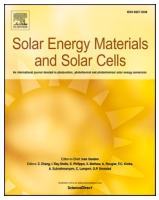
Solar Energy Materials & Solar Cells 274 (2024) 112950

Contents lists available at ScienceDirect

Solar Energy Materials and Solar Cells

journal homepage: www.elsevier.com/locate/solmat



Enhancing solar cell production line monitoring through advanced statistical analysis

Gaia M.N. Javier [a,*], Rhett Evans [b], Thorsten Trupke [a], Ziv Hameiri [a]

[a] The University of New South Wales, Sydney, Australia
[b] 5B, Sydney, Australia

ARTICLE INFO

Keywords:
Lag sequential analysis
Variance
Process control
Production line
Real-time monitoring

ABSTRACT

Efficient monitoring of solar cell performance in high-volume production lines is crucial to ensure consistency and stability. However, this task faces challenges as many manufacturing processes introduce efficiency variations. This study proposes a method, based on lag sequential analysis, to monitor and evaluate these variations. The proposed method is based on the analysis of time-series electrical measurements (such as open-circuit voltage, short-circuit current, fill factor, and efficiency) to identify the degree of randomness, trace process-induced batch variations, and assess line stability. Real-time application of the method can flag anomalies. Furthermore, the suggested method can be extended to image analysis by extracting relevant features from time-series luminescence images, enabling the study of whether cell defects in manufacturing exhibit a random pattern or possess distinguishable characteristics. With its various possible applications, the proposed method has significant potential in enhancing solar cell production line monitoring systems, enabling early identification of production issues and process improvement by manufacturers.

1. Context and motivation

Solar cell production is an intricate process [1] that demands strict adherence to design specifications [2]. However, solar cell production lines, like other manufacturing lines, are vulnerable to variations [3] that can stem from multiple sources, such as equipment operating outside of its specifications [4], inconsistencies in materials [5], and environmental factors [6]. Variations in a production line can be broadly categorised into (1) common cause and (2) special cause variations [7]. Common cause variations are inherent in the production process and occur **randomly**, while special cause variations originate from specific factors, such as equipment drifts or operator errors. Common cause variations are inevitable but predictable. In contrast, special cause variations are erratic and difficult to foresee. Therefore, identifying random variations is often the preferred approach for the optimisation of production processes [7].

In large-scale solar cell production, solar cells are classified based on their performance at the end of the line [8,9]. Variations in manufacturing processes can impact the quality and production costs if left untracked. Therefore, implementing robust monitoring and control measures [10] throughout the production process is essential to minimise variations and ensure consistent and high-quality solar cells. By doing so, manufacturers can increase efficiency, reduce waste, and produce reliable solar cells that meet industry standards [11].

End-of-line characterisation techniques in solar cell manufacturing include current-voltage (I–V) measurements [12] and electroluminescence (EL) imaging [13] [in some cases, photoluminescence (PL) imaging [14]]. Current approaches for analysing these measurements have their own limitations. Manually assessing these measurements for abnormalities or deviations from the mean performance [15], is often subjective and prone to errors. Time-averaging [16], a common statistical method, involves calculating the mean performance for certain time intervals and identifying long-term trends. However, this method may not reliably detect non-random short-term fluctuations [17]. Statistical control charts involve mapping measurements and checking whether they fall within acceptable limits, which can be prone to false alarms [18]. With the advent of machine learning, new approaches have emerged for anomaly detection in production lines [19,20]. However, the effectiveness of these methods often depends on the data available for training [21]; they may also have limitations when facing new types of irregularities. Although autocorrelation [22] can detect previously unseen anomalies, it falls short in quantifying the proportion of random

* Corresponding author.
E-mail address: g.javier@unsw.edu.au (G.M.N. Javier).

https://doi.org/10.1016/j.solmat.2024.112950
Received 30 October 2023; Received in revised form 15 April 2024; Accepted 17 May 2024
Available online 13 June 2024

and non-random variations. Note that most of these methods have primarily focused on one-dimensional data. Indeed several studies [23–25] have delved into the analysis of images to evaluate spatial features, such as cell defects. However, none of them have specifically investigated the randomness associated with these defects.

This study seeks to address the limitations of existing methods by proposing an alternative approach to monitor variations in solar cell production lines. The objectives of this study are: (1) to develop a simple and reliable metric for evaluating critical aspects of production lines and (2) to evaluate the feasibility of this metric through simulations and actual measurements.

2. Theory

2.1. Background concepts

In this study, two key concepts are used: (1) lag sequential analysis [26] and (2) the variance sum law [27].

The lag sequential analysis is a common method for analysing time series data [26]. This approach involves creating multiple datasets by shifting the original measurements at specific time intervals. Lag differenced datasets are calculated by subtracting the shifted data from the original data. In the example shown in Fig. 1, the lag-1 differenced data is generated by shifting the original data by one-time stamp and subtracting the result from the original data. This process yields a new dataset that reflects the change between adjacent periods. Similarly, to produce the lag-2 differenced data, the original data is shifted by two-time stamps instead of one. These lag differenced datasets provide an overview of the cell-to-cell variation [28] in a production line. In the case of stable solar cell performance (i.e., when the various electrical parameters remain constant over time), the lag differenced datasets should ideally display zero values for all their elements.

The variance sum law [27] is a useful tool to investigate the relationship between the variance (σ^2) of datasets. In the case of two datasets (A, B), the variance of their difference [σ^2(A-B)] is given by:

$$\sigma^2(A-B)=\sigma^2(A)+\sigma^2(B)-2\cdot Cov(A,B) \quad (1)$$

where σ is the standard deviation and Cov is the covariance. Equation (1) considers both the individual variances of A and B and their covariance, which indicates the degree to which the two datasets vary together. If there is no linear relationship between A and B (Cov = 0), Equation (1) is simplified to:

$$\sigma^2(A-B)=\sigma^2(A)+\sigma^2(B). \quad (2)$$

Equation (2) indicates that the variance of the difference between two statistically independent datasets is simply the sum of their variances. This equation is particularly relevant to understanding statistical independence, and its implications will be further explored below.

2.2. The Rhett factor

In this study, a metric named the Rhett factor (RF) was developed by integrating principles from the lag sequential analysis and the variance sum law. Substituting A as the original dataset (X) and B as the lagged dataset ($X_{lag,n}$) into Equation (2) yields:

$$\sigma^2(X-X_{lag,n})=\sigma^2(X)+\sigma^2(X_{lag,n}). \quad (3)$$

If there is no systematic variability introduced by lagging X, then $\sigma^2(X)$ is approximately equal to $\sigma^2(X_{lag,n})$. This means that Equation (3) can be further simplified to:

$$\sigma^2(X-X_{lag,n})=2\cdot\sigma^2(X). \quad (4)$$

The RF is then defined as:

$$RF_n=\frac{\sigma^2(X-X_{lag,n})}{2\cdot\sigma^2(X)} \quad (5)$$

The RF at lag-n is denoted as RF_n, while the variances of the lag-n differenced dataset and the original dataset are represented by σ^2(X-$X_{lag,n}$) and σ^2(X), respectively.

The RF is a quantitative measure that captures the extent to which a time series deviates from randomness. This is accomplished by considering both the random and non-random components of the variation. Specifically, an RF of unity for a given lag indicates that the data is statistically independent at that lag, implying that the mean of the data does not exhibit any systematic shifts and that any variation observed at that lag is purely random. The RF values at different lags can be calculated to generate an RF vs lag graph as shown in Fig. 2. This graph shares similarities with a variogram [29], commonly employed in geostatistics for assessing spatial information and in process control for assessing temporal information [30]. The RF vs lag graph normalises the variogram to identify the statistical batch, which will be discussed further in this section. Furthermore, the normalised values at different lags

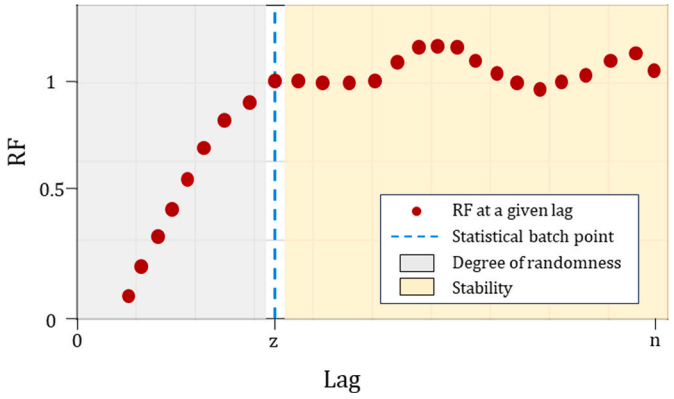


Fig. 2. An illustration of an RF vs lag graph.

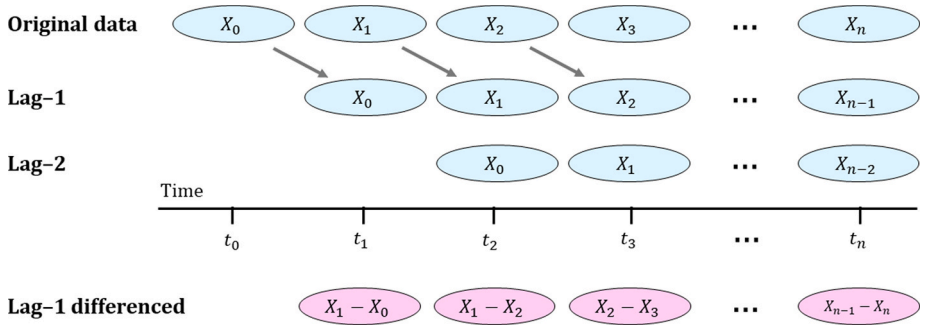


Fig. 1. The general framework for the lag sequential analysis.

quantify the proportion of random versus non-random variations.

Valuable insights into three aspects of a production line can be gained from an RF vs lag graph. Firstly, the statistical batch point, z (in blue), represents the first lag where the RF reaches unity. In the context of a production line, this is the batch size at which cells become statistically independent of each other. When noise is present in the data, the statistical batch point may not reach exactly one. Hence, additional conditions must be established to accurately determine the appropriate batch point. The second aspect, represented by the grey region prior to the statistical batch point, provides information on the proportion of random and non-random data. For instance, an RF value of 0.8 for a given lag indicates that 80% of the variance is random, while 20% is non-random. The final aspect, represented by the yellow region after the statistical batch point, is a qualitative measure of changes in the production line's variance relative to the mean. While values above unity may indicate a negative correlation between the lagged and original data (as derived from Equation (1)), fluctuations in RF values in this region are more significant than the values themselves. Strongly fluctuating RF values in this region suggest high instability within the production line. An ideal production line has a statistical batch point close to lag-1, high RF values in the region before the statistical batch point, and minimal RF fluctuations in the region after the statistical batch point. Table 1 summarises these aspects.

The RF method offers advantages over traditional approaches for analysing time series data. It effectively captures both non-random and random variations. Furthermore, it identifies changes in the process over time by considering within- and between-batch variability.

3. Methodology

3.1. Simulations

The feasibility of the RF method is evaluated by simulating 10,000 time-series cell electrical measurements across different scenarios as listed in Table 2. In all cases, the base simulation includes normally distributed data, to which additional variations were incorporated. The RF is calculated for lag-1 to lag-2,000.

3.2. Experimental implementation

The RF method was applied to analyse 17,000 I–V measurements and EL images of multi-crystalline silicon (mc-Si) cells from an industrial production line. The electrical measurements contain time-series electrical data such as V_{oc}, short-circuit current (I_{sc}), FF, efficiency, series resistance (R_s), and shunt resistance (R_{sh}). The RF values from lag-1 to lag-2,000 were calculated for each of these parameters.

A dynamic analysis was also conducted on the experimental data by implementing a sliding window. The window size was set to accommodate 10,000 cells across various time stamps. Within each window, the RF vs lag was calculated. The graphs obtained from different time frames were then compared.

The RF method was further expanded to include EL image analysis, wherein the image data was converted into scalar metrics, such as the mean and standard deviation of pixel intensities. The RF vs lag was calculated for various image features. Furthermore, an analysis was conducted to assess how the variance in each image feature correlates with the variance in the overall cell performance.

The above analyses can be categorised into: (1) basic analysis, (2) dynamic analysis, and (3) image-based analysis, as shown in Fig. 3.

Table 2
Case studies evaluating the feasibility of the RF method.

Case No.	Description	Details
1	Completely random data	Efficiency values were generated randomly, without underlying patterns.
2	Data with outliers	Similar to Case 1, however, occasional outliers were introduced.
3	Varying mean data	The mean efficiencies of every 500 cells were randomly varied within a range of ±0.3 %.
4	Data with two sources of variance	The mean open-circuit voltage (V_{oc}) was randomly varied for every 200 cells, while the mean fill factor (FF) was varied for every 500 cells. The resulting efficiency was then calculated.
5	Dynamic data	The efficiency values were initially set to a stable mean. After 5,000 cells, the mean efficiency was randomly varied for every 200 cells.

4. Results and discussion

4.1. Simulation results

Fig. 4 presents the results of the first three case studies on the simulated data. Fig. 4(a) shows the efficiency as a function of time of 10,000 cells from Case 1, which involves completely random data. As expected, the corresponding RF vs lag [Fig. 4(d)] exhibits a statistical batch point at lag-1, which remains constant up to lag-2,000. For Case 2, which examines data with outliers [Fig. 4(b)], most of the efficiency data points are concentrated between 22.5 % and 23.5 %. However, a few outliers with efficiencies as low as 15 % were introduced. The RF vs lag graph [Fig. 4(e)] shows a statistical batch point of unity; however, beyond lag-1, the RF values exhibit fluctuations, demonstrating the influence of outliers on the RF vs lag graph, which can lead to misinterpretations. Therefore, to ensure a comprehensive understanding of the trends in the majority of cells, it is recommended to eliminate outliers before performing the RF analysis. In Case 3, a non-random data scenario is considered; Fig. 4(c) displays the efficiency as a function of time. The efficiency distribution remains within a narrow range of 22.5–23.5 %. However, after certain batches of cells, a visible mean shift occurs. The corresponding RF vs lag graph [Fig. 4(f)] indicates a statistical batch point of 500 (RF = 1.01). At lag-1 (grey region), RF = 0.9, indicating that 90 % of the variance is already random and the remaining 10 % is related to the mean shifts. In terms of stability (yellow region), the RF value remains relatively stable at unity from lag-500 to

Table 1
Production line insights obtained from an RF vs lag graph.

Insight	Description	Physical meaning	Ideal scenario
1 – statistical batch point (z)	The first lag where the RF reaches unity NOTE: Further considerations are necessary when analysing data that contains noise or uncertainties	The batch size at which cells are statistically independent of each other	A lower statistical batch point is preferred as it means that the variance among cells is more randomly distributed
2 – degree of randomness	The region before the statistical batch point	Indicates the level of randomness at a given lag, and the remainder (1-RF) is related to the non-random factors (e.g., mean shifts)	High RF values (above 0.9) are desirable as they indicate a smaller proportion of variance related to non-random factors
3 – stability	The region after the statistical batch point	A qualitative measure of the changes in the production line's variance in relation to the mean	Small fluctuations are preferred as they suggest a stable production line with minimal deviations from the mean

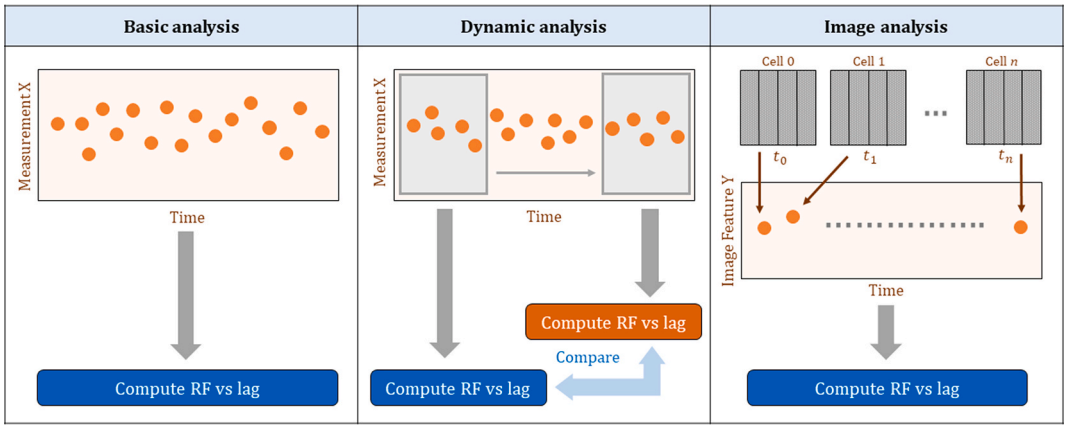


Fig. 3. Different types of analysis using the RF method on the industrial measurements.

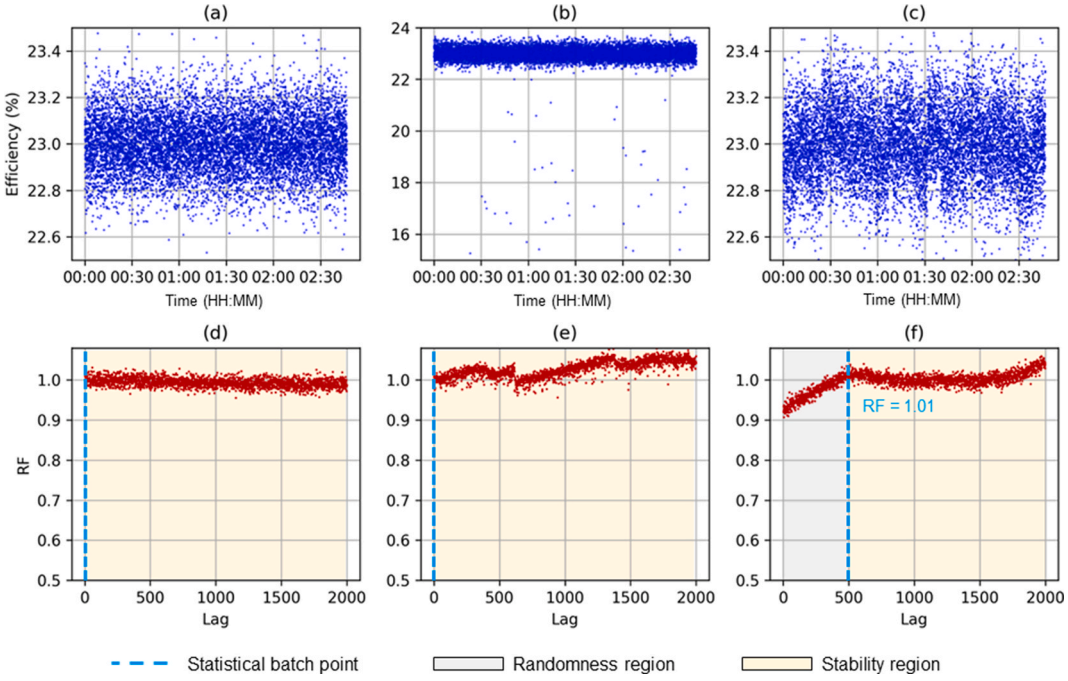


Fig. 4. Simulation results of (a) random data, (b) data with outliers, and (c) non-random data with varying mean. Graphs (d–f) are the corresponding RF vs lag plots.

lag-1,500. However, it then increases to 1.04 from lag-1,500 to lag-2,000. The fact that the RF does not remain at unity indicates instability in the line. Decisions regarding the significance of this instability and the required actions should be defined by the manufacturers.

Fig. 5 presents the results for Case 4, which considers two sources of variance. In Fig. 5(a), the mean V_{oc} for every 200 cells was shifted within ±3 mV. As expected, the corresponding RF vs lag plot, shown in Fig. 5 (d), exhibits a statistical batch point of 200. The randomness within the batches of 200 cells is above 90%, with the remaining 10% accounting for the mean shifts. After the statistical batch point, RF fluctuates only slightly between 0.97 and 1.02. Fig. 5(b) illustrates FF vs time. In this case, the mean FF was varied for every 500 cells, with a mean shift within ±0.3 %. As expected, the RF vs lag plot in Fig. 5(e) indicates a statistical batch point of 500. Similar to the V_{oc} analysis, the degree of randomness within the statistical batches is high (>90 %), and the stability between batches exhibits minimal fluctuations. The resulting efficiency, considering both sources of variance, is shown in Fig. 5(c), with values ranging from 22.5% to 23.5%. Some similarities can be observed between the FF and efficiency curves, such as the dip in value around the 02:35 time stamp (indicated by the orange boxes). However, the RF vs lag plot in Fig. 5(f) reveals that the statistical batch point of the efficiency better aligns with the statistical batch point of the V_{oc} (i.e., 200). Further analysis indicates that the result for the efficiency reflects a consolidated batch effect, encompassing the weighted average of all individual batch effects observed. Given the relatively stronger influence of V_{oc} in this case, the statistical batch point is determined primarily by this parameter. Correlation analysis was utilised to validate this conclusion. Indeed, the correlation coefficient between V_{oc} and the efficiency was found to be 0.77 while the correlation coefficient between FF and the efficiency was 0.63 (see the Appendix), confirming this finding. The randomness within batches is consistently high (>90%). Furthermore, the fluctuations observed after lag-500 closely mirror those exhibited by the V_{oc}.

The final simulated case involves dynamic data analysis. Fig. 6(a) presents the efficiency of 10,000 cells, plotted against time. It is important to note that for the first 5,000 cells, random efficiency values

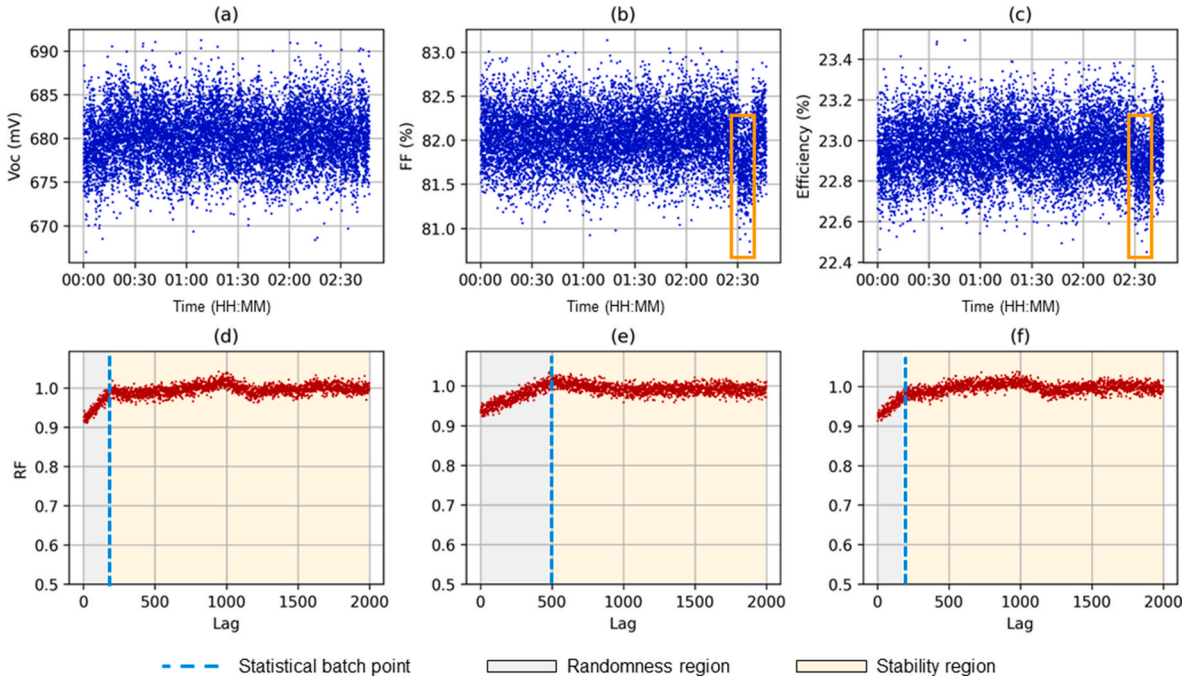


Fig. 5. Simulation results of Case 4: (a) V_{oc}, (b) FF, and (c) the resulting efficiency vs time. Graphs (d–f) are the corresponding RF vs lag plots. The orange boxes indicate the dip in value around 02:35.

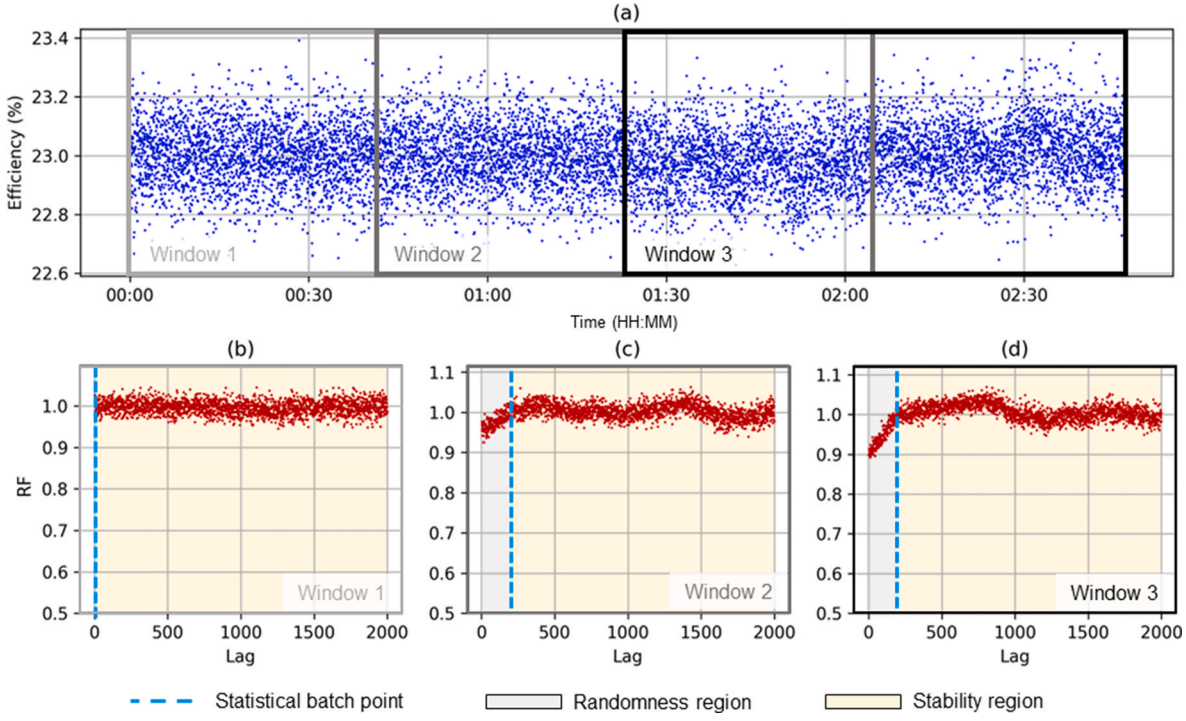


Fig. 6. Simulation results of Case 5: (a) Efficiency vs time with three different window frames and (b–d) corresponding RF vs lag plots for each window frame.

were generated. Consequently, from 00:00 to 01:25, the values followed a normal distribution. However, beyond these initial 5,000 cells, the mean efficiency of every batch of 200 cells was varied. Although not easily observable in the graph, there is a distinct difference in the distribution patterns between the first 5,000 cells and the last 5,000 cells. To further investigate these patterns, three window frames (containing 5,000 cells each) are examined. Window 1 covers the first 5,000 cells, and its corresponding RF vs lag graph is shown in Fig. 6(b). Since this window only covered the random cells, the statistical batch point is at lag-1 and the RF remained constant until lag-2,000. Window 2 includes the latter 50% of the random cells and the initial 50% of the non-random cells. The RF vs lag graph for this window is presented in Fig. 6(c). There

is a shift in the statistical batch point from 1 to 200, demonstrating that even with a window frame that is only 50% non-random, the RF method can still capture variations induced by batch effects. At lag-1, the RF value is approximately 0.95 and the region after the statistical batch point exhibited a relatively stable RF. Lastly, Window 3 contains the last 5,000 cells, and the RF vs lag graph is shown in Fig. 6(d). As expected, the statistical batch point is at 200. At lag-1, the RF value is around 0.90, slightly lower than that of Window 2, indicating a shift in the randomness of the data. Additionally, the same level of RF fluctuations is observed after the statistical batch point, indicating sustained stability.

The results indicate the potential of using the RF analysis along with the sliding window approach to study variations in dynamic data. Note that mean drifts are prevalent in production and with the appropriate window size, this dynamic analysis can potentially flag these drifts. Manufacturers can also establish thresholds for acting based on differences observed in the RF vs lag graphs across different window frames.

4.2. Experimental results

4.2.1. Basic analysis

The above simulation results showed that the presence of outliers can disrupt trends, which might lead to misinterpretations. To mitigate this, an Isolation Forest method [31] with a contamination rate of 2% was utilised to remove outliers from a set of I–V data measured on 17,000 industrial cells. This method has been proven to be effective in managing high-dimensional data and does not require prior knowledge of the characteristics of normal points and outliers [32]. The normalised efficiency (scaled to mean = 0 and variance = 1), after outlier removal, is illustrated in Fig. 7(a). Analysing the graph visually poses considerable challenges in accurately gauging the degree of randomness in the cell measurements, identifying statistical batches, and ultimately assessing production line stability. Fig. 7(b) shows the normalised average efficiency per hour. While this method can identify macro-trends and thus assess line stability, it does not provide insights into batch processing or randomness. Fig. 7(c) displays a statistical control chart [18]. The central line (dashed grey) represents the average at zero while the upper and lower control limits (dashed red) are set at 3 and -3, respectively. This approach facilitates the detection of special cause variations, identified by data points falling below the lower control limit, highlighting non-random variations. However, it does not offer additional insights into batch processing or stability. The RF method is then applied to determine if it can yield better insights.

The corresponding RF vs lag graph is presented in Fig. 7(d). Note that noise is inherent in measurements and can impact the statistical batch point. To address this, in this study, the statistical batch point is defined with an uncertainty range, outlined as follows:

- The lower limit of the range is set at the instance when the RF exceeds 0.99, and there should be no fluctuations in the RF for the subsequent five lags following this threshold. These fluctuations should be within the range of 0.99 ± 0.005.
- Starting from the lower limit, the statistical batch range includes RF values within the range of 1 ± 0.03. The upper limit of this range is defined by either the point at which the RF exceeds 1.03 or falls below 0.97. If the RF remains within these defined limits (1 ± 0.03) for many lags, then the range is restricted to 250 lags from the lower limit.

These definitions can vary depending on the specific manufacturing environment and the nature of the noise or uncertainty involved.

With these definitions, the statistical batch range for the efficiency is at lag-(875 ± 125). Interestingly, the diffusion process [33] in this specific manufacturing line is done in batches of 850 wafers. The region before the statistical batch point (in grey) exhibits an RF ranging from 0.70 to 0.98. In this region, the slope of the RF values is comparatively less steep than what was observed in the simulations. This discrepancy suggests that a higher proportion of variance at lower lag values is linked to mean shifts along with other non-random factors, happening in the measurements compared to simulations. A decreasing trend in RF is observed beyond the statistical batch point (in yellow), suggesting a possible instability in the line.

The obtained efficiency is influenced by various electrical parameters, each of which can be linearly correlated with it. The RF vs lag is calculated separately for V_{oc}, I_{sc}, FF, R_{sh}, and R_s, and compared with the

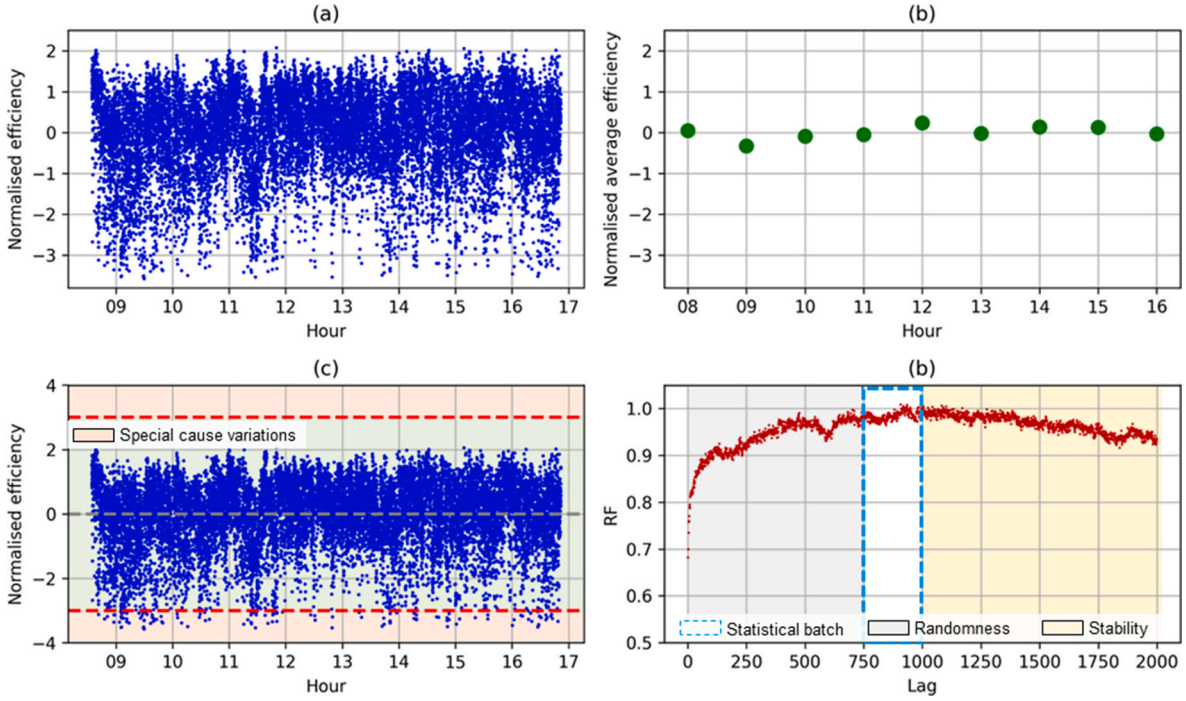


Fig. 7. (a) Normalised efficiency vs time, (b) normalised average efficiency per hour, (c) statistical control chart, and (d) the corresponding RF vs lag.

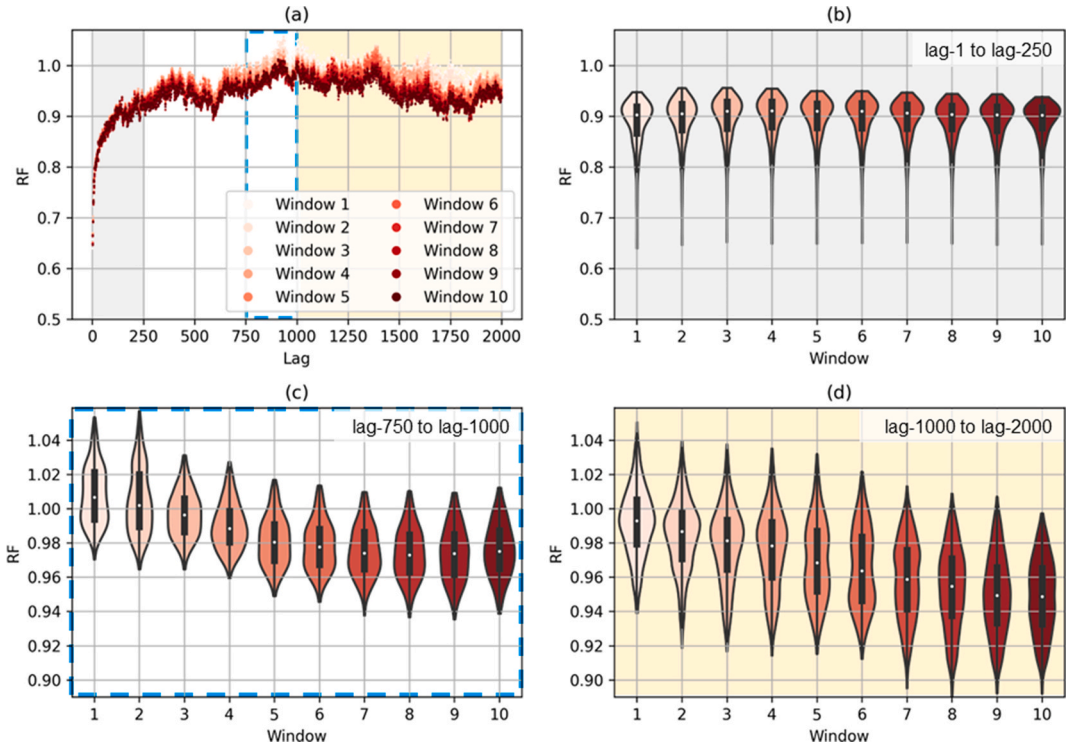


Fig. 8. (a) RF vs lag of windows across various time stamps and the RF distribution for (b) lag-1 to lag-250, (c) lag-750 to lag-1000, and (d) lag-1,000 to lag-2,000.

RF vs lag of the efficiency. Correlation analysis was conducted for two categories: (1) lower lags, providing insights into the primary driver of non-random variations in the efficiency, and (2) higher lags, indicating the primary driver of the instability in the efficiency. It was found that V_{oc} has the strongest correlation in both categories, making it the primary driver of non-random variations and instability of the efficiency. For non-random variations, I_{sc} is ranked second in importance, while R_s is second for the instability in the efficiency. Implementing the RF method on relevant parameters may facilitate root-cause analysis of efficiency variance; however, further investigation is required. Another set of analyses is provided in the Appendix where the RF method is used to investigate an industrial tunnel oxide passivated contact (TOPCon) solar cell manufacturing line.

4.2.2. Dynamic analysis

In the previous sections, the application of RF to analyse static data was demonstrated. To assess the efficiency dynamically, a window containing 10,000 cells was shifted across various time stamps. Within each window, the RF vs lag was calculated, and the resulting RF vs lag for ten consecutive windows is presented in Fig. 8(a). At lower lags (lag-1 to lag-250), no significant difference among the curves from the different windows can be observed. However, at higher lags (lag-1,750 to lag-2,000), higher RF values are observed for earlier windows (Windows 1–2) compared to the later ones (Windows 9–10).

To gain further insights, the RF distribution for different lag regions was analysed: (1) lag-1 to lag-250 [Fig. 8(b)], (2) lag-750 to lag-1,000 [Fig. 8(c)], and (3) lag-1,000 to lag-2,000 [Fig. 8(d)]. At lower lags [Fig. 8(b)], all windows exhibited a mean RF of approximately 0.9, suggesting that around 90% of the variance is attributed to randomness, while the remaining 10% is associated with non-random factors. The statistical batch point was identified to have a range of 875 ± 125 for the majority of the windows, consistent with the results presented in Section 4.2.1. Around this region [Fig. 8(c)], the mean values of the RF hovered around 0.97 to 1.01. However, it is important to note that there is an observed trend of decreasing mean RF for higher window numbers. To fully understand the implications of this trend, further investigation is required. For higher lags [Fig. 8(d)], there also appeared to be a slight reduction in the mean RF as the window was swept. However, since the mean RF value remained within 1 ± 0.05, the production line can still be considered relatively stable.

It is important to note that the window size should be optimised to effectively monitor dynamic variations. Furthermore, the determination of thresholds to decide when action should be taken is left to the discretion of the manufacturer. In this case, it is assumed that the differences observed in the RF vs lag graphs across different time frames are insignificant, thereby indicating that no immediate action is required.

4.2.3. Image analysis

The experimental dataset also included EL images. Fig. 9 shows three representative EL images of cells from the investigated industrial line. These images reveal the presence of defects, manifested as multiple low-intensity areas [34] spread across the cells. Despite the existence of several methods utilised for detecting faulty cells [35], none of these approaches have yet delved into analysing images in the time domain to understand the randomness associated with these observed defects.

Several features were derived from the EL images, including the EL mean intensity. The normalised EL mean intensity vs time is shown in Fig. 10(a) and the corresponding RF vs lag is shown in Fig. 10(b). The statistical batch point for this parameter is 700 ± 100. At lower lags (lag-1 to lag-250), the RF ranges from 0.8 to 0.9, indicating that the contribution of non-random factors to the variance is 10–20%. Moreover, an instability in the RF is observed at higher lags. Further analysis showed a similarity between this graph and the RF vs lag of the V_{oc} [Fig. 10(b) inset]. Note that EL mean intensity and V_{oc} are expected to exhibit a high

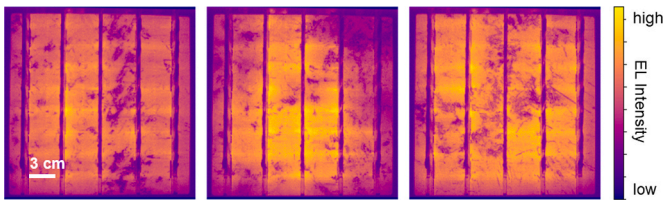


Fig. 9. Representative EL images from the studied industrial line.

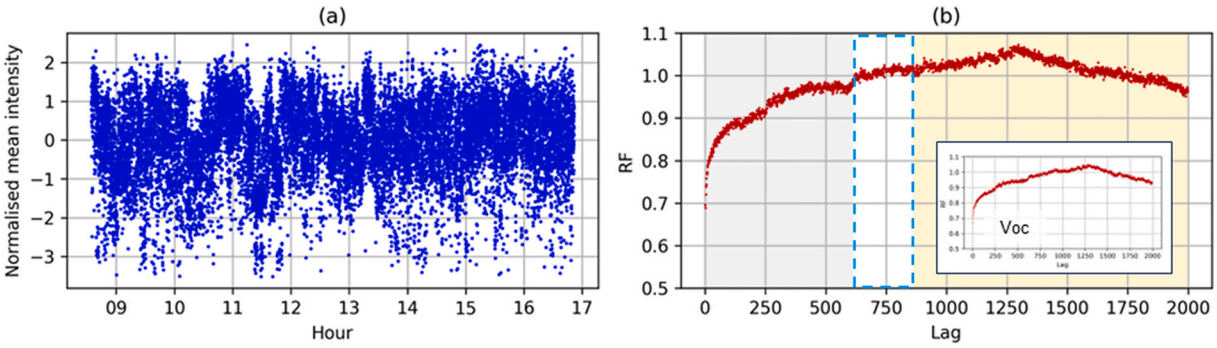


Fig. 10. (a) Normalised mean EL intensity vs time and (b) the corresponding RF vs lag. The inset of (b) shows the RF vs lag for V_{oc}.

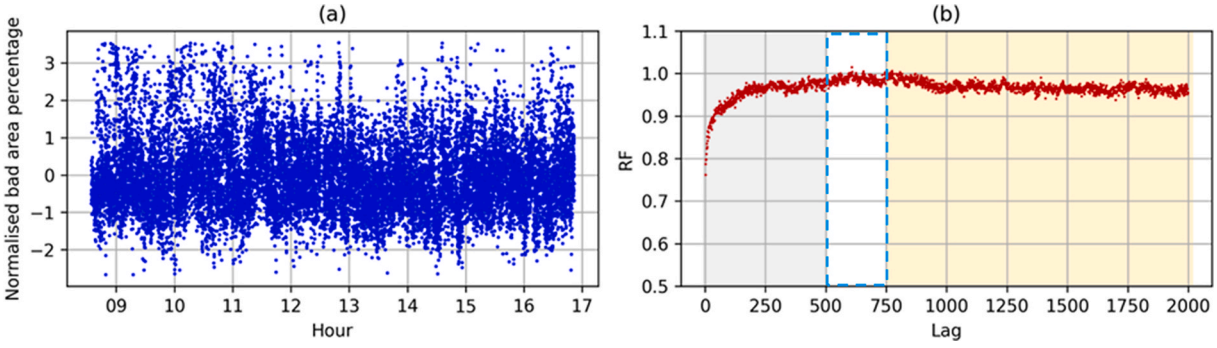


Fig. 11. (a) Normalised dark ("bad") area percentage vs time and (b) the corresponding RF vs lag.

degree of correlation with each other (0.99 in this study). Therefore, by exploring the variability in the mean EL intensity, valuable insights can be gained regarding the variability in V_{oc}.

As defects in cells often manifest as dark areas, a valuable metric for studying their presence is the percentage of dark regions within a cell (excluding the busbars). In this study, a pixel with intensity below 90% of the mean is defined as "dark". Fig. 11(a) shows the normalised dark ("bad") area percentage as a function of time, while Fig. 11(b) shows the corresponding RF vs lag. For this parameter, the statistical batch point is within the range of 625 ± 125. The degree of randomness before the statistical batch range is relatively high since RF exceeds 0.9 as early as lag-50. At higher lags, RF gradually decreases to around 0.97 but no other significant fluctuations are observed after this subtle dip. This suggests that defects in this specific production line are not entirely random, demonstrating that using the RF analysis on time-series images not only correlates image features to variations in electrical parameters but also with the presence of defects along the production line.

5. Conclusion

This study proposed a method for monitoring variations in production lines. The developed method was evaluated on both simulations and actual measurements. Results demonstrate that the method effectively gauges the level of randomness, probable batching, and stability within the production process. Its potential applications include root-cause analysis (the identification of dominant sources of variance) and dynamic analysis (facilitating real-time monitoring of the production line). Additionally, the method can be implemented in conjunction with image analysis to assess spatial information, such as variations in defects. The established capabilities are set to substantially enhance manufacturing process control within solar cell production lines.

CRediT authorship contribution statement

Gaia M.N. Javier: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Rhett Evans:** Writing – review & editing, Methodology, Formal analysis, Conceptualization. **Thorsten Trupke:** Writing – review & editing, Methodology, Formal analysis. **Ziv Hameiri:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgment

This work was supported by the Australian Government through the Australian Renewable Energy Agency [ARENA; Projects 2020/RND016 and 2022/TRAC002]. The views expressed herein are not necessarily the views of the Australian Government, and the Australian Government does not accept responsibility for any information or advice contained herein.

The authors would also like to thank Gordana Popovic from UNSW Stats Central for a fruitful discussion regarding the study.

Appendix

A.1. Correlation analysis for Case 4

Figure A.1(a) presents the FF vs efficiency while Figure A.1(b) illustrates the plot of V_{oc} vs efficiency. The distribution of points in the FF vs efficiency plot appears slightly broader than that in the V_{oc} vs efficiency plot. A correlation coefficient of 0.63 was extracted between the FF and efficiency, while a higher correlation of 0.77 was noted between V_{oc} and efficiency. These correlation values indicate that V_{oc} exhibits a stronger linear relationship with efficiency, supporting the initial observations derived from the RF method.

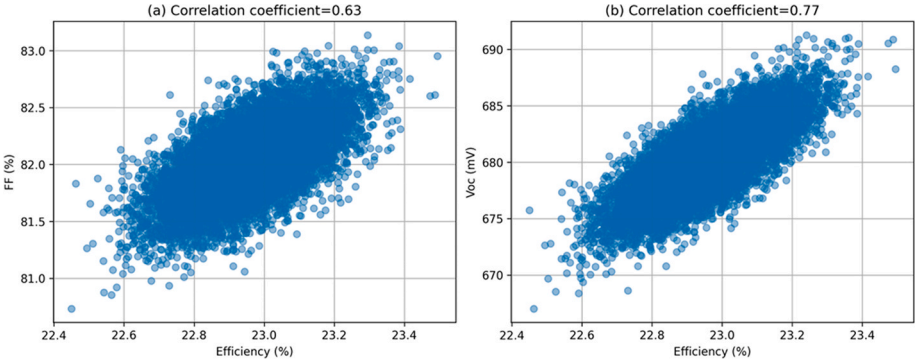


Fig. A.1. Case 4 scatter plots: (a) FF and (b) V_{oc} versus efficiency.

A.2. Application of the RF method in a TOPCon mature production line

Figure A.2(a) shows the normalised efficiency as a function of time for 15,000 cells from an undisclosed industrial TOPCon manufacturing line while Figure A.2(b) presents the corresponding RF versus lag. Compared to the mc-Si line discussed in Section 4.2.1, the TOPCon line has a relatively low statistical batch range (200 ± 125). Furthermore, at lag-10, the RF has already reached 0.95, indicating that even for these small batches, only 5% of the variance stems from non-random factors like mean shifts, while the remaining 95% is due to random factors. Lastly, only minor fluctuations are visible in the yellow region. Hence, this line can be defined as very stable.

The RF method was also applied to other electrical parameters such as the V_{oc}, I_{sc}, and FF. It was found that the FF has the strongest influence on the randomness, batch processing, and stability of cell efficiency, followed by the V_{oc}.

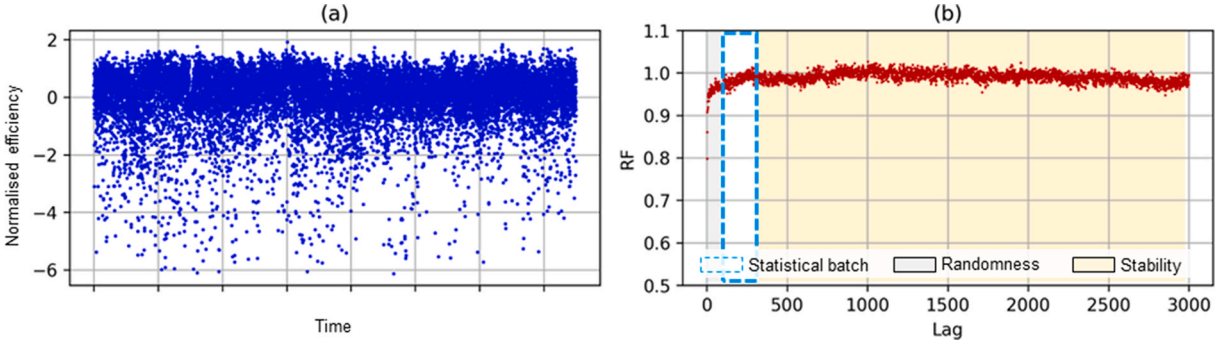


Fig. A.2. (a) Normalised efficiency as function of time and (b) the corresponding RF vs lag from an undisclosed TOPCon manufacturing line.

References

[1] Solar cell production: From silicon wafer to cell. https://sinovoltaics.com/solar-basics/solar-cell-production-from-silicon-wafer-to-cell/.
[2] S. Kurtz, et al., Marrying quality assurance with design engineering – a winning partnership! But, a cultural divide?, in: 44th IEEE Photovoltaic Specialist Conference, 2017, pp. 1275–1279.
[3] Y. Buratti, C. Eijkens, Z. Hameiri, Optimization of solar cell production lines using neural networks and genetic algorithms, ACS Appl. Energy Mater. 3 (11) (2020) 10317–10322.
[4] R. Evans, M. Boreland, Multivariate data analytics in PV manufacturing—four case studies using manufacturing datasets, IEEE J. Photovoltaics 8 (1) (2018) 38–47.
[5] H. Wagner-Mohnsen, P.P. Altermatt, A combined numerical modeling and machine learning approach for optimization of mass-produced industrial solar cells, IEEE J. Photovoltaics 10 (5) (2020) 1441–1447.
[6] M.J. Hossain, et al., A comprehensive methodology to evaluate losses and process variations in silicon solar cell manufacturing, IEEE J. Photovoltaics 9 (5) (2019) 1350–1359.
[7] X. Lei, C.A. MacKenzie, Distinguishing between common cause variation and special cause variation in a manufacturing system: a simulation of decision making for different types of variation, Int. J. Prod. Econ. 220 (2020) 107446.
[8] Y. Buratti, A. Sowmya, R. Evans, T. Trupke, Z. Hameiri, Half and full solar cell efficiency binning by deep learning on electroluminescence images, Prog. Photovoltaics Res. Appl. 30 (3) (2022) 276–287.
[9] P. Kunze, et al., Contactless inline IV measurement of solar cells using an empirical model, Sol. RRL 7 (8) (2022) 2200599.
[10] H. Mohamad, R. Jenal, D. Genas, Quality control implementation in manufacturing companies: motivating factors and challenges, in: Applications and Experiences of Quality Control, InTech, 2011.
[11] C. Ballif, F.-J. Haug, M. Boccard, P.J. Verlinden, G. Hahn, Status and perspectives of crystalline silicon photovoltaics in research and industry, Nat. Rev. Mater. 7 (8) (2022) 8.
[12] K.A. Emery, Solar simulators and I–V measurement methods, Sol. Cell. 18 (3–4) (1986) 251–260.
[13] T. Fuyuki, H. Kondo, T. Yamazaki, Y. Takahashi, Y. Uraoka, Photographic surveying of minority carrier diffusion length in polycrystalline silicon solar cells by electroluminescence, Appl. Phys. Lett. 86 (26) (2005) 262108.

[14] T. Trupke, R.A. Bardos, M.C. Schubert, W. Warta, Photoluminescence imaging of silicon wafers, Appl. Phys. Lett. 89 (4) (2006) 044107.
[15] B. Smith, Six-sigma design (quality control), IEEE Spectrum 30 (9) (1993) 43–47.
[16] pandas.DataFrame.resample — pandas 1.5.2 documentation. https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.resample.html.
[17] D.W. Zimmerman, A note on the influence of outliers on parametric and nonparametric tests, J. Gen. Psychol. 121 (4) (1994) 391–401.
[18] T.M. Margavio, M.D. Conerly, W.H. Woodall, L.G. Drake, Alarm rates for quality control charts, Stat. Probab. Lett. 24 (3) (1995) 219–224.
[19] Z. Kang, C. Catal, B. Tekinerdogan, Machine learning applications in production lines: a systematic literature review, Comput. Ind. Eng. 149 (2020) 106773.
[20] A. Chaudhary, R. Agarwal, Machine learning techniques for anomaly detection application domains, in: Paradigms of Smart and Intelligent Communication, 5G and beyond, Springer Nature, Singapore, 2023, pp. 129–147.
[21] T. Chen, et al., Machine learning in manufacturing towards industry 4.0: from 'for now' to 'four-know', Appl. Sci. 13 (3) (2023) 3.
[22] D. Dondurur, Chapter 4 - Fundamentals of data processing, in: Acquisition and Processing of Marine Seismic Data, Elsevier, 2018, pp. 211–239.
[23] J. Fioresi, et al., Automated defect detection and localization in photovoltaic cells using semantic segmentation of electroluminescence images, IEEE J. Photovoltaics 12 (1) (2022) 53–61.
[24] H. Han, C. Gao, Y. Zhao, S. Liao, L. Tang, X. Li, Polycrystalline silicon wafer defect segmentation based on deep convolutional neural networks, Pattern Recogn. Lett. 130 (2020) 234–241.
[25] B. Su, Z. Zhou, H. Chen, X. Cao, SIGAN: A novel image generation method for solar cell defect segmentation and augmentation, ArXiv 14 (8) (2021). https://arxiv.org/abs/2104.04953.
[26] M. Allen, The SAGE Encyclopedia of Communication Research Methods, SAGE Publications, 2017.
[27] M.H. DeGroot, M.J. Schervish, Probability and Statistics, fourth ed., Addison-Wesley, Boston, 2012.
[28] G.M. Javier, R. Evans, P. Dwivedi, T. Trupke, Z. Hameiri, Advanced production line monitoring with time-lag sequential analysis, in: 50th IEEE Photovoltaic Specialists Conference, 2023.
[29] C.V. Deutsch, Geostatistics, in: Encyclopedia of Physical Science and Technology, third ed., Academic Press, New York, 2003, pp. 697–707.
[30] G. Box, T. Kramer, Statistical process monitoring and feedback adjustment: A discussion, Technometrics 34 (3) (1992) 251.
[31] sklearn.ensemble.IsolationForest, scikit-learn. https://scikit-learn/stable/modules/generated/sklearn.ensemble.IsolationForest.html.
[32] W.S. Al Farizi, I. Hidayah, M.N. Rizal, Isolation forest based anomaly detection: A systematic literature review, in: 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE), 2021, pp. 118–122.
[33] A.G. Prasad, S. Saravanan, E.V. Gijo, S.M. Dasari, R. Tatachar, P. Suratkar, Six Sigma-based approach to optimise the diffusion process of crystalline silicon solar cell manufacturing, Int. J. Sustain. Energy 35 (2) (2016) 190–204.
[34] J. Ahmad, A. Ciocia, S. Fichera, A.F. Murtaza, F. Spertino, Detection of typical defects in silicon photovoltaic modules and application for plants with distributed MPPT configuration, Energies 12 (23) (2019) 4547.
[35] R. Al-Mashhadani, G. Alkawsi, Y. Baashar, A.A. Alkahtani, F.H. Nordin, W. Hashim, Deep learning methods for solar fault detection and classification: A review, Information Sciences Letters 10 (2) (2021).