Identifying Customer Returns in a Printed Circuit Board production line using the Mahalanobis Distance

Endre Sølvsberg[1✉ [0000-0001-6916-9793]], Simone Arena[2[0000-0002-9932-6080]], Fabio Sgarbossa[1[0000-0002-9541-3515]], Per Schjølberg[1]

[1] Norwegian University of Science and Technology, Trondheim, Norway
[2] University of Cagliari, Cagliari, Italy
✉endre.solvsberg@ntnu.no

Abstract. This paper discusses as its primary research question the viability of using the Mahalanobis Distance as a multivariate method for detecting outliers in an industrial setting. An algorithm is used to detect future customer returns in a printed circuit board production line situated in Sibiu, Romania. From the literature, there is a lack of methods, tools and guidelines concerning the paradigm of Zero-Defect Manufacturing. The novelty of the method presented includes separation of highly specialized, future outliers from other outliers, and further automation using Python, a Docker container, a graphical user interface, a search-engine and a reporting tool. This allows the method to be used without external assistance. The data used is extracted industrial datasets from Continentals datalake. The algorithm detects 20% of future outliers and has been implemented by Continental. This can possibly be improved by increasing domain knowledge. The generality of the algorithm in principle allows for use at any of Continental's production lines. There are strong assumptions regarding the requirements for the method, including benefits of employing domain knowledge critical variable identification and detection rate improvements. Further improvements of detection rate are also discussed. The paper concludes that the algorithm can detect a percentage of highly specialized outliers with simple automation in Python, but also acknowledges limitations in terms of increased demands from data quality and domain knowledge.

Keywords: Industry 4.0, Multivariate Analysis, Automation.

1 Introduction

Nowadays, the notable emphasis on sustainability of manufacturing companies has determined a specific focus on production aspects including flexibility, quality, reliability, productivity, operational efficiency, and cost performance [1]. To meet the growing challenges of today's dynamic environment, these companies must prioritize the implementation of innovative and competitive strategies to achieve the goal of sustainability improvement of their processes and systems. In this context, one of the main challenges is related to product quality since it could lead to severe consequences on customer satisfaction, companies' reputation, financial performance, environmental impact, and resource usage. Therefore, production processes should be continuously improved through the implementation of the so-called zero-waste and zero-defect strategies aiming at preventing waste, reducing manufacturing lead times, limiting resources, and producing high-quality goods [2].

The advent of Industry 4.0 and the maturing of its enabling technologies have substantially boosted the rapid digital transformation to manage production complexity and enhance data management for processes understanding and advanced problem-solving. This has contributed to the evolution of the traditional manufacturing industries to a new class of smart factories characterized by fully interoperated, automated, and optimized production flow through the adoption of emerging technologies such as Cyber-Physical Systems (CPS), Artificial Intelligence (AI), Internet of Things (IoT) and Big Data [3]. The integration between physical and digital processes, where sensors, connected devices, equipment, and production systems continuously collect and share data, has enabled a consistent approach based on data-driven decision-making strategies. The valuable information and knowledge provided by this vast amount of available data can be leveraged into a plethora of potential applications.

One promising application concerns production quality improvement and waste reduction. In this scenario, Zero Defect Manufacturing (ZDM) is a disruptive paradigm focused on data-driven approaches aiming at realizing the vision of zero defects following the concept of "to do the right thing the first time" [4]. Thus, ZDM exploits Industry 4.0 tools to prevent defects and errors in a production process through detection, prediction, prevention, and repair strategies [5]. As precisely reported by Wang [4] and Psarommatis et al. [6], these strategies include the following steps: (i) data acquisition from sensor-equipped machinery, collection, storage, and cleaning; (ii) automatic signal processing, filtering, and feature extraction; (iii) data mining and knowledge discovering for diagnosis and prognosis; (iv) gathering information about monitored defects; (v) online predictive maintenance and (vi) re-configuration and re-organisation of the production process.

In the last years, by exploiting the potential of the enabling technologies of Industry 4.0, two specific strategies, i.e. detection and prediction, have significatively attracted the researchers' interest [7]. Especially, the great potential of AI and Machine Learning (ML) algorithms have created new opportunities for more effective quality management and advanced problem-solving since they are capable to process complex datasets analysing different factors and scenarios, identifying structures and patterns, predicting future behaviours, and making optimal decisions [2].

Concerning ZDM detection strategy based on AI approaches, Tabernik et al. [8] proposed a deep learning technique for surface-anomaly detection within electric commutator production, Okaro et al. [9] presented a ML algorithm for the automatic detection of faults for addictive manufacturing applications, Soualhi et al. [10] used an unsupervised classification technique for outliers detection and diagnosis for quality assessment purpose.

Concerning ZDM prediction strategy, Peres et al. [11] adopted different ML algorithms aiming at predicting dimensional defects in a real automotive multistage assembly line, Wang K. et al. [12] proposed a deep learning approach for batch process quality prediction, Ranjan et al. [13] presented an AI-based technique for quality prediction in micro-drilling.

Finally, thanks to the increasing interest in these topics, several comprehensive reviews are recently proposed aiming at providing the state-of-the-art perspective, the emerging open challenges, and the future directions [1, 2, 6, 7, 14-16]. As emerged from [16], the current literature on ZDM shows a lack of methods, tools, and guidelines for its proper implementation in manufacturing facilities, especially the adoption of data-driven approaches and technique is still a challenging aspect. Thus, this paper presents the preliminary development of a Machine Learning algorithm for the detection of customer returns from a Printed Circuit Board (PCB) production line. The main goal is to provide an easy-to-implement tool to perform effective diagnostic tasks in supporting an agile and informed decision-making process. The anomaly detection process is carried out by adopting the Mahalanobis Distance (MD) to monitor and recognize the anomalous observations on the sensors' data.

The remainder of this paper is organised as follows: Section 2 describes the production line used as a pilot case. Section 3 illustrates the methodology used to detect customer returns and the ML algorithm implemented. In Section 4, the achieved results are presented and discussed. Finally, conclusions and future research are reported in Section 5.

2 Continental Pilot Case

This chapter will first outline the Continental pilot, with challenges and goals. The chapter also presents the resulting research problems from Continental's side. Subchapters include a description of the EU project QU4LITY and a detailed description of the pilot.

The main challenge concerning the Continental pilot case deals with faulty PCBs undetected by the in-line testing suite and manual testing, resulting in customer returns. The main goal was to use available test data from Continental's datalake to identify specific PCBs that are candidates for future customer returns through extraction of critical-to-quality (CTQ) variables, multivariate statistical analysis and detection of specific outliers related to future customer returns. The resulting research problems were then to: (i) find a suitable statistical method for identifying outliers specifically related to customer returns and separate these from other statistical outliers and normal units, and; (ii) examine the viability of automating the method to allow for rapid in-line sampling and analysis. This would ideally also allow for Continental personnel to utilize the method on site with little or no external expert assistance.

2.1 QU4LITY

QU4LITY, or "Autonomous Quality Platform for Cognitive Zero-defect Manufacturing Processes through Digital Continuity in the Connected Factory of the Future" was an EU project, part of the Horizon 2020 program and was active between 2019 and 2022. The project lead was ATOS Spain, with 45 partners including large industrial corporations, SMEs, research institutes, universities, digital innovations hubs and industrial associations. In QU4LITY, there were 14 lighthouse projects, of which five were manufacturing equipment pilots and the remaining nine process pilots.

2.2 Pilot description

The Continental pilot, one of the nine process pilots, concerns a combined Surface-Mount Device (SMD) and final assembly line situated in Sibiu, Romania. The manufacturing capacity is on average 30-40 million units per year. The four expected outcomes from the pilot were data mining in production systems to provide early indicators and trends from process signals, facilitate the creation of new applications that include the entire value chain, digital modelling and zero-defect strategies and physical interpretation and initiation of real-time reaction plans for shop floor visualization management. This paper focuses on the first expected outcome. The pilot group consisted of members from Continental, ATB Bremen, SINTEF and the Norwegian University of Science and Technology, and included multidisciplinary expertise within the fields of IT, quality engineering, domain expertise and statistical analysis.

The production line guides the PCBs through laser marking, paste printing, automatic placement machines and reflow ovens. In addition, there are three in-line testing stations; solder paste inspection, automatic optical inspection and a final in-circuit test. Finally, there are sensors placed along the production line. These are product specific, environmental and line equipment sensors.

3 Methodology

This chapter describes the methodology employed in the paper. It also describes some challenges and requirements and an outline of the data extraction limitations. The chapter discusses the exploratory analysis through descriptive statistics and both single- and multivariate methods in detail. It also explains how the Mahalanobis Distance (MD) was selected, with some justifications. A subchapter describes the MD with a bit more detail.

Initial pilot group discussions included identifying a suitable unit for analysis. Data for the unit had to be available from the datalake, and the data should also be as complete as possible, with little to no missing variable data. PCBs for automotive cameras were selected as the best candidate. External pilot members were then given access to the part of the datalake where this unit was situated. This was again restricted to test data, and no production data was made available.

A challenge when extracting data from the datalake at Continental was structure. Data is structured so that 20-30 columns and around 1700 lines describe one specific unit. With no initial domain expertise, a decision was made to focus on data from one of the in-line testing stations. The data then had to be formatted to allow for the use of statistical software. To avoid including non-pertinent variables and to maximize probable CTQ variables, domain expertise from on-site in Sibiu was subsequently included in the pilot group, resulting in a dataset of confirmed customer returns, increased knowledge on how data was organized for the different production and test steps throughout the production process, in addition to the identification of probable CTQ variables from all steps of the process.

These variables were extracted from the datalake and formatted to allow for analysis. Historical datasets were used based on test data from the production line, datasets with confirmed customer returns and constructed datasets with various numbers of confirmed customer returns and sample sizes from the production line.

An initial descriptive analysis was done to get an overview of the datasets. This included means, standard deviations, min, max and quartile values, skewness and kurtosis. A Kolmogorov-Smirnov test was conducted to test for normal distribution [17]. A correlation matrix was constructed to detect linear relationships between variables using Pearson correlation.

As part of the exploratory analysis an Individual Value and Moving Range (I-MR) analysis was performed for all variables to look at performance over time and to identify anomalies like outliers, trends, shifts, oscillations and so on, based on the tests utilized. I-MR is an individual counterpart to the traditional X-Bar R chart used in Statistical Process Control (SPC) [18]. The analysis detected anomalous data, most importantly samples exceeding the 3-sigma upper control limit. The testing was done using 100 sample time series from start to end of the datasets. An example chart from testing is shown in Fig. 1. The software used is Minitab.

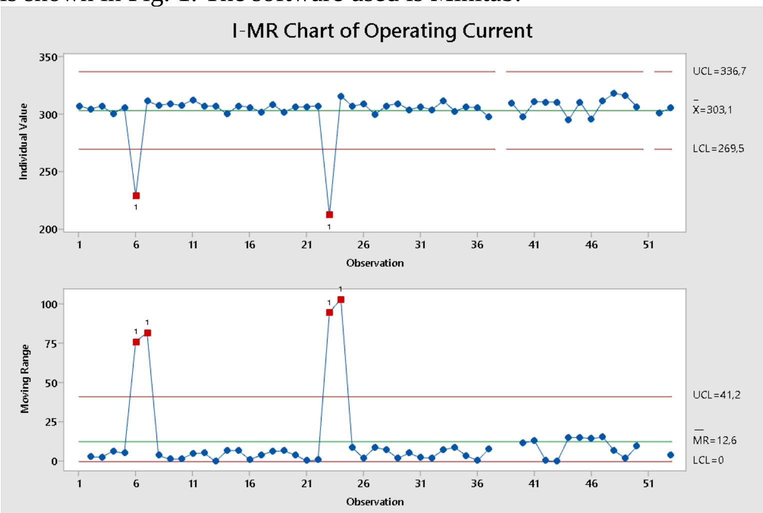

Fig. 1. Example I-MR chart.

Other single variate exploratory tests include Run Chart tests, an F-test comparing a random dataset with a confirmed customer returns dataset and Grubbs' tests on all variables. This initial exploratory testing and analysis using univariate methods was able to identify two of the units from the confirmed customer returns dataset in the F-test and Grubbs' tests. No other single variate testing found any significant difference between normal and customer returns datasets.

The first multivariate test conducted used a Z-transformation on all variables in a random and a customer returns dataset, using absolute values to determine mean and cumulative sigma. An F-test comparing the two datasets found them to be significantly different, and a boxplot found two extreme outliers in the customer returns dataset.

The second test conducted was the Hotelling's T^2 test for Generalized Variance, which is a multivariate counterpart to I-MR and X-Bar R charts [19]. The chart is shown by:

$$T^2 = n(\bar{X} - \bar{\bar{X}})S^{-1}(\bar{X} - \bar{\bar{X}}), \quad (1)$$

where

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_{ijk}, \quad (2)$$

$$\bar{\bar{X}}_j = \frac{1}{n}\sum_{k=1}^{m} \bar{X}_{jk} \quad (3)$$

and

$$S = \begin{bmatrix} \bar{S}_1^2 & \bar{S}_{12} & \dots & \bar{S}_{1p} \\ \dots & \bar{S}_2^2 & \dots & \bar{S}_{2p} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \bar{S}_p^2 \end{bmatrix}, \quad (4)$$

where

$$\bar{S}_j^2 = \frac{1}{m}\sum_{k=1}^{m} S_{jk}^2, \quad (5)$$

$$S_{jk}^2 = \frac{1}{n-1}\sum_{I=1}^{n} (X_{ijk} - \bar{X}_{jk})^2, \quad (6)$$

$$S_{jh}^2 = \frac{1}{m}\sum_{k=1}^{m} S_{jhk} \quad (7)$$

and

$$S_{jhk} = \frac{1}{n-1}\sum_{i=1}^{n} (X_{ijk} - \bar{X}_{jk})(X_{ihk} - \bar{X}_{hk}) \quad (8)$$

$|S|$ is the determinant of the sample covariance matrix and $|S_i|$ is the determinant of the sample covariance matrix for sample i. $|S|$ is the center line of the chart. Upper and lower control limits are calculated by

$$LCL = \frac{|S|}{b_1}(b_1 - 3b_2^{1/2}), \quad (9)$$

where

$$b_1 = \frac{1}{(n-1)^{\mu}} \prod_{i=1}^{\mu} (n-i), \tag{10}$$

$$b_2 = \frac{1}{(n-1)^{2\mu}} \prod_{i=1}^{\mu} (n-i) \left[\prod_{j=1}^{\mu} (n-j+2) - \prod_{j=1}^{\mu} n-j \right], \tag{11}$$

and

$$UCL = \frac{|S|}{b_1} (b_1 + 3b_2^{1/2}) \tag{12}$$

The T^2 chart shows a clear outlier, and the chart for Generalized Variance shows two clear outliers. The chart is shown in Fig. 2.

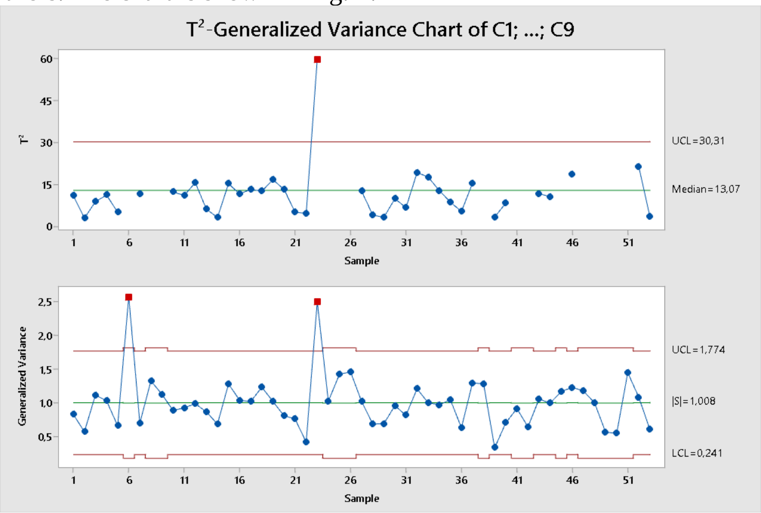

Fig. 2. T2-Generalized Variance chart.

The last multivariate method employed before settling on a preferred method was the Principal Component Analysis (PCA) [20]. The method is most often used to reduce the dimensionality of datasets by identifying a smaller number of uncorrelated variables from a large dataset but can also be used to identify outliers in non-Euclidian space. The test calculates a set of orthogonal eigenvectors of the covariance matrix of the covariance matrix of the variables. The result is a principal component matrix, where the first component accounts for the largest data variation and so on. The main goal is to explain the maximum amount of variance using the minimum number of components.

The eigenvectors contain coefficients corresponding to each variable and are the weights for each variable to calculate principal component score. The scores are calculated by Z = YV, where Z is the principal component matrix, Y a raw data matrix (n * p) and V a matrix of eigenvectors. The proportion of sample variance explained by the kth principal component is calculated by

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_p} \tag{13}$$

The analysis showed that 96,3% of the variance of the original dataset could be represented by using two principal components.

As part of the PCA analysis, the Mahalanobis Distance was employed to find unit distances to a center point P. This testing provided a good way to identify more of the confirmed customer returns. This includes easy checks of datasets, change of limits, change of variables and so on. Easy of automation is also a consideration that will be discussed in a later part. A final decision was made to focus solely on this method. The Mahalanobis Distance is a useful multivariate outlier detection method, in that it removes correlation in the dataset, which can be a challenge in high-dimensional datasets [21].

3.1 Mahalanobis Distance

The Mahalanobis Distance (MD) outlier detection method is unitless and scale invariant. It measures the distance of a point P from a distribution D in an n-dimensional non-Euclidean space [21]. The distance from a point to the distribution is calculated by

$$MD = \sqrt{\left(\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^T V_n^{-1}\left(X_i - \overline{X}_n\right)\right)} \tag{14}$$

where Xi is the data value vector at row I, $\overline{X}$ the mean vector and Vn-1 the inverse of the covariance matrix [22]. The square of the MD is approximately chi-squared distributed with n degrees of freedom, where the degrees of freedom are equal to the number of variables [23], and therefore it is possible to find a suitable critical distance based on the confidence level and degrees of freedom by using a chi-squared table.

4 Results and discussion

This chapter presents the results from the MD analysis, and a subchapter describes how the Python automation was done. The chapter describes how the algorithm was placed in a virtual container with other tools inside the Continental network. The discussion includes perspectives on detection rate, requirements and limitations.

Initial analysis using the MD was done by inserting 50 confirmed customer returned PCBs into a random dataset, using confidence levels of 95, 99 and 99,9. Setting the confidence level too low would result in a critical limit inside normal PCBs in addition to the outlying units. A sufficiently high confidence level would ensure that the critical limit avoids normal units, and still identify the outliers, as shown in Fig. 3. Experimenting with different sample sizes provided insight into how to better detect customer return units with a minimized amount of false alarms.

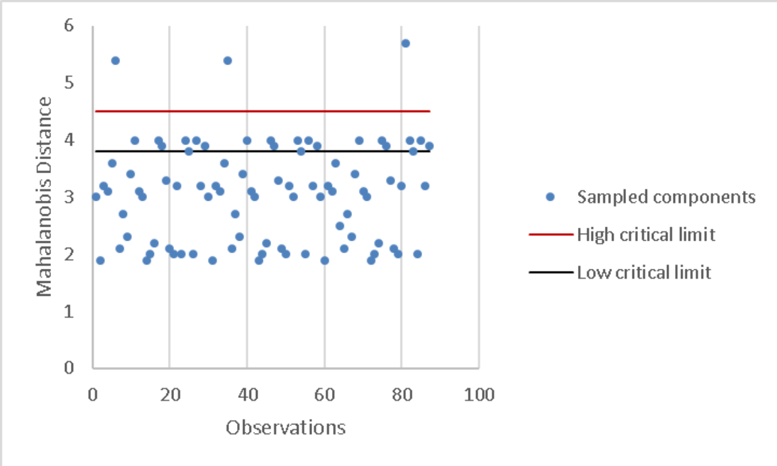

Fig. 3. Mahalanobis Distance with limit examples.

The initial detected about 4% of the customer returned units. Experimenting with different critical limits, sample sizes and CTQ variables improved the detection rate to 10 and finally 20%. There is also an economic constraint in terms of selecting the critical limit. There is a cost to removing units from production and a benefit to removing the customer returns from production. The goal was to find a point that maximizes the units detected and minimizes the units removed from production.

4.1 Automation using Python

A decision was made to automate the MD algorithm using Python. The reasoning included Python being open source, the data extraction and formatting team mainly used Python, and it was the preferred tool for the Continental IT engineers working on the datalake. The data was extracted from the Continental datalake and formatted to fit the algorithm. The formatted data were provided in .csv format, and then transformed into a matrix object (df1) with N * M dimensions. The rows with any missing data were excluded. The degrees of freedom (DoF) are equal to the number of variables or the number of columns. The mean vector of is then calculated for every column. The algorithm calculates the covariance matrix and its inverse, and performs an identity test. Maintaining the matrix notation, the means vector is subtracted from the initial data matrix df1, and the result (df1*) is multiplied with the inverse covariance matrix. The intermediate result is stored in a temporary object (temp).

Further, the square of the MD is calculated as the product of temp and the transposed matrix of df1*. The result is a diagonal matrix (MD_squared), where the elements on the diagonal are the squares of the MD for each row of the initial data matrix df1. These elements are extracted from MD_squared and appended to df1, which then becomes a matrix of N * (M+1) dimensions. The Chi square critical value is calculated (critical_value), with a 0.9xx level of confidence and degrees of freedom

DF. The outliers are identified as any data points above the set critical value. The Python code for the MD calculation is shown in Fig. 4.

```
df1 = df1.dropna(how='any',axis=0)
DF = len(df1.columns)
mean = df1.mean(axis=0)
covar_mat = df1.cov()
covar_mat = nm.around(covar_mat,6)
inv_covar = nm.linalg.inv(covar_mat)
inv_covar = nm.around(inv_covar,6)
test_identity = nm.matmul(inv.covar,covar_mat)
test_identity = nm.around(test_identity)
sample_minus_mean = df1-mean
temp = nm.matmul(sample_minus_mean,inv_covar)
temp = nm.around(temp,6)
transp = nm.transpose(sample_minus_mean)
MD_squared = nm.dot(temp,transp)
df1['MD_squared'] = nm.around((nm.diagonal(MD_squared)),3)
critical_value = nm.around((scipy.stats.chi2.ppf(0.9xx, DF)),3)
```

Fig. 4. Python code calculating Mahalanobis Distance, MD2 and critical limit.

The Python code was tested both manually and using software. Another test included conducting a data pipeline test, including data extraction, formatting, MD algorithm activation and output of MD for all included units and a table of all units above the critical limit. After a successful test, a Docker container [24], which is a virtual environment that allows for applications, dependencies and related libraries to function in any environment, was placed inside the Continental network. The container included the MD algorithm, a Graphical User Interface (GUI) including a search engine and a reporting tool. The goal was that the quality team on-site and any other relevant personnel at Continental could use the algorithm autonomously. An overview of the method is shown in Fig. 5.

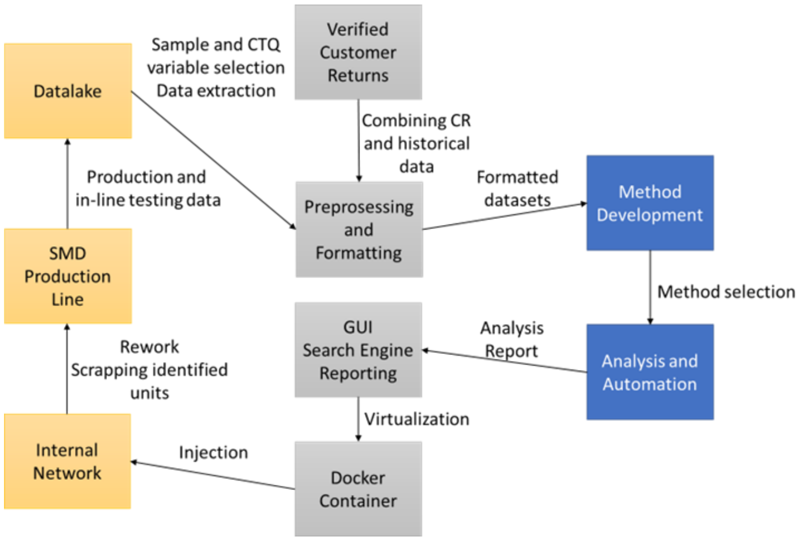

Fig. 5. Method architecture.

In the final review meeting of the QU4LITY project, pilot management reported successful algorithm implementation in Sibiu, and a continuing detection rate of 20%. This implementation seems to support the viability of the method in an industrial Big Data environment. Using the MD also solves Big Data challenges like heterogeneity, statistical accuracy and efficiency. The confidence level will remain unchanged at the chosen level with dimensionality upscaling, and the method is able to analyze datasets up to a day's production of units on-site in less than a second. The generality of the method allows for usage in other production lines at Continental and may also be interesting for other manufacturers and in other sectors for detecting customer returns or other specialized outliers in big datasets.

Although these results provided a promising starting point that ultimately led to the implementation of the solution, there were several requirements observed to make that possible. Domain knowledge through process and process line knowledge is assumed to be highly important to be able to filter out non-pertinent variables. Domain knowledge also contributes towards identifying the CTQ variables for the use-case in question and discovering if there are any CTQ variables missing or not measured. Including non-pertinent variables or missing CTQ variables most likely will skew the analysis output.

Cloud storage solutions, including the Continental datalake, contain large amounts of data that can be hard to navigate without any form of domain knowledge. Including domain knowledge through on-site expertise seems to be essential in providing insight into the production process from start to finish, variables importance and linking data structure with the physical production line and where in the process they stem from. Using available domain expertise can help with data filtering and using expertise to identify CTQ variables and suggest experiments to better detect outliers. Still, the domain expertise is probably not complete, and, there is still likely incomplete knowledge and insight related to the customer return outliers, leading to some variables being left out of the analysis or some that are perhaps not even measured. Increasing domain knowledge and a deeper knowledge on the CTQ variables in the analysis would likely improve the 20% detection rate significantly, and potentially lead to increased success in research related to condition monitoring or even prediction capability.

The Mahalanobis Distance and other multivariate analysis methods are important and often more powerful tools when the univariate methods can't produce significant results and the outliers result from combinations of variables, not unique variables. This importance can further be improved when basic automation is easy to implement. Some apparent negatives are increased demands on use-case domain knowledge, structure, data completeness, synchronicity, capture rate, sample sizes and CTQ variable knowledge to mention some.

5 Conclusion

The work in the Continental pilot, where they presented a challenge with undetected customer returns from a PCB production line, has resulted in an outlier detection method using the Mahalanobis Distance. The method can detect 20% of the customer returns as specific multivariate outliers. The method, combined with Python automation, a GUI, a search-engine and a reporting tool has to the implementation of the method at Continental factory in Sibiu, Romania. The general nature of the method also provides interesting possibilities in utilization on other process lines at Continental, in other companies and other industrial sectors. Although the assumption is that increased domain knowledge can help improve identification of CTQ variables, and increase the detection rate, the method has successfully been able to identify a significant percentage of the before undetected outliers and is therefore considered to be viable as a method. The automation using Python allows for big data analysis, with MD allowing for effective, simultaneous analysis of multiple variables and units. The inclusion of a domain expertise seems to have been of great importance in finding part of the CTQ variables and in aiding with continuous improvement of the method through rapid prototyping.

5.1 Future work

Continental has shown interest in continuing work with the method. This can open interesting possibilities in examining the importance of increased process knowledge and a higher percentage of CTQ variables, and how this can affect the detection rate and the false positives rate. The MD method described in this paper may not be feasible as a solution in many cases. The quality of the output is limited by data quality, especially concerning CTQ variables and the exclusion of non-pertinent variables. A high level of domain and process knowledge is also helpful to ensure the proper inputs. Application of requirements and the inclusion of domain expertise may also be interesting for future approaches involving AI and ML methods.

Acknowledgements

This work has been done as part of the BRU21 program at the Norwegian University of Science and Technology, with the independent oil company OKEA as a sponsor. The contributions of Chiara Caccamo and Ragnhild Eleftheriadis, who were both co-authors in a conference paper that served as a foundation for this paper, are greatly appreciated and acknowledged.

References

1. G. Fragapane, R. Eleftheriadis, D. Powell, and J. Antony, "A global survey on the current state of practice in Zero Defect Manufacturing and its impact on production performance," Computers in Industry, vol. 148, p. 103879, 2023.
2. N. Leberruyer, J. Bruch, M. Ahlskog, and S. Afshar, "Toward Zero Defect Manufacturing with the support of Artificial Intelligence—Insights from an industrial application," Computers in Industry, vol. 147, p. 103877, 2023.
3. R. Kunst, L. Avila, A. Binotto, E. Pignaton, S. Bampi, and J. Rochol, "Improving devices communication in Industry 4.0 wireless networks," Engineering Applications of Artificial Intelligence, vol. 83, pp. 1-12, 2019.
4. K.-S. Wang, "Towards zero-defect manufacturing (ZDM)—a data mining approach," Advances in Manufacturing, vol. 1, pp. 62-74, 2013.
5. H.-T. Chen, "Quality function deployment in failure recovery and prevention," The service industries Journal, vol. 36, no. 13-14, pp. 615-637, 2016.
6. F. Psarommatis, G. May, P.-A. Dreyfus, and D. Kiritsis, "Zero defect manufacturing: state-of-the-art review, shortcomings and future directions in research," International journal of production research, vol. 58, no. 1, pp. 1-17, 2020.
7. B. Caiazzo, M. Di Nardo, T. Murino, A. Petrillo, G. Piccirillo, and S. Santini, "Towards Zero Defect Manufacturing paradigm: A review of the state-of-the-art methods and open challenges," Computers in Industry, vol. 134, p. 103548, 2022.
8. D. Tabernik, S. Šela, J. Skvarč, and D. Skočaj, "Segmentation-based deep-learning approach for surface-defect detection," Journal of Intelligent Manufacturing, vol. 31, no. 3, pp. 759-776, 2020.
9. I. A. Okaro, S. Jayasinghe, C. Sutcliffe, K. Black, P. Paoletti, and P. L. Green, "Automatic fault detection for laser powder-bed fusion using semi-supervised machine learning," Additive Manufacturing, vol. 27, pp. 42-53, 2019.
10. A. Soualhi, G. Clerc, and H. Razik, "Detection and diagnosis of faults in induction motor using an improved artificial ant clustering technique," IEEE Transactions on Industrial Electronics, vol. 60, no. 9, pp. 4053-4062, 2012.
11. R. S. Peres, J. Barata, P. Leitao, and G. Garcia, "Multistage quality control using machine learning in the automotive industry," IEEE Access, vol. 7, pp. 79908-79916, 2019.
12. K. Wang, R. B. Gopaluni, J. Chen, and Z. Song, "Deep learning of complex batch process data and its application on quality prediction," IEEE Transactions on Industrial Informatics, vol. 16, no. 12, pp. 7233-7242, 2018.
13. J. Ranjan et al., "Artificial intelligence-based hole quality prediction in micro-drilling using multiple sensors," Sensors, vol. 20, no. 3, p. 885, 2020.

14. D. Powell, M. C. Magnanini, M. Colledani, and O. Myklebust, "Advancing zero defect manufacturing: A state-of-the-art perspective and future research directions," Computers in Industry, vol. 136, p. 103596, 2022.
15. V. Azamfirei, F. Psarommatis, and Y. Lagrosen, "Application of automation for in-line quality inspection, a zero-defect manufacturing approach," Journal of Manufacturing Systems, vol. 67, pp. 1-22, 2023.
16. F. Psarommatis and G. May, "A practical guide for implementing Zero Defect Manufacturing in new or existing manufacturing systems," Procedia Computer Science, vol. 217, pp. 82-90, 2023.
17. H. W. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," Journal of the American statistical Association, vol. 62, no. 318, pp. 399-402, 1967.
18. K. Mukundam, D. R. Varma, G. R. Deshpande, V. Dahanukar, and A. K. Roy, "I-MR control chart: A tool for judging the health of the current manufacturing process of an API and for setting the trial control limits in phase I of the process improvement," Organic Process Research & Development, vol. 17, no. 8, pp. 1002-1009, 2013.
19. M. A. Djauhari, "Improved monitoring of multivariate process variability," Journal of Quality Technology, vol. 37, no. 1, pp. 32-39, 2005.
20. H. Abdi and L. J. Williams, "Principal component analysis," Wiley interdisciplinary reviews: computational statistics, vol. 2, no. 4, pp. 433-459, 2010.
21. R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, "The mahalanobis distance," Chemometrics and intelligent laboratory systems, vol. 50, no. 1, pp. 1-18, 2000.
22. I. Ben-Gal, "Outlier detection," in Data mining and knowledge discovery handbook: Springer, 2005, pp. 131-146.
23. G. Gallego, C. Cuevas, R. Mohedano, and N. Garcia, "On the Mahalanobis distance classification criterion for multidimensional normal distributions," IEEE Transactions on Signal Processing, vol. 61, no. 17, pp. 4387-4396, 2013.
24. A. M. Potdar, D. Narayan, S. Kengond, and M. M. Mulla, "Performance evaluation of docker container and virtual machine," Procedia Computer Science, vol. 171, pp. 1419-1428, 2020.