PID could be detected. With the examination of the datasets from these 4 systems and several others, the correct detection was obtained in the span of 10 months. Thus, this method can accurately detect PID in a small-scale system in an operation time of 10 months. For detecting PID in a smaller period of time, PR_{corr} was used. If the other 3 relations are true and the PR_{corr} has a negative slope with the use of LR in the timeseries, then PID can be detected shorter time. In the systems examined, this method made the detection time decrease from 10 months' time to 4 months. The plot of PR_{corr} on months with LR is presented in Figure 14, in which 14(a) presents a system affected only from aging and 14(b) presents a PID affected system.

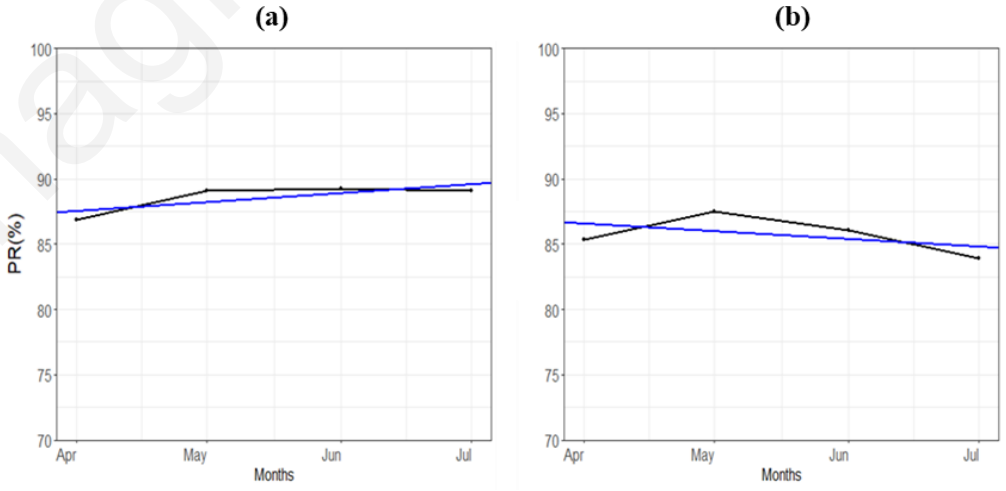

Figure 14. PR_{corr} plotted over 4 months with LR on a system (a) affected by aging (b) affected by PID.

Furthermore, PID detection was utilized with the use of boxplots. Boxplots were constructed for yearly data considering the voltage of the modules. The boxplot represents the maximum, minimum and median value of voltage per year. Figure 15 presents the boxplot representation of two modules, (a) a module which PID did not affect and (b) a module which suffers from PID. As it can be observed, PV modules affected by PID, have values of voltage that decrease over time and their distribution is smaller. In year 2015, in the PID affected system, the maximum voltage reaches *56 V*, having a median value of *51 V*, this value is decreased year by year in contrast to the not affected system on which the voltage remains in similar and in some years more than the year before. The median value of each year is compared to the median value of the voltage of the following year. If the median value is bigger than the following value and this is observed in each year, then PID is detected. In the PID affected systems, considering the 5-year period of operation, a linear voltage decline is observed, with magnitude *-1.07 V/yr* for system C and *-0.88 V/yr* for system D. In 2019 the PID effects are less than the previous years, and if that year is ignored and the first 4 years are considered, then the magnitude of the linear decline is *-1.28 V/yr* for system C and *-1.1 V/yr* for system D. In this method, year 2020 was not considered due to the small number of data that contained (only 3 months).

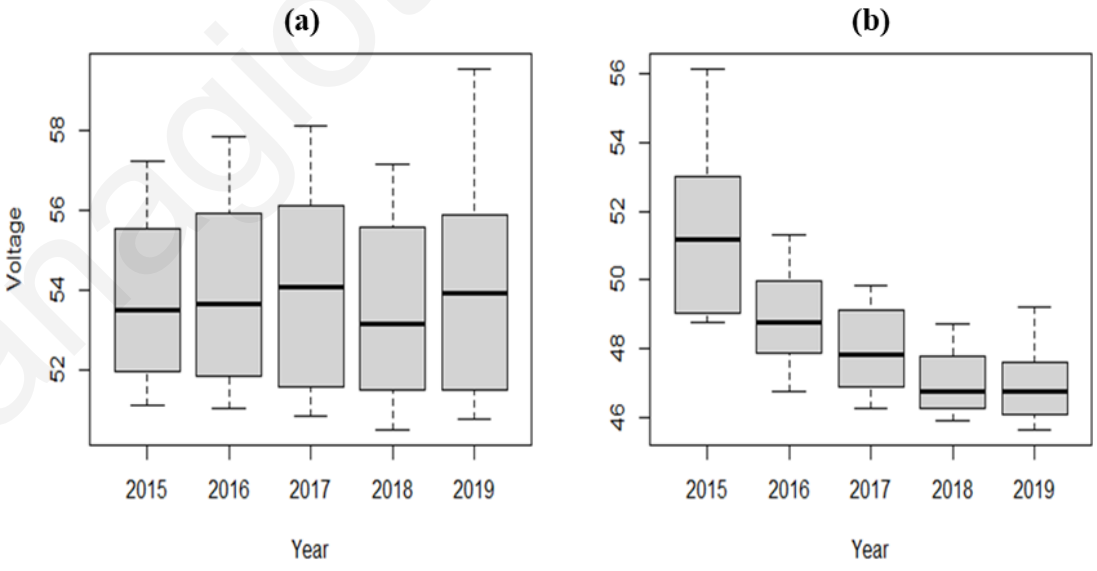

Figure 15. Boxplot detection of PID on a system (a) affected by aging (b) affected by PID.

5. Conclusion and Future Work

By examining these two different degradation mechanisms, conclusions were made regarding the degradation rate, forecasting and the electrical parameters in each mechanism. The LR, LOESS and RPCA, yielded similar results. For the modules affected from aging degradation, the R_D computed was lower than *-0.5 %/yr* and for the modules affected from PID the R_D was approximately *-1.5 %/yr* taking into consideration a mean value of the three methods. Further research revealed that PID affected modules exhibit most of the R_D in the starting years of operation, having over *-3 %/yr* after the second year, indicating that PID damages the modules in the first years of operation more than in the following years. Thus, fast detection of PID is necessary. In the modules affected from aging degradation, the effects are the same as years pass, having a stationary degradation of *-0.5 %/yr*. Regarding forecasting of future effects, a PID affected module was observed that needs at least 4 years of training to obtain more accurate results in contrast to 3-year period that an aging degradation module needs. Although, even with a 4-year period of training, the PID affected modules presented a bigger difference in %/yr between the actual degradation and the forecasted degradation than the aging affected modules.

The main purpose of this thesis was to devise a detection methodology for PID which is easy to use and does not need external equipment. Furthermore, the method was examined in different modules that did and did not suffer from PID to study the time period needed for detection. The method made use of the relations between voltage over time, power over time and voltage over power. In the PID affected modules, the voltage over time relation had a decreasing slope in contrast to the voltage over time relation of the non-affected modules. In the case of the power over time relation, both modules exhibited decreasing slope, although in the PID affected module the decrease was observed to be more. In the third relation, the voltage over power, in the PID affected module the voltage was proportional to the power of the module, in contrast to a non-affected in which the relation was inversely proportional. With the use of the LR and the 3 relations aforementioned, the detection method could be made inside the first 10 months of operation, and with the addition of PR_{corr}, the detection method decreased to 4 months. Fast detection is necessary to prevent the effects of PID because prevention is preferred over cure because, if PID is

detected and it is not cured quickly, then the effects are irreversible, and the only solution is to replace the module. Moreover, boxplot method was also proposed for detecting PID on a yearly level in terms of the decreasing voltage exhibited. Linear voltage decrease was observed in the systems with PID, with a decrease of *-1.07 V/yr* for system C and *-0.88 V/yr* for system D. If the more affected years were examined and the last year was not considered due to the less effect of PID, then the voltage decrease per year increases to *-1.28 V/yr* for system C and *-1.1 V/yr* for system D. The systems that were not affected by PID did not exhibit linear voltage drop, in contrast to the affected PID system. Thus, it can be observed that the detection of PID can be focused on electrical parameters such as, voltage and power and if a quicker detection is preferred then PR can be introduced in which for the calculation meteorological parameters such as irradiance and temperature are used.

Further work will be performed to study the effects of other degradation mechanisms such as LID and LeTID. Finally, the detection of different degradation modes in large scale PV systems is proposed as future work. Methods and observations can be implemented that will assist in countering and detecting quickly the occurance on a larger scale.

References

[1] J. H. Wohlgemuth and S. Kurtz, "Using accelerated testing to predict module reliability," *Conference Record of the IEEE Photovoltaic Specialists Conference*, pp. 003601–003605, 2011.

[2] D. C. Jordan and S. R. Kurtz, "Photovoltaic degradation rates - An Analytical Review," *Progress in Photovoltaics: Research and Applications*, Volume 21, Number 1, pp. 12–29, Jan. 2013.

[3] D. C. Jordan and S. R. Kurtz, "Analytical improvements IN PV degradation rate determination," *Conference Record of the IEEE Photovoltaic Specialists Conference*, pp. 2688–2693, 2010.

[4] I. T. Robert B. Cleveland, William S., Jean E. McRae, "STL: A Seasonal-Trend Decomposition Procedure Based on Loess - ProQuest," *Journal of Official Statistics*, 1990. https://www.proquest.com/docview/1266805989?pq-origsite=gscholar&fromopenview=true (accessed Dec. 13, 2021).

[5] EAC, "Tariffs for Commercial and Industrial Use."

[6] A. Phinikarides, N. Philippou, G. Makrides, and G. E. Georghiou, "Performance loss rates of different photovoltaic technologies after eight years of operation under warm climate conditions," *29th European Photovoltaic Solar Energy Conference and Exhibition*, pp. 2664–2668, 2014.

[7] D. C. Jordan, B. Sekulic, B. Marion, and S. R. Kurtz, "Performance and Aging of a 20-Year-Old Silicon PV System," *IEEE Journal of Photovoltaics*, Volume 5, Number 3, pp. 744–751, May 2015.

[8] S. Lindig, I. Kaaya, K. A. Weis, D. Moser, and M. Topic, "Review of statistical and analytical degradation models for photovoltaic modules and systems as well as related improvements," *IEEE Journal of Photovoltaics*, Volume 8, Number 6, pp. 1773–1786, Nov. 2018.

[9] G. Makrides, B. Zinsser, M. Schubert, and G. E. Georghiou, "Performance loss rate of twelve photovoltaic technologies under field conditions using statistical techniques," *Solar Energy*, Volume 103, 2014.

[10] C. R. Osterwald, A. Anderberg, S. Rummel, and L. Ottoson, "Degradation analysis of weathered crystalline-silicon PV modules," *Conference Record of the IEEE Photovoltaic Specialists Conference*, pp. 1392–1395, 2002.

[11] C. R. Osterwald, J. Adelstein, J. A. Del Cueto, B. Kroposki, D. Trudell, and T. Moriarty, "Comparison of degradation rates of individual modules held at maximum power," *Conference Record of the 2006 IEEE 4th World Conference on Photovoltaic Energy Conversion, WCPEC-4*, Volume 2, pp. 2085–2088, 2006.

[12] H. T. C. Pedro and C. F. M. Coimbra, "Assessment of forecasting techniques for solar power production with no exogenous inputs," *Solar Energy*, Volume 86,

Number 7, pp. 2017–2028, Jul. 2012.

[13] 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 Renewable Power Generation*, Volume 11, Number 10, pp. 1245–1252, Aug. 2017.

[14] H. F. Barreto Miranda, L. P. Da Costa, S. O. Soares, and J. V. Da Silva, "Potential induced degradation (PID): Review," *2020 IEEE PES Transmission and Distribution Conference and Exhibition - Latin America, T and D LA 2020*, Sep. 2020.

[15] P. Hacke, M. Kempe, J. Wohlgemuth, J. Li, and Y.-C. Shen, "Potential-Induced Degradation-Delamination Mode in Crystalline Silicon Modules: Preprint," Accessed: Oct. 27, 2021. [Online]. Available: www.nrel.gov/publications.

[16] P. Hacke *et al.*, "Characterization of Multicrystalline Silicon Modules with System Bias Voltage Applied in Damp Heat," 2010, Accessed: Oct. 27, 2021. [Online]. Available: http://www.osti.gov/bridge.

[17] S. Pingel *et al.*, "Potential induced degradation of solar cells and panels," *Conference Record of the IEEE Photovoltaic Specialists Conference*, pp. 2817–2822, 2010.

[18] P. Hacke *et al.*, "Testing and analysis for lifetime prediction of crystalline silicon PV modules undergoing degradation by system voltage stress," pp. 1–8, Oct. 2014.

[19] A. Phinikarides, N. Kindyni, G. Makrides, and G. E. Georghiou, "Review of photovoltaic degradation rate methodologies," *Renewable and Sustainable Energy Reviews*, Volume 40, pp. 143–152, Dec. 2014.

[20] A. Kaul, S. Pethe, and N. G. Dhere, "Long-term performance analysis of copper indium gallium selenide thin-film photovoltaic modules," *Https://doi.org/10.1117/1.JPE.2.022005*, Volume 2, Number 1, p. 022005, Nov. 2012.

[21] R. Swanson *et al.*, "the Surface Polarization Effect in High- Efficiency Silicon Solar Cells," pp. 4–7.

[22] W. Luo *et al.*, "Potential-induced degradation in photovoltaic modules: A critical review," *Energy and Environmental Science*, Volume 10, Number 1, pp. 43–68, 2017.

[23] H. V. Barreto Câmara, H. Ferreira Barreto Miranda, and J. V. Da Silva, "Recovery of modules affected by potential induced degradation (PID) applying a reverse potential in the laboratory without temperature control," *2019 IEEE PES Conference on Innovative Smart Grid Technologies, ISGT Latin America 2019*, Sep. 2019.

[24] M. Schwark *et al.*, "Investigation of potential induced degradation (PID) of solar modules from different manufacturers," *IECON Proceedings (Industrial Electronics Conference)*, pp. 8090–8097, 2013.

[25] M. A. Islam, M. Hasanuzzaman, and N. A. Rahim, "A comparative investigation on in-situ and laboratory standard test of the potential induced degradation of crystalline silicon photovoltaic modules," *Renewable Energy*, Volume 127, pp. 102–113, Nov. 2018.

[26] S. Pingel, S. Janke, S. Pingel, S. Janke, and O. Frank, "Recovery Methods for Modules Affected by Potential Induced Degradation (PID)."

[27] J. Kapur, K. M. Stika, C. S. Westphal, J. L. Norwood, and B. Hamzavytehrany, "Prevention of potential-induced degradation with thin ionomer film," *IEEE Journal of Photovoltaics*, Volume 5, Number 1, pp. 219–223, Jan. 2015.

[28] H. Nagel, A. Metz, and K. Wangemann, "Crystalline si solar cells and modules featuring excellent stability against potential-induced degradation," *26th European Photovoltaic Solar Energy Conference and Exhibition*, Number May 2016, pp. 3107–3112, 2011.

[29] A. Livera *et al.*, "Data processing and quality verification for improved photovoltaic performance and reliability analytics," *Progress in Photovoltaics: Research and Applications*, 2020.

[30] T. Dierauf, A. Growitz, S. Kurtz, and C. Hansen, "Weather-Corrected Performance Ratio Jose Luis Becerra Cruz Weather-Corrected Performance Ratio," Number April, 2013.

[31] D. C. Jordan, C. Deline, S. R. Kurtz, G. M. Kimball, and M. Anderson, "Robust PV Degradation Methodology and Application," *IEEE Journal of Photovoltaics*, Volume 8, Number 2, pp. 525–531, 2018.

[32] G. Makrides, M. Theristis, J. Bratcher, J. Pratt, and G. E. Georghiou, "Five-year performance and reliability analysis of monocrystalline photovoltaic modules with different backsheet materials," *Solar Energy*, Volume 171, Number June, pp. 491–499, 2018.

[33] R. J. Hyndman and Y. Khandakar, "Automatic Time Series Forecasting: The forecast Package for R," *Journal of Statistical Software*, Volume 27, Number 3, pp. 1–22, Jul. 2008.

[34] JCGM, "Evaluation of measurement data-Guide to the expression of uncertainty in measurement Évaluation des données de mesure-Guide pour l'expression de l'incertitude de mesure," 2008, Accessed: Oct. 11, 2021. [Online]. Available: www.bipm.org.

[35] W. S. Cleveland and S. J. Devlin, "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting," *Journal of the American Statistical Association*, Volume 83, Number 403, pp. 596–610, 1988.

[36] 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 Journal of Photovoltaics*, Volume 5, Number 6, pp. 1698–1705, 2015.

[37] R. H. Shumway and D. S. Stoffer, “Time series analysis and its applications : with R examples,” p. 575, 2006.

[38] X. Chang, M. Gao, Y. Wang, and X. Hou, “Seasonal Autoregressive Integrated Moving Average Model for Precipitation Time Series,” *Journal of Mathematics and Statistics*, Volume 8, Number 4, p. 2012, 2012.

[39] B. Lance, *(TOP) Machine Learning with R: Expert techniques for predictive modeling, 3rd Edition*, Volume 34, Number 4. 2020.

[40] S. A. McLeod, “What does a box plot tell you?,” *Simply psychology: https://www.simplypsychology.org/boxplots.html*. .

[41] M. Florides, G. Makrides, and G. E. Georghiou, “Early Detection of Potential Induced Degradation by Measurement of the Forward DC Resistance in Crystalline PV Cells,” *IEEE Journal of Photovoltaics*, Volume 9, Number 4, pp. 942–950, Jul. 2019.

[42] I. Romero-Fiances *et al.*, “Impact of duration and missing data on the long-term photovoltaic degradation rate estimation,” *Renewable Energy*, Volume 181, pp. 738–748, Jan. 2022.