Diagnosis of quality management systems using data analytics – a case study in the manufacturing sector

Sanchez-Marquez R[1], Albarracin Guillem JM[2], Vicens-Salort E[3], Jabaloyes Vivas J[4]

Abstract: The main objective is to improve customer satisfaction by developing and testing a method to study quality management systems by analysing the key performance indicators of balanced scorecards in manufacturing environments. The methodology focuses on the identification and quantification of relationships between internal and external metrics that allow moving from performance measurement to effective performance management. It has been tested as a case study approach using real data from two complete years of the balanced scorecard of a leading manufacturing company. The results provided a new understanding of how the quality management system works that was used to make systemic and strategic decisions to improve the long-term performance of the company. Industry practitioners with a moderate level of data analytical skill can use it to help managers and executives improve management systems.

Keywords: Manufacturing, quality management system, data analytics, balanced scorecard, key performance indicators

The final version of this article with full bibliographic details is available online at: https://doi.org/10.1016/j.compind.2019.103183

1. Introduction

Although performance measurement is not an end in itself, the literature identifies it as an essential part of performance management, since a lack of appropriate performance measurement can be a barrier to change and improvement [1]. Bititci et al. [2] claim that reviewing and prioritising internal goals when changes in the internal and

---

[1] Rafael Sanchez-Marquez (email: rsanch18@ford.com)
Doctorate Student (corresponding author)
Production Management and Engineering Research Centre
Universitat Politecnica de Valencia
Camino de Vera, s/n. 46021 Valencia. Spain
ORCID: 0000-0001-9071-9550

[2] José Miguel Albarracín Guillem (email: jmalbarr@omp.upv.es)
Departamento de Organización de Empresas
Universitat Politècnica de València

[3] Eduardo Vicens-Salort (email: evicens@cigip.upv.es)
Production Management and Engineering Research Centre
Universitat Politècnica de València
Camino de Vera, s/n. 46021 Valencia. Spain
ORCID: 0000-0002-8111-5269

[4] José Jabaloyes Vivas (email: jabaloye@eio.upv.es)
Department of Statistics, Operational Investigation and Quality
Universitat Politecnica de Valencia
ORCID: 0000-0003-3411-2062

Declarations of interest: none

external environment are significant is an important feature of effective performance measurement systems. As a performance management system, the balanced scorecard (BSC) establishes the importance of knowing and using the cause-and-effect relationships between internal and external metrics to move from performance measurement to effective performance management [3].

This paper develops a methodology that allows practitioners to identify and quantify those relationships by using key performance indicators (KPIs) of the BSC. Since the objective is to develop a practical methodology that uses real BSC data, its development focuses on the integration of appropriate methods and tools for data analytics.

The present method was developed and tested in a leading multinational manufacturing company, which had implemented a balanced scorecard for the production facilities composed of seven management/operating systems [4]: safety; quality; delivery; cost; people; maintenance; and environment. The quality management system (QMS) was selected by the directors of the company to develop and test the validity of the method, since it was the system with the highest level of complexity. Nevertheless, with small adjustments the method can be applied in the other six management systems in the same way as in quality.

According to the international standard ISO 9001:2015 that specifies the requirements for a quality management system, industrial products and their manufacturing processes must be designed to meet customer expectations through the specific engineering specifications of critical product characteristics. These are typically specified in terms of a nominal (ideal) value and a tolerance interval (upper specification limit - lower specification limit). Controlling and managing these critical characteristics is a fundamental task of the quality measurement system and, therefore, of the quality management system. When critical characteristic measurements meet engineering specifications, they also meet customer expectations, leading to customer satisfaction. These measurements are summarised in the internal KPIs of the QMS. Therefore, the QMS includes internal KPIs, which summarise compliance with engineering specifications, and external KPIs, which include customer complaint indicators. Consequently, if the quality management system works well, internal and external KPIs must reflect customer satisfaction and, therefore, both sets of indicators must be highly correlated.

Identifying which internal KPIs drive customer satisfaction (external KPIs) and quantifying such relationships allows executives and managers to design strategies to improve customer satisfaction, which is the main objective of this research work. In addition, the results serve as a start point to reduce the complexity of the quality management system (QMS). Simplification of performance management systems (PMSs) is a recurring topic in the literature [5, 6, 7]. Therefore, the following two main research questions were established:

- How do the KPIs of the QMS relate to each other?
- How can these KPIs help improve customer satisfaction?

There is some research on the development of analytical methods based on the key performance indicators of balanced scorecards in the manufacturing environment. However, the results of these works [8, 9, 10, 11] are qualitative rather than quantitative (which should be the nature of any analytical method). Therefore, the development of robust analytical methods for manufacturing systems based on proven scientific tools is an issue that has not been covered in the literature. This paper focuses on the diagnosis of a management system to improve its capabilities and this implies a novel approach.

This work was carried out as part of a collaborative research project between the company (which requested to keep its identity and data confidential) and the Centre for Research and Production Management of the Polytechnic University of Valencia (Spain) to improve management methods in manufacturing environments.

The company decided to use the findings of the present study to make changes in the balanced scorecards of all production facilities worldwide. Although these changes are detailed in the results section of this paper, they can be summarised as a reduction in the complexity of the operating system and the inclusion of new KPIs, as well as the elimination of some existing indicators that have shown less strategic weight. The new insight provided by this study was used to prioritise some strategies over others and start new strategies to improve customer perception about the quality of company products.

The method was validated using real data from two complete years of key quality performance indicators as a case study approach.

2. Literature review

The literature review was structured to cover the relevant topics:

- Analytical methods applied to key performance indicators using actual data

    - Regression, multiple linear regression (MLR), partial least squares (PLS), principal component analysis (PCA), time series, artificial neural networks (ANN), data mining
- Analytical methods applied to building balanced scorecards as a proactive tool
    - Fuzzy logic, analytic network process (ANP)
- The balanced scorecard in the manufacturing environment
- Limitations of the analytical tools mentioned above
- Limitations of the balanced scorecard model
- Quality management systems in the manufacturing environment

The main objective of the literature review was to identify the best possible approach and the strengths and limitations of each method available in the literature. As discussed in the introduction section, the present method covers a new objective, although to some extent it is based on improvements in existing methods developed by other authors and applied for other purposes. In addition, it addresses the limitations already commented by the authors themselves.

2.1. Analytical methods applied to KPIs using actual data

The available works use analytical tools such as MLR [12], PCA and PLS [13, 14, 15], and graphic methods [16], to assess the effectiveness of the strategies in place and quantify their impact on the output metrics. Sanchez-Marquez et al. [16] suggest previously selecting the output metrics among all the key performance indicators (KPIs) included in the scorecard to streamline the method as a key step in any method that addresses the KPIs. While some comments are made about the need for more perspective to understand how the system works, this goal is beyond the scope of those works.

2.2. Analytical methods applied to the BSC as a proactive tool

Other works focus on proactive methods to build a balanced scorecard by selecting the best key performance indicators when sufficient data is not yet available. These works use other techniques – such as ANP [17] or fuzzy logic [18, 19]. Although the effectiveness of these methods proves that it works in the construction of new information systems as a proactive approach, this document focuses on making the most of the data available from existing information systems.

2.3. Regression methods

2.3.1. Multiple linear regression

MLR has been used to quantify the effect of input metrics on the output [12, 15] with good results in terms of model predictability (R^2). However, the main objective of the present study, which is to discover systemic relationships, can be compromised by the effect of collinearity. MLR when affected by collinearity, which can be measured by the variance inflation factor (VIF), can produce an unstable model since coefficients are overestimated when VIF > 5. In addition, the MLR, as a regression technique, must assume cause and effect relationships between the variables before evaluating the model, which are not sufficiently clear in this case, at least as a starting point.

2.3.2. Partial least squares

For complex models (e.g., high-order constructs) or cases with multi-collinearity, PLS is more appropriate [20]. Moreover, PLS can be used even if the number of observations is smaller than the number of variables to study [13]. However, the uncertainty of the construct in the initial stages of the study is the most difficult obstacle to overcome [20]. Rodriguez-Rodriguez et al. [13] highlighted this uncertainty in a study where the research team had to evaluate different constructs together with the team of the board of the company where the study was made.

Although PLS is generally the preferred method when a regression analysis is required, MLR also has some points in its favour, such as the possibility of evaluating non-linear relationships between predictors and dependent variables. PLS is a multivariate technique, so it uses linear algebra, and although the transformations of the variables can be used to explain nonlinear relationships, it is not recommended, since the number of variables increases exponentially, and multivariate techniques are not adequate for such models in practical terms [21].

2.3.3. Simple linear regression

Simple linear regression (SLR) can also be an option when the problem is to understand the relationships between different levels or dimensions and only two variables are being studied. However, depending on the nature of the problem, several regression techniques can be applied, and the practitioner will always have to consider the principle of parsimony (which is to keep the model as simple as possible). The principle of parsimony can generally be considered a good guide when applying statistical tools [22, 23]. However, in social sciences, Gunitsky [24] recommends distinguishing between three different views of the concept according to the objective. He emphasises the epistemological conception of parsimony – abstract from reality – to highlight recurring patterns and construct verifiable propositions. Therefore, Gunitsky [24] suggests that to prove a specific hypothesis, the principle of parsimony is justified, coinciding fundamentally with Coelho et al. [22] and Nalborczyk et al. [23].

2.4. Principal component analysis

Several studies [13, 14, 15] have shown that PCA is an effective tool for selecting KPIs. Bi-dimensional plots of principal components can be used to screen the main KPIs for their weight, but also to perform a more comprehensive correlation analysis than just looking at the table with the loads of each variable for each component. Rencher [25] pointed out that this analysis can be an integral result by itself if a qualitative analysis is carried out together with the quantitative analysis.

2.5. ANN and other data mining techniques

ANN and other data mining techniques are more suitable in big data contexts [26], since these methods work well when number of instances is much bigger than the number of variables (KPIs), which in principle is not the case when dealing with KPIs of the BSC. In addition, ANN does not provide an explicit regression equation compared to other regression techniques, which was considered essential for the purpose of this research. Therefore, the present methodology does not use data mining methods such as ANN.

2.6. Quality management systems in manufacturing and the BSC

The main studies on QMSs are more qualitative than empirical and analytical [27, 28, 29], mainly in the manufacturing sector [30]. Although the quantitative analysis was performed in the QMS, the approach was to generate a construct using PLS-SEM or CB-SEM techniques based on established theoretical frameworks [30, 20].

Norreklit [31] points out that one of the main problems of the balanced scorecard model is the assumption of fixed cause and effect relationships between variables of different dimensions. Instead, she proposes a model with systemic relations where the different dimensions do not have a defined hierarchy or a fixed model. She also mentions the problem of potential delayed effects on the system of some variables. Kaplan [5] recognises that these problems can be present in the model and invites the scientific community to study how they can be discovered and thereby improve the model using analytical techniques and empirical systems dynamics. Hoque [6], in a comprehensive review of the use and limitations of the balanced scorecard, suggests that the existence of potential trade-offs between KPIs from different dimensions or levels is among the most cited unresolved problems.

2.7. Time series techniques

Time series techniques should be applied to address and solve the problems that this type of data tends to have. The most common problems are autocorrelation or working with non-stationary time series. A hybrid method that combines analytical and graphical tools is the most convenient in these cases [15].

2.8. Synthesis of the literature review

Table 1 summarises the literature review on the existing methods explained in detail in the previous sections. The tools and techniques selected for the proposed methodology are underlined. This selection is based on the characteristics of each tool and those of the problem addressed in this study.

Tool / technique	Type of data	Multivariate/ univariate	Suitable for variable selection	Typical applications
SLR	Actual data	Univariate	Yes	Hypothesis testing, causal models
MLR	Actual data	Univariate	Yes	Medium-complexity models
PLS	Actual data	Multivariate	No	High-complexity models / machine learning
PCA	Actual data	Multivariate	Yes	Feature extraction
ANN	Actual data	Multivariate	No	Deep learning / data mining
Fuzzy logic	Subjective data (from experts)	Multivariate	Yes	Proactive methods / decision support systems
ANP	Subjective data (from experts)	Multivariate	Yes	Decision support systems
Time series	Actual data (time domain)	Both	Yes	Data pre-processing, econometrics, forecasting

Table 1. Analytical methods

In the next section, the method used to carry out the study is presented as a multi-phase model. This method was designed to include all the characteristics and, as far as possible, improve the limitations of the techniques selected from those identified in the literature review.

3. Data and methods

The methodology developed has been tested in a case study approach using real data from two full years of the balanced scorecard of a leading manufacturing company. The company where this work was done considers the raw data used to be confidential and its representatives and the university research team signed a confidentiality agreement. For this reason, this paper only shows the result of the statistical analyses, but not specific values of the key performance indicators of the QMS. To preserve its confidentiality, the scale of the original data has been changed by dividing all data points in the entire original dataset by the same figure. It has been confirmed that by dividing by the same number, all analyses give the same result with the original and the transformed data, since the scales change, but not the ratios between the KPIs. This paper provides the reference to the transformed dataset to allow replication of the main results shown in section 4.1. To ease interpretation, Table 2 shows detailed definitions for all the KPIs used in the study.

KPI	Designation	Units	Definition
D1000 or D1000 ONLINE	Online defects per thousand units	# of defects / 1000 units	Number of defects detected online at any production stage every 1000 units produced
EL D1000	End of line defects per thousand	# of defects / 1000 units	Number of defects detected at the end of the production line every 1000 units produced
EL FTT	End of line first time through	%	Proportion of units produced without defects that need offline repairs detected at the end of the line
FTT	First time through	%	Proportion of faultless produced units that need offline repairs detected at any stage of the production line
ONLINE or ONLINE %	Online percentage	%	Proportion of units repaired online with at least one defect
PA D1000	Final product audit defects per thousand	# of defects / 1000 units	Number of defects detected in the final product audit every 1000 units
PA FTT	Final product audit first time through	%	Proportion of faultless units needing offline repairs detected in the final product audit
PA ONLINE	Final product audit online percentage	%	Proportion of units repaired online with at least one defect detected in the final product audit
PA TGW	Final product audit things gone wrong	# of claims / 1000 units	Number of customer claims per thousand units predicted based on the severity and probability of defects detected in the final product audit
PA TGW A	Final product audit things gone wrong type A	# of claims / 1000 units	Number of customer type A claims per thousand units predicted based on the severity and probability of defects detected in the final product audit
PA TGW AB	Final product audit things gone wrong type A and B	# of claims / 1000 units	Number of customer claims of type A and B per thousand units estimated based on the severity and probability of defects detected in the final product audit
PA TGW B	Final product audit things gone wrong type B	# of claims / 1000 units	Number of customer type B claims per thousand units estimated based on the severity of defects detected in the final product audit
R1000 0MIS	Repair per thousand at zero months in service	# of claims / 1000 units	Number of customer claims per 1000 units due to repairs at zero months in service after product sale
R1000 1MIS	Repair per thousand at one month in service	# of claims / 1000 units	Number of customer claims per 1000 units due to repairs at one month in service after product sale
R1000 3MIS	Repair per thousand at three months in service	# of claims / 1000 units	Number of customer claims per 1000 units due to repairs at three months in service after product sale

Table 2. Definition of the quality management system KPIs

The multi-phase methodology is shown in Figure 1 and the details of each phase are explained below.

The statistical analyses were performed using the statistical software packages Minitab, Stata, and the data analysis tool of Excel.

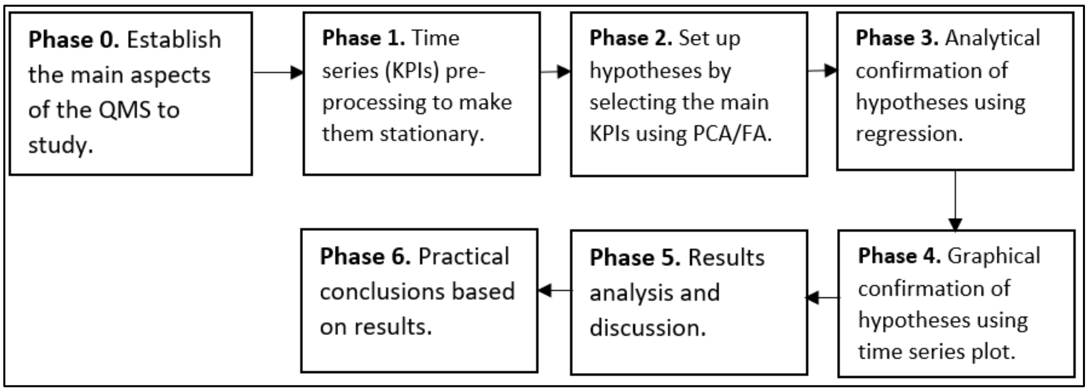

***Figure 1**. Multiphase methodology of the study*

In **phase 0**, the research team together with company experts established that the main aspects of the study were the 'predictability of the quality system' and the 'feedback capability of the quality system'. The predictability of the quality system can also be understood as the ability to control customer satisfaction through internal KPIs. If there were internal KPIs with good predictability, causality, or correlation with external KPIs (related to customers), it would be easy to implement strategies to improve customer satisfaction indexes.

Quality feedback is the ability of the system to recalibrate internal controls in an environment of continuous improvement. The ability to recalibrate quality inspection is vital to ensure the system continues predicting, reacting, and preventing future customer complaints.

In **phase 1**, the raw data must be processed before starting statistical analyses [15]. The main problems when dealing with time series (KPIs) are the autocorrelation and the seasonality of the data. The time series must be stationary before performing statistical analyses that use correlation or regression [32, 33]. Sanchez-Marquez et al. [15] use the Dickey-Fuller analytic t-test augmented for stationary time series [34, 35] complemented by a graphical analysis of the time series with the time series chart, the autocorrelation function (ACF), and the partial autocorrelation function (PACF) [32, 33]. If any sign of non-stationarity is observed, a transformation of the unprocessed data must be performed to obtain a stationary time series. The most common transformation is to take differences, but in some cases, other transformations are needed, such as the logarithmic ones [32, 33, 35, 15].

The main objective of **phase 2** is to select the main KPIs that explain most of the variability observed. Rodriguez-Rodriguez et al. [13] use the two-dimensional plot of the PCA to select those KPIs with the highest loadings (coefficients) before performing regression analyses (PLS). This eliminates the noise produced in the system by the discarded variables, which results in a more precise estimation of the regression coefficients. In this paper, this quantitative analysis is complemented with a qualitative analysis using the vector view of the two-dimensional plot. As shown in the results section, the closer the vector direction, the more similar are the variables explained by those vectors. This means that there is a high correlation between variables represented by vectors with close directions. These variables with high correlation together with the known manufacturing flow (Fig. 2), which establishes the cause-and-effect relationships between the variables, are used to carry out the regression analyses in **phase 3**, which will quantify the variables relationships in terms of strength (regression coefficients) and stability (R^2-predictive).

As mentioned by Rencher [25], the PCA can be a result in itself when the objective is a descriptive or qualitative analysis. Starting with the data matrix (multidimensional observations), the variance-covariance matrix (usually called the covariance matrix as its shortest form) can be computed as follows:

$$\boldsymbol{S} = \frac{1}{n-1}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}} \tag{1}$$

where:

- $\boldsymbol{S}$ is the covariance matrix.
- n is the number of observations or multidimensional instances
- $\widetilde{\boldsymbol{X}}$ is the data matrix centred by subtracting from each data point the mean of each variable (column). Therefore, $\widetilde{\boldsymbol{X}} = \boldsymbol{X} - \boldsymbol{1}\bar{\boldsymbol{x}}'$, where $\boldsymbol{X}$ is the raw data matrix, $\boldsymbol{1}$ is a column vector composed of n observations or instances, and $\bar{\boldsymbol{x}}'$ is the row vector composed of the means of the m variables in the study. Therefore, since $\boldsymbol{X}$ is an n x m matrix, $\widetilde{\boldsymbol{X}}$ is also an n x m matrix, where n is the number of multidimensional instances or observations, and m is the number of variables considered in the study.

Since $\boldsymbol{S}$ is a square and symmetric matrix, the Eigen analysis can be performed to obtain the eigenvalues and eigenvectors. According to Peña [21], this can be shown in its matrix form as follows:

$$\boldsymbol{SU} = \boldsymbol{UD} \tag{2}$$

where:

- $\boldsymbol{S}$ is the covariance matrix
- $\boldsymbol{U}$ is a square matrix where each value u_{nm} represent the loadings or coefficients of the original m variables in each principal component (p components). The principal components (also known as latent variables) are the column vectors.
- $\boldsymbol{D}$ is a diagonal matrix where each diagonal value (λ_p) represents the eigenvalue of each p component.

Initially, from the Eigen analysis, we obtain the same number of components as original variables (p = m), since $\boldsymbol{U}$ is square. In practical terms, the eigenvalues of some of the components are almost zero ($\lambda \approx 0$), because some variables are not linearly independent of others (high correlation between the variables), so $p \leq m$ and this implies a reduction of complexity.

Since $\boldsymbol{U}$ is a square matrix composed of orthogonal vectors [21], then $\boldsymbol{U'U}=\boldsymbol{U}^{-1}\boldsymbol{U}=\boldsymbol{I}$. If one pre-multiplies (2) by $\boldsymbol{U'}$ on each side of the equation, then

$$\boldsymbol{U'SU} = \boldsymbol{D} \quad (3)$$

and therefore

$$\boldsymbol{S} = \boldsymbol{UDU'} \quad (4)$$

Equation (4) is known as the spectral decomposition of the covariance matrix [21]. The covariance matrix is decomposed into orthogonal vectors (principal components) where each explains a certain amount of variance (λ_p). Therefore, all the variance observed in the original data can be explained by these new variables (components/dimensions).

To obtain the value of the new variables in each observation (principal component scores), the original variables must be projected in the new space, which normally has fewer dimensions due to the reduction in complexity explained above, therefore

$$\boldsymbol{T} = \widetilde{\boldsymbol{X}}\boldsymbol{U} \quad (5)$$

where $\boldsymbol{T}$ is a matrix n x p that represents the projected observations in the new space. Note that, as explained above, $p \leq m$ due to the reduction in complexity.

Rodriguez-Rodriguez et al. [13] only use the coefficients ($\boldsymbol{U}$) as the weight to select the variables. Since an original variable can be projected in more than one component, the original variables are characterised by their coefficients and by its direction when they are projected. Therefore, the present method uses the vector view as a graphical method, not only the coefficients.

Peña [21] and Rencher [25] recommend using the correlation matrix instead of the covariance to perform PCA when the variables have different scales, which is a way of standardising the scale of the variables. The balanced scorecard, including each of its operating systems, is composed of heterogeneous groups of variables; therefore, this method must use the correlation matrix as follows:

$$\boldsymbol{C} = \boldsymbol{PLP'} \quad (6)$$

where

- $\boldsymbol{C}$ is the correlation matrix, where the elements outside the diagonal are the correlation coefficients between the variables and the elements of the diagonal are all equal to one.

- ***P*** is a square m x p matrix (square since initially p=m), which represents the standardised loadings / coefficients.
- ***L*** is the diagonal matrix where the values in the diagonal (eigenvalues) represent the amount of variance explained by each principal component. In this case, the variance is standardised as well.

Therefore, using ***C*** instead of ***S*** also changes the scores of the principal components (the new projected variables) from absolute to standardised units. To compare and select variables, which is a qualitative analysis, it is recommended to use the standardised units (***C*** instead of ***S***) when scales are different as already mentioned [21, 25]. Since the KPI scales are typically different, the present method should use the correlation matrix (***C***) to extract the principal components. However, once the selection is made (**phase 2**), start with the regression analysis **(phase 3**) as the objective is usually to interpret the coefficients in absolute terms – rather than just the statistical significance (p-value vs. α) and the predictive power (predictive R^2). The study must be done with the original variables and so their original units must be used (original scales). The present method uses regression analysis in this sense, and therefore the original scales of the variables are used. However, other methods use, for instance, multivariate regression analysis as PLS for qualitative analysis. In these cases, the dichotomy of standardised versus non-standardised is present, and researchers have to decide on the objectives of the study and the nature of the variables. Marin-Garcia & Alfalla-Luque [20] make an in-depth analysis on this topic and propose a series of recommendations for researchers using the PLS analysis.

Since a two-dimensional vector chart can only represent two dimensions, the method uses the two first principal components, $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$. A verification of the variability explained by these two components is needed to ensure that the variance is at least 80% of the total [21]. For practical reasons, if the variation is not 80%, but is close, it is advisable to use the first two components. As part of this method, when more than two components are needed, factor analysis (FA) can be used instead of PCA [36]. First, according to Joliffe & Morgan [36], it is necessary to select the number of components (explaining at least 80% of the total variance) and rotate the vectors, usually using the 'varimax' rotation method, which facilitates the interpretation of the results. However, wherever possible, bi-dimensional vector visualisation is recommended, since a graphical method is always more intuitive, mainly, considering that the results are interpreted not only by the researchers, but also by company staff. The use of the 'varimax' rotation, which maximises the variance explained by the new projected variables (called factors instead of components in FA), is equivalent to using the direction of the vectors when using the two-dimensional plot. These new coefficients are maximised when they are rotated and so the effect of having the original variables explained by several components or factors is solved, or at least minimised [36].

From the two-dimensional plot, the variables are selected according to weight criteria and correlation (vectors in the same direction, regardless of the sense) and considering which hypotheses are related to the aspects established in phase 0 – predictability and feedback of the QMS.

Once the variables are selected, a regression analysis is performed in **phase 3**. Following the principle of parsimony, the simplest regression technique is selected to test the hypotheses. The hypotheses related to the predictability of quality will always be a cause and effect relationship between the internal and external variables in the direction from inside the company towards the customers (outwards). The quality feedback hypotheses go in the other direction (inwards).

In this phase, the principle of parsimony is not the only aspect to select the simplest technique. Simple linear regression (SLR) models can be represented graphically; however, when there is more than one predictor in the model, the graphical representation is not clear or is not possible.

The practical application of the principle of parsimony is to select the simplest possible model, i.e. with as few variables as possible. The application of this principle will ensure that the selected model is the easiest to interpret, which is essential for the objectives of the methodology. On the other hand, a good quality of the model in terms of a high R^2 must be ensured. Therefore, if two regression models are comparable in terms of predictability (R^2), the simplest will be selected.

Akoglu [37] provides guidance for deciding the strength of the relationship between variables based on the correlation coefficient (ρ). Since it is well known that in simple linear regression $\rho^2 = R^2$, for each value of ρ we can compute an equivalent for R^2. Although this relationship can only be proven mathematically for SLR, the same interpretation of R^2, at least in terms of strength (quality), can be used for any regression model. Table 3 summarizes the criteria to decide between different models.

Strength	**Correlation**	**Regression**
Very strong	$0.8 < \lvert\rho\rvert \leq 1$	$64\% < R^2 \leq 100\%$
Strong	$0.7 < \lvert\rho\rvert \leq 0.8$	$49\% < R^2 \leq 64\%$
Moderate	$0.5 < \lvert\rho\rvert \leq 0.7$	$25\% < R^2 \leq 49\%$
Weak	$0.3 < \lvert\rho\rvert \leq 0.5$	$9\% < R^2 \leq 25\%$
Negligible	$0 \leq \lvert\rho\rvert \leq 0.3$	$0\% \leq R^2 \leq 9\%$

Table 3. Interpretation of ρ and R^2

In **phase 4**, the hypotheses proven/disproven by the regression models are confirmed by graphically comparing the behaviour of the time series of the variables included in the regression models. If there is correlation, the regression model is significant (p-value < α) and the predictability power of the regression model is at least moderate according to table 3. It can then be said that there is a good model. If there is a good model, the behaviour of the variables and, therefore, of the time series should be similar. For each significant regression model whose strength is moderate to very strong, we will confirm that the behaviours of KPIs are similar by comparing the trends of the time series charts of each KPI included in the model – see figures 9, 16, and 17. This will help make the decision that the strength of the relationship is not only mathematical, but practical. Like any graphical analysis, it is essentially qualitative, since the confirmation of the quality (strength) of the regression model will depend on the nature of each KPI and its practical meaning. Therefore, the management team will conduct the analysis with the support of data analysts.

To graphically compare KPIs that have different scales, the size of the chart bars should be the same regardless of the range shown by the data, so KPIs can be compared in terms of trends regardless of the scale of the data. In practice, this can be done using automatic chart scaling that most computer packages with graphical tools incorporate.

In **phase 5**, researchers together with subject matter experts (SME) from the company discuss the results in detail. The main objective is to develop a specific statement for each significant regression model confirmed in the previous phase. This declaration should include an interpretation of the regression coefficient and the strength of the model based on Table 3.

For example, if we have the following regression model with p-value < 0.05 and R^2-pred = 76.86%:

R1000 0MIS=15.52-0.1966 *FTT*, the team will present the following statement:

'It has been found that there is a very strong relationship between the internal KPI of first time through (*FTT*) and the external KPI for warranty repairs at zero months in service (*R1000 0MIS*). A 1% increase in FTT causes a decrease of approximately 0.2 warranty repairs per 1000 units sold.'

Finally, in **phase 6**, these discussions are summarised in solid and practical conclusions with the aim of proposing strategic changes to improve customer satisfaction – which is the ultimate goal of the QMS.

The team must draw at least two main types of conclusions, one based on the confirmed regression models between internal and external KPIs, and one based on KPIs of the same stage, either external or internal – see figure 2 for process stages. The latter will be based on strong or very strong relationships based on the correlation coefficient. For example, if a strong relationship is confirmed between two external metrics, as they belong to the same stage (the customer's), it is not a causal relationship, but a correlation. The same could happen for internal metrics of the same stage.

Only those KPIs that appear in the confirmed regression models will be considered as strategic, therefore the management team should exclude the rest. In addition, the team will reduce the complexity of the QMS by choosing only one KPI for each strong or very strong correlation between metrics of the same stage. These decisions will lead to a simplified QMS composed with KPIs with a strong impact on customer satisfaction.

4. Results of the case study and discussion

The aspects that were selected in **phase 0** of the study, which were the predictability of the quality management system and its feedback capability, have been explained in the previous section. In this phase, it was also decided to separate the study into two sub-studies, one with variables that include all the models produced in the company and the other, by the model.

In the hybrid analysis (graphical and analytical) of the time series [15], corresponding to **phase 1**, the conclusion was that they were stationary series and, therefore, the transformation of the data was not necessary.

Figure 2 shows the process flow of the case study and locates each group of KPIs (internal and external). The process flow is necessary to establish input and output variables for the regression analyses of **phase 3**, which is carried out on the KPIs previously selected in **phase 2** (see section 3).

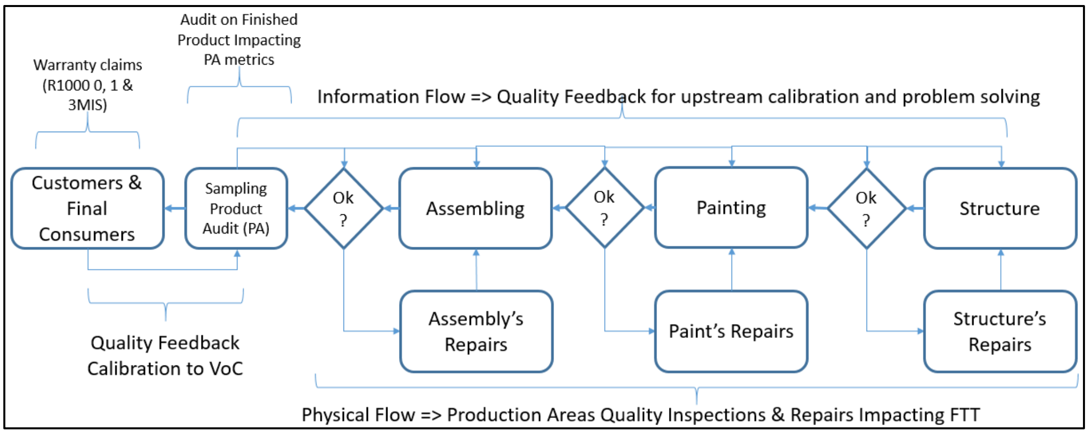

***Figure 2**. High-level process flow of the quality management system*

Phases from **2** to **6** are detailed in the following sections.

4.1. Results including all models

4.1.1. Quality predictability

The predictability of quality is the relationship between the internal metrics and the voice of the customer as measured by warranty repairs at 0 months in service (*R1000 0MIS*), *R1000 1MIS* and *R1000 3MIS*.

Figures 3 and 5 are the bi-dimensional plots of the principal component analysis (PCA). Figures 4 and 6 show the amount of variance explained by each principal component (eigenvalues). In both study periods, bi-dimensional plots could explain about the 80% of the total variance observed [25]. By comparing the period from August 2017 to January 2018 (Fig. 3) to the period from January 2017 to January 2018 (Fig. 5), it can be seen that the relationship between the variables 'online product auditing' (*PA ONLINE*) and 'repairs per thousand at 0 months in service' (*R1000 0MIS*) is not maintained. The more orthogonal the vectors are, the less correlation there is between the variables. It is also denoted by the fact that the predictive R^2 (R^2-pred) was low (<30%) in the period beginning in August 2017. Therefore, when more data points are taken, that relationship disappears because the model is unstable.

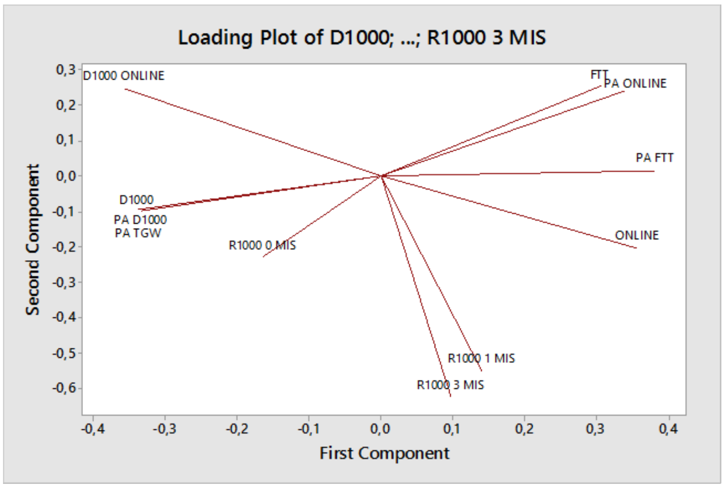

Figure 3*. PCA for all models (data from Aug 2017 to Jan 2018)*

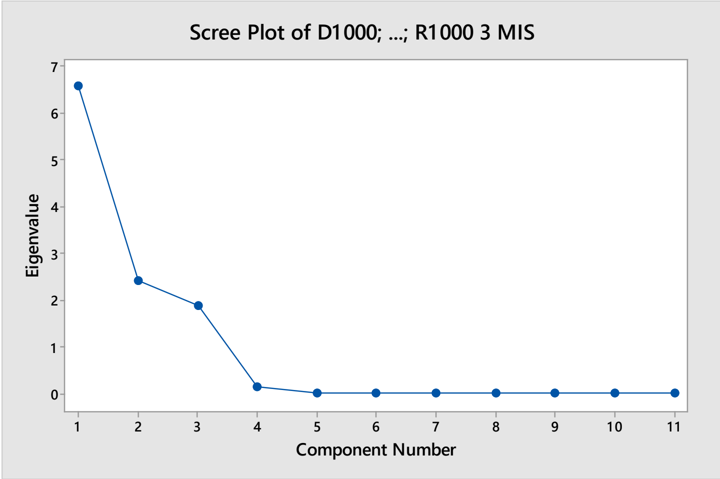

Figure 4. Scree plot (Aug 2017 to Jan 2018). 82% of variance in the two first components

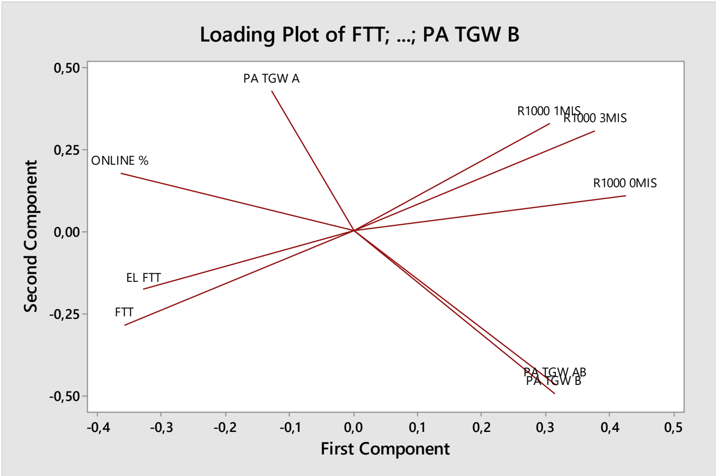

Figure 5. PCA for all models from Jan 2017 to Jan 2018

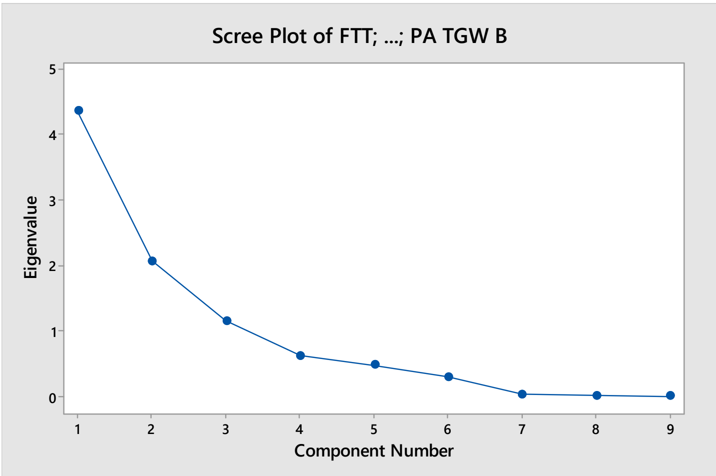

Figure 6. Scree plot (Jan'17-Jan'18). 72% of variance in the two first components

The most powerful relationship that appears is that of warranties with almost all internal metrics – first time through (*FTT*), *end-of-line FTT* (*EL FTT*) and even with on-line metrics, but especially with *FTT*, with a predictive R^2 for the period from August 2017 to January 2018 of 89.3%. For the period beginning in January 2017, R^2-pred was 75%. These values of R^2-pred mean a high predictive power and a high stability of the model.

A good quality of the model implies a good calibration of the internal quality controls with the voice of the customer (VoC). Therefore, the variability in the R^2 could mean differences in the level of calibration within different periods. These changes in the calibration of the internal controls require a recalibration of the quality controls, which is a key function of the quality improvement teams. Another highlight of this result is the potential use of the R^2 of this regression model to evaluate the level of calibration of internal controls in a given period. However, the limitation of sample size will always be present in this type of study, although the possibility of having more data points should also be explored, for example, by increasing the frequency of data points.

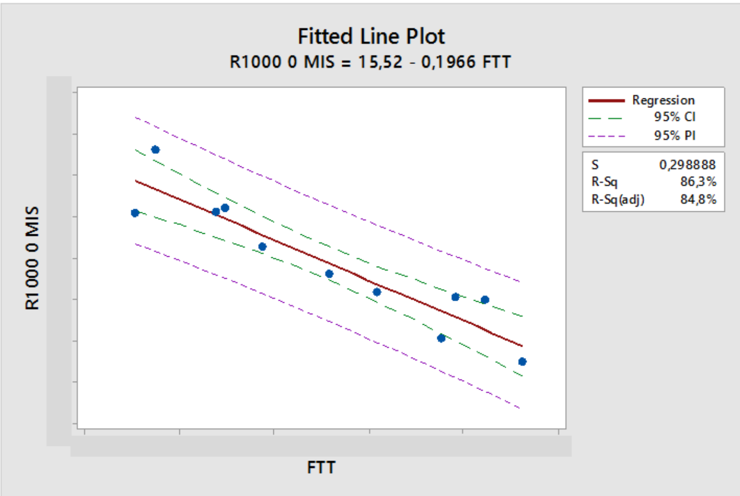

***Figure 7**. All models from Aug'17 to April'18 (R^2-pred=76.86%)*

The regression equation (also shown in Fig. 7) for this model is:

$$R1000\ 0MIS = 15.52 - 0.1966\ FTT \quad (7)$$

In Table 4, a complete analysis of variance and a model summary of the regression analysis of the Figure 7 is presented.

Analysis of variance							
Source	**DF**	**Seq SS**	**Contribution**	**Adj SS**	**Adj MS**	**F-Value**	**P-Value**
Regression	1	5.0753	86.32%	5.0753	5.0753	56.81	0.000
FTT	1	5.0753	86.32%	5.0753	5.0753	56.81	0.000
Error	9	0.8040	13.68%	0.8040	0.0893		
Total	10	5.8793	100.00%				
Model summary							
S	**R^2**	**R^2(adj)**	**PRESS**	**R^2(**pred)			
0.2988	86.32%	84.81%		1.3603	76.86%		
Coefficients							
term	**Coef**	**SE Coef**	**95% CI**	**T-Value**	**P-Value**	**VIF**	
Constant	15.52	1.60	(11.89; 19.14)	9.67	0.000		
FTT	-0.197	0.026	(-0.26; -0.14)	-7.54	0.000	1.00	

Table 4. Analysis of variance and model summary for the period Aug 2017 to April 2018

The coefficient of *FTT* means that an increase of one percentage point in the *FTT* equals a decrease of approx. 0.2 *R/1000 0MIS* and vice versa. However, the extrapolation of the linear function beyond the inference space should be used with caution even with such a high model quality, which would imply assuming that the linearity of the model remains beyond the inference space.

The model shows that there is no need to reach 100% of the *FTT* to eliminate warranty claims at *0MIS (R1000 0MIS)*. Although it is not entirely possible, since the probability model based on continuous distributions and product specifications is asymptotic, the linear approximation is good and thinking of a defect reduction very close to zero in the customer before 100% of *FTT* is not completely illogical. This objective, in relation to the transfer function of the regression model, was established at a certain *FTT* point (not shown due to confidentiality reasons) for this case study. The assumptions of normality, equal variance, and independence of the residuals have been verified to validate the model. The autocorrelation for the independent variables has also been verified by up to 12 lags to rule out the overestimation of the regression coefficient due to the time relationships (lack of independence of the estimators). The assumptions were verified for *FTT* and ‘defects per thousand’ KPIs (*D1000*) – see Figure 8.

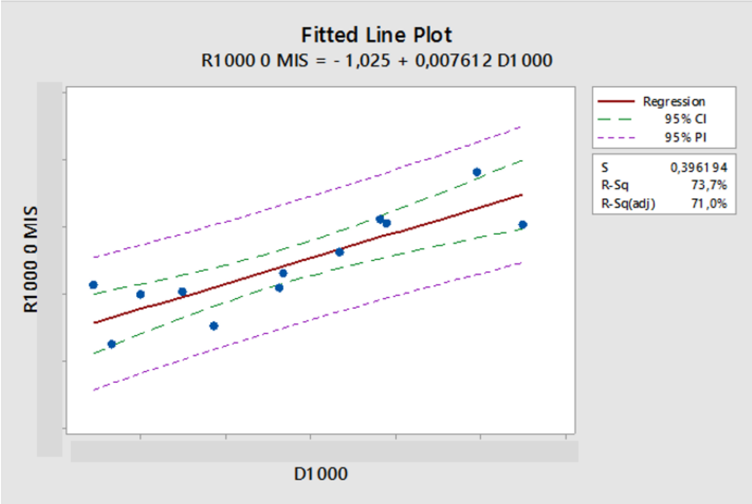

Figure 8. *Regression R1000 0MIS vs. D1000 (Aug'17 – Apr'18) (R^2-pred=57%)*

In Table 5, a complete analysis of variance and a model summary of the regression analysis of the Figure 8 is presented.

Analysis of variance							
Source	**DF**	**Seq SS**	**Contribution**	**Adj SS**	**Adj MS**	**F-Value**	**P-Value**
Regression	1	4.393	73.68%	4.393	4.3933	27.99	0.000
D1000	1	4.393	73.68%	4.393	4.3933	27.99	0.000
Error	10	1.570	26.32%	1.570	0.1570		
Total	11	5.963	100.00%				
Model summary							
S	**R^2**	**R^2(adj)**	**PRESS**	**R^2-pred**			
0.3962	73.68%	71.04%	2.5812	56.71%			
Coefficients							
Term	**Coef**	**SE Coef**	**95% CI**	**T-Value**	**P-Value**	**VIF**	
Constant	-1.025	0.848	(-2.915;0.864)	-1.21	0.254		
D1000	0.0076	0.0014	(0.0044;0.011)	5.29	0.000	1.00	

Table 5. Analysis of variance and model summary for the period Aug 2017 to April 2018

A likely interpretation of this result is that all failure modes at 0 MIS (impact on customer's warranty claims) are the same as those detected within the production facilities during internal verifications (those related to the KPIs of *FTT*, *EL* and *ONLINE %)*. Another possible reason is that the relationship between *R1000 0MIS* and *D1000* remains stable regardless of the chosen study period, which was also confirmed by a regression model. There was a slight fluctuation in the value of the regression coefficient that turned out to be between 0.008 and 0.01. It means that the quality leak can be estimated around that proportion, which is the Type-II error. An improvement strategy may be to reinforce internal quality controls based on objective measures using Gage R & R for both variables and attributes. However, a Type II error of less than 1%

is more than 10 times better (smaller) than the industry average, which is approximately 10%. Negative values of *R1000 0MIS* are not possible, but the negative coefficient of the equation implies that before *D1000* reaches zero we will have zero *R1000 0MIS*, which is the same conclusion as for the equation with *FTT*, due to the linear assumption.

Another point to consider is the relationship between *R1000 1MIS* and *R1000 3MIS*, which also remains constant with an R^2-pred of 80%. This means that both are, in fact, the same indicator, at least in their dynamic behaviour. Both indicators could be summarised – as one or one of them can be eliminated to reduce the complexity of the balanced scorecard.

In the following lines and figures (see Figure 9), as part of **phase 4**, it is graphically confirmed that when there is a good regression model or a high correlation, the dynamic behaviour of the variables on both sides of the equal sign of the equation is very similar, since this method uses time series as variables.

In Figure 9, where the warranties at 0MIS (*R1000 0MIS*) are compared with the complementary of the *FTT*, we can see the correlation between both KPIs in a more intuitive way.

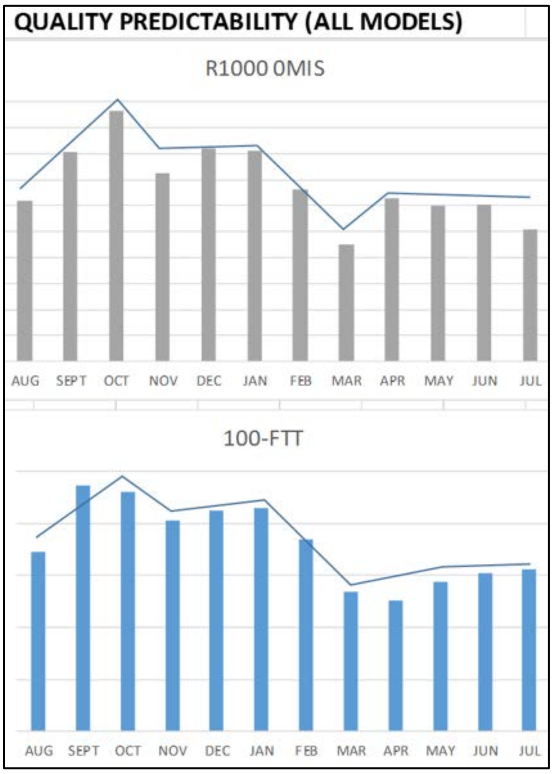

Figure 9*. Graphical confirmation of the predictability of the quality system*

4.1.2. Quality feedback

While the quality predictability can be understood as the ability to predict customer warranties based on internal metrics, the quality feedback is the ability of the system to feed customer claims back to production facilities in the form of quality controls during the audits of finished products (*PA*). These audits, since they are based on small samples, are designed to calibrate the upstream system, but not to predict the behaviour of the market.

To carry out this study, it was necessary to transform some variables, applying a certain time delay. The time series related to customer complaints were transformed with different delays of t-1, t-2 and t-3, which means delays of 1, 2, and 3 months. This transformation allowed the study of the hypothetical delayed correlation between the customer claims and the product audit KPIs (*PA*). Delays of more than three months were also tested in the study although they are not shown here for reasons of clarity. However, the results showed that there were no relationships between the variables with such delays.

In Figures 10 and 11, we can see a clear relationship between R1000 *0MIS t-3* and type B alerts of *PA* (*PA B*) with 70% of R^2-pred, slightly weaker than with *R1000 1MIS t-3* and *R1000 3MIS t-3*, which have an R^2-pred of 50%. With the time series with a delay of less than three months, which is t-1 and t-2, there was no significant relationship; as shown by the analysed data.

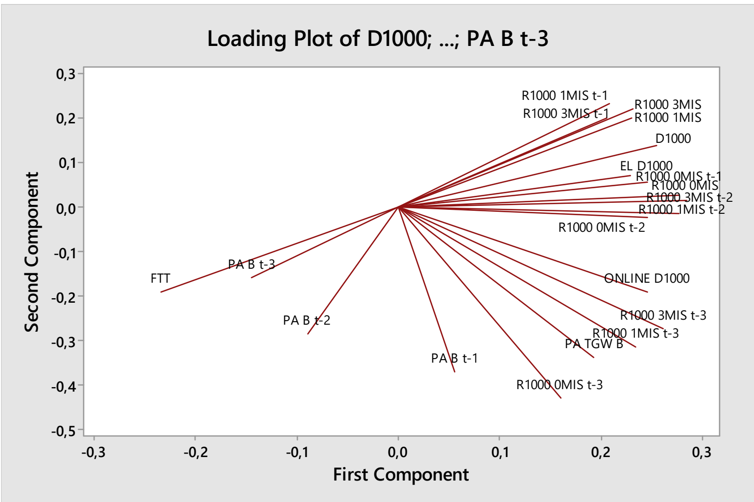

Figure 10*. Quality feedback for all models. From Jan 2017 with lagged variables*

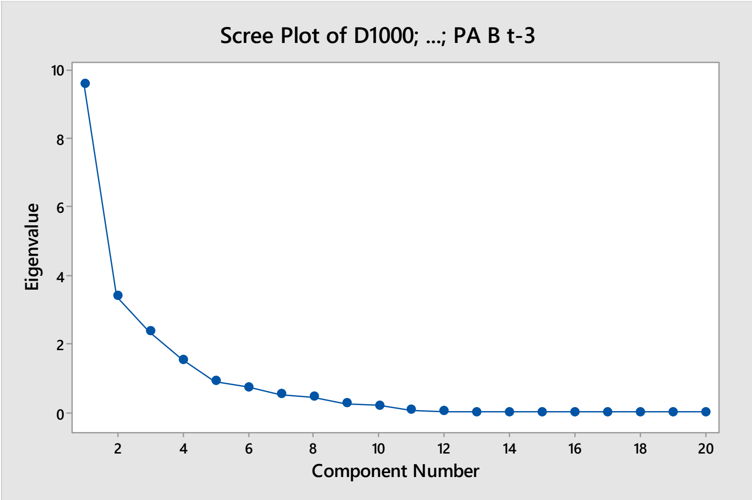

Figure 11. *Scree Plot from Jan 2017. 70% of variance in the two first components*

Figures 12 to 14 show the relationship between customer claims and *PA* in terms of quality feedback.

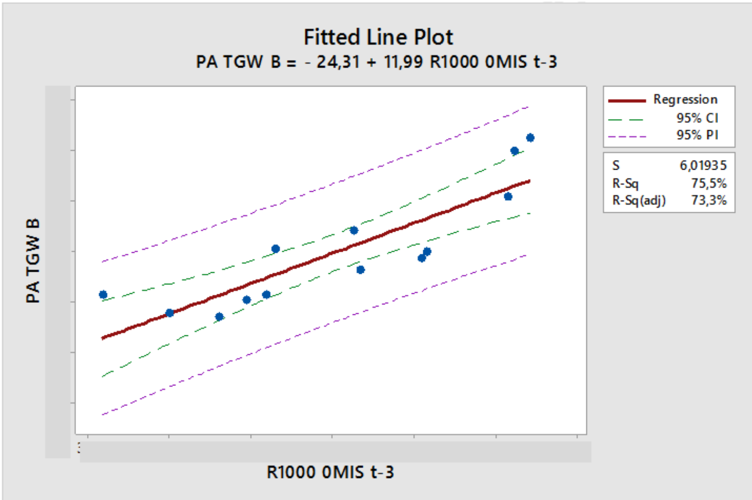

Figure 12. *PA TGW B vs. customer claims at 0MIS after 3 months.* R^2*-pred = 62%*

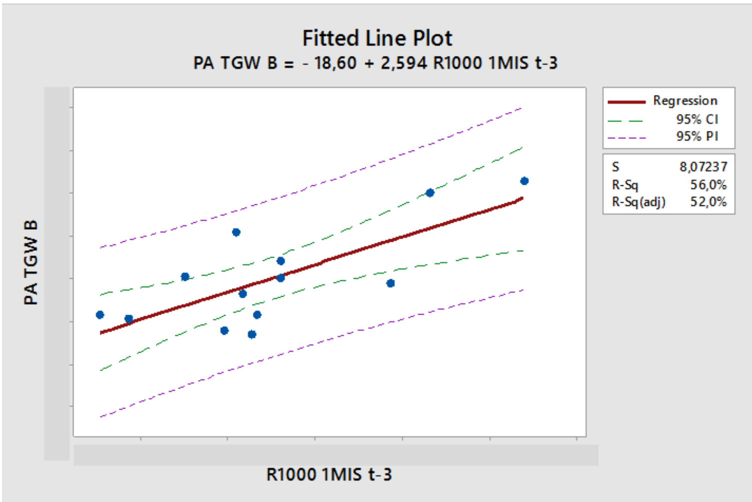

Figure 13. *PA TGW B vs. customer claims at 1MIS after 3 months.* R^2*-pred = 41%*

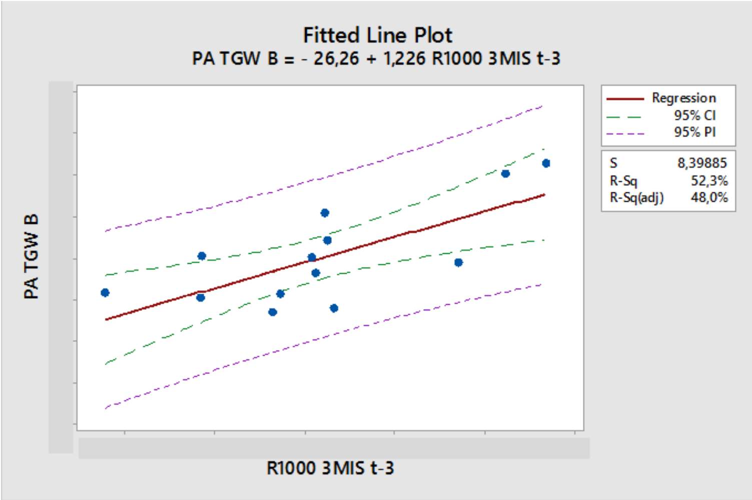

Figure 14. *PA TGW B vs. customer claims at 3MIS after 3 months. R^2-pred=32%*

The main interpretation of these results is that it takes around three months to provide feedback to the product audits. In addition, failure modes claimed by customers at 1MIS and 3MIS do not feed back with the same efficiency to product audits as those at 0MIS. This could be because these failure modes are not based on verifications in the production plant, but in special actions to increase the robustness of the product or in verifications related to reliability. In addition, these failure modes are sometimes latent or functional problems that cannot be detected in regular internal inspections, but only in product audits.

Negative values of *PA TGW B* are not possible, but the negative coefficient tells us that before *R1000 0MIS* reaches zero, *PA* must be zero. This means that product audits do not capture all failure modes. Only after a certain value of *R1000 0MIS* do product audits detect those failure modes three months later.

Before adjusting the simple regression models, a multiple linear regression (MLR) model was tested that included all the variables in the three different MIS (R1000 *0MIS*, *R1000 1MIS* and *R1000 3MIS*) and the quadratic terms. This model was ruled out due to a much lower R^2-pred than the SLR models. In addition, the variance assumptions and the independence of the residuals were verified to validate the regression model.

The model *PA TGW B* = -24.31 + 11.99 *R1000 0MIS t-3* was chosen as the only model valid from a systemic and structural point of view. The reasons were the following:

- When applying MLR and reducing the model using the stepwise algorithm, only the *R1000 0MIS* term remains in the model. Such a result was replicated for the model with and without constant – as well as when using standardised variables and absolute scales. Therefore, the conclusion was always the same – only *R1000 0MIS* remained in the model.

- The coefficient of *R1000 0MIS* is greater than the others, which also means greater sensitivity and power of explanation. The same occurred when using standardised variables.
- It makes physical sense that the 0MIS warranty claims explain most of the *PA* defects.
- The direct correlation between the *PA* indicators and *1MIS* & *3MIS* is lost according to the study period, which is also supported by the evidence shown in Figures 2, 4 and 16. In Figure 16, we can see that there is no clear correlation between *PA* and the warranties, but the correlation between *1MIS* and *3MIS* is never lost regardless of the study period (see also Figures 2 and 4).

However, the fact that, although only in some specific periods, *PA* KPIs may have some relation with *R1000 1MIS* and *R1000 3MIS* could be interesting and may be the objective for a future study on this topic.

Figure 15 summarises the three models in one picture.

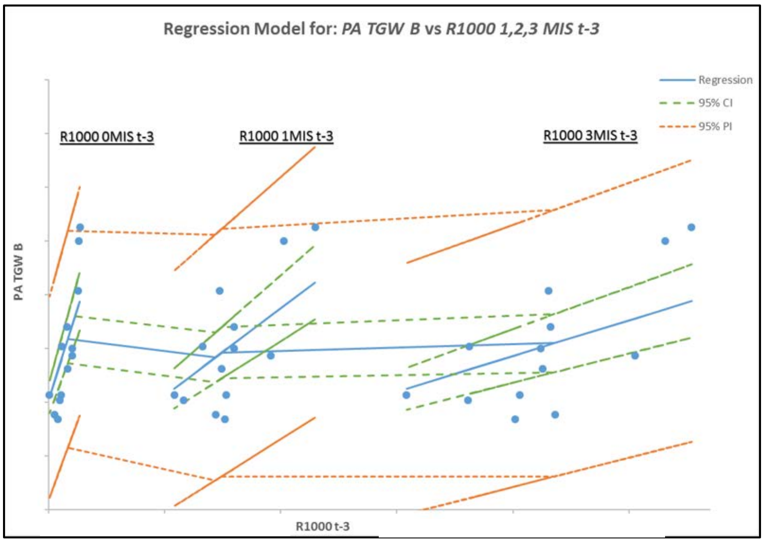

Figure 15. *Regression models for PA TGW B vs. R1000 at 1, 2 and 3MIS t-3*

Figure 16 shows the graphic confirmation of the correlation between *PA TGW B* and *R/1000 0 MIS*.

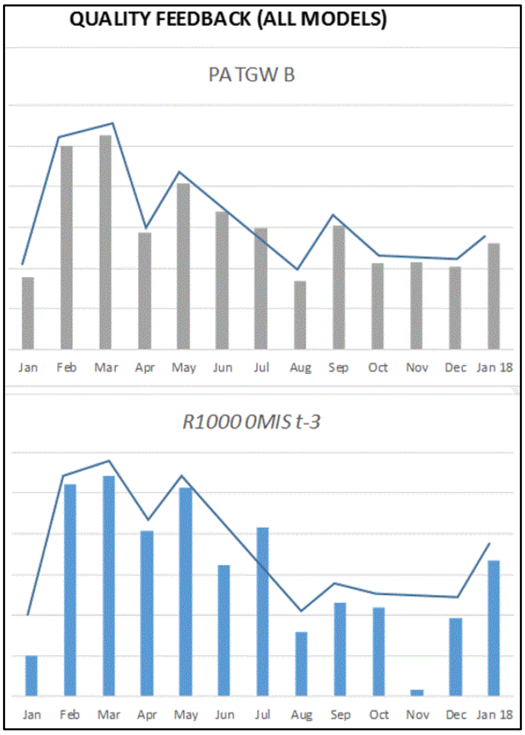

Figure 16. *Correlation between PA TGW B & R1000 0MIS with 3-month delay (t-3)*

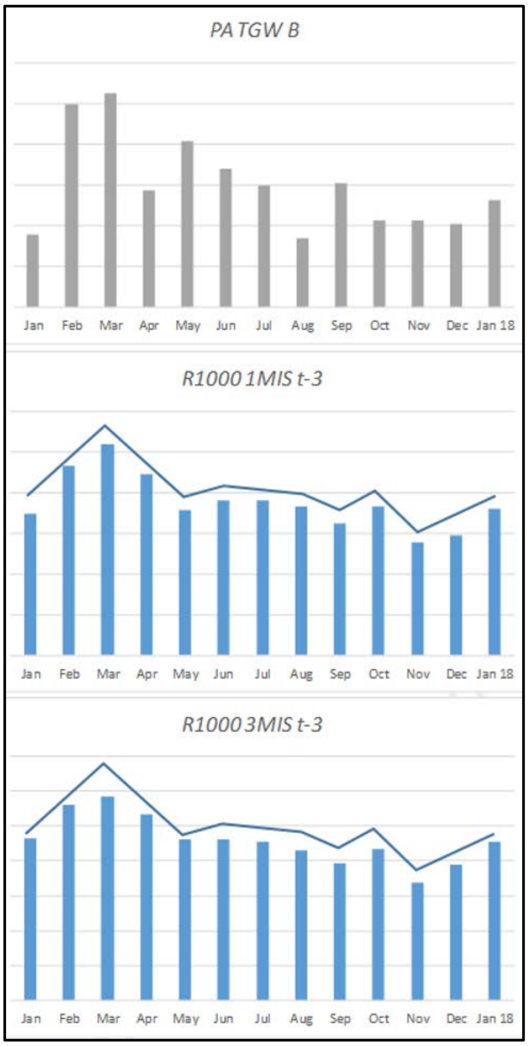

Figure 17. *Correlation among PA TGW B, R1000 1MIS and R1000 3MIS with 3-month delay (t-3)*

Figure 17 clearly shows the absence of correlation between *PA* indicators and *R1000 1MIS* & *R1000 3MIS*. In addition, the correlation between *R1000 1MIS* and *R1000 3MIS* is again evident and has been confirmed in each study period, which means that it is a solid structural relationship.

To validate the models, the assumptions of independence and equality of variance of the residuals were verified. In addition, the presence of autocorrelation of up to 12 delays in the predictors was ruled out.

It is interesting to quantify in a time period the ability to capture the modes of failure of warranty claims. The time period has been estimated as approximately three months and the ability to capture faults per *PA* could be estimated at a rate of 12 for *R1000 0MIS*, 2.6 for *R1000 1MIS*, and 1.23 for *R1000 3MIS*, which are the coefficients of the regression models shown in Figures 11 to 13. The higher the MIS, the lower the detection capacity in PA. Such a conclusion derived from the models is logical, since the higher MIS failure modes are more difficult to detect within the inspections of the production plant.

4.2. Results by model

4.2.1. Quality predictability

Analysis by model gives similar results, although less consistent in terms of stability and the power of relationships between variables. This first unexpected result is probably because the uncertainty due to working with proportions of internal and external metrics is much greater than that of continuous variables. This uncertainty increases as the proportion or size of the sample decreases, so for models with small proportions (defect rate) and/or small production volumes (sample size), the uncertainty of the data increases. Therefore, more data points may be necessary to establish relationships based on regression / correlation techniques.

The above-mentioned characteristic, confirmed by the results, has meant that conclusions of the aspect of quality predictability were only obtained when the relationships between the variables were significant enough. Therefore, it was not possible to obtain any meaningful model for the aspect of quality feedback when the KPIs were split by model.

Figure 18 shows the regression model for the production model A. We can see a similar relationship between *R1000 0MIS* and *EL D1000*. Although there are more relationships between internal and external metrics, the relationship shown is the strongest, regardless of the study period. The regression coefficient is around 0.0206.

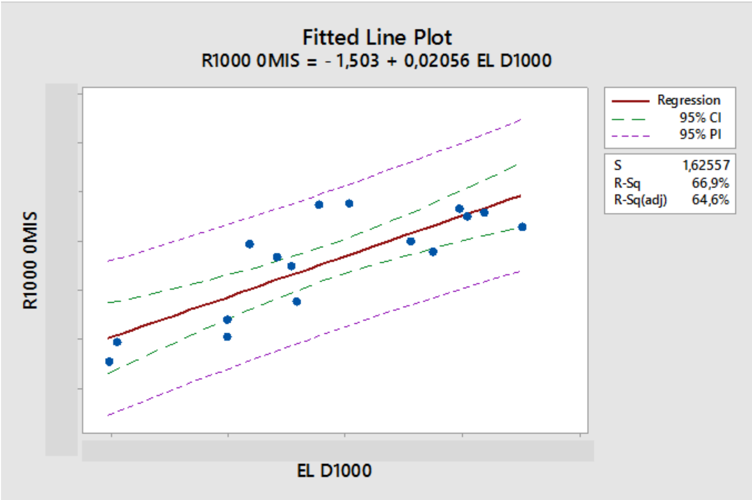

Figure 18. Data for the period from Jan'17 to Apr'18. R^2-pred ≈ 60%

For production model B, it was not possible to confirm such relationships between internal and external metrics. Figure 19 shows some new metrics between different MIS, which, interestingly, were different from what was seen when working with all the models. *0MIS* warranties (*R1000 0MIS*) had a moderate to strong correlation with *1MIS* and *3MIS* (*R1000 0MIS* & *R1000 3MIS*), with a Pearson correlation coefficient of 0.8 ($R^2 \approx 64\%$) for the case of *1MIS* and 0.7 ($R^2 \approx 50\%$) for *3MIS*. A more detailed analysis of the failure mode could establish physical reasons for these relationships if it is confirmed that some related failure modes are appearing in different MIS (at least in this production model).

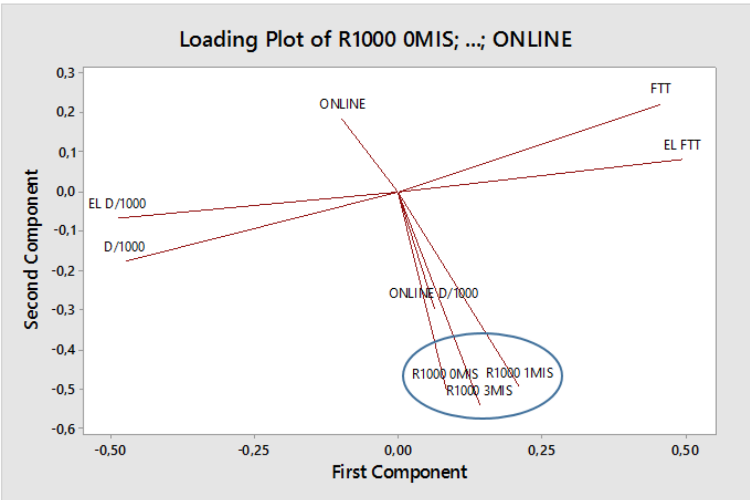

Figure 19. Production model B bi-plot of PCA for the period starting in Aug 2017

Figure 20 illustrates the results for the production model C. A similar relationship was found between *R1000 0MIS* and *D1000*, although its coefficient was only 0.7% and its R^2-pred was slightly greater than 30%. Therefore, it seemed to confirm the relationship between internal and external metrics with a moderate quality of the regression model.

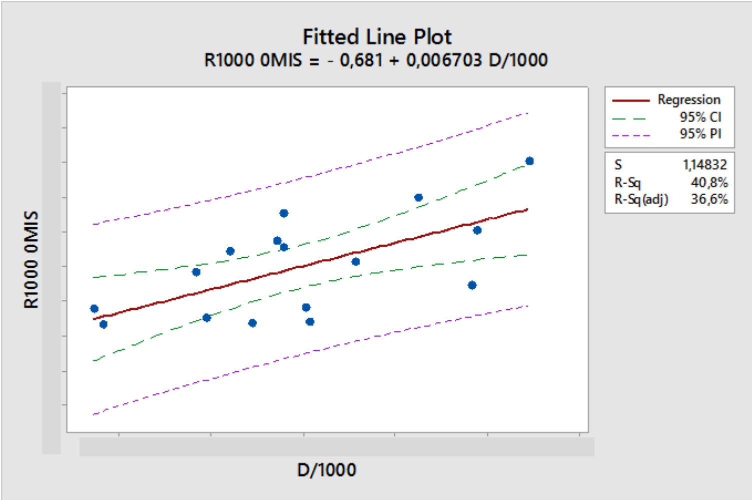

Figure 20. *Regression of R1000 0MIS vs D1000 – R^2-pred ≈ 30%*

Figure 21 shows the results for the production model D. Two relationships between the metrics were found, although the most interesting is that this it is the only model that establishes a correlation of *R1000 3MIS* and an internal metric. It was the *ONLINE* metric expressed as a percentage. This relationship had a Pearson correlation coefficient of 0.636 (R^2-pred of 20%), which can be considered moderate to weak, but with a p-value of 0.019 – although its stability would not be particularly good, and it would have a high-risk level if used to make predictions. Despite this, there were additional correlations that, although weak, were present in other metrics: such as *EL* with a Pearson coefficient of -0.539 and a p-value of 0.057. Based on these findings, we could say that production model D would be the only model where it is possible to detect some failure modes that appeared after three months in service (*3MIS*).

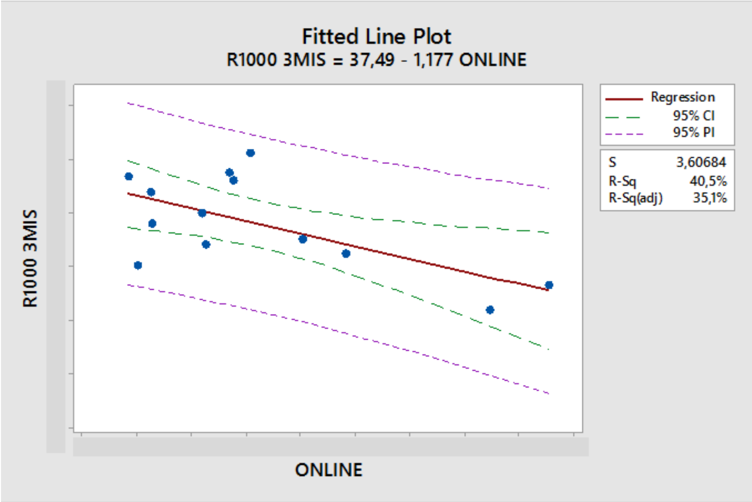

***Figure 21**. Regression of R1000 3MIS vs ONLINE – R^2-pred ≈ 21%*

5. Conclusions

Based on the results, the main conclusions are summarised in the following lines. Two different sections are presented for the aspects of quality predictability and quality feedback.

The executive board of the company followed most of the recommendations made in **phase 6** of this study, which are included in this section. For example, the *FTT* was included in the balanced scorecard for all production facilities around the world and strategies were initiated to improve the *FTT*. The improvement actions derived from these strategies caused the customer quality complaint metrics to improve within a few months. Due to this, the *FTT* was considered as a strategic KPI. In addition, the balanced scorecard was simplified by eliminating the KPIs of *R1000 3MIS* and the quality improvement teams began to only monitor *R1000 1MIS* and this implied a faster reaction time that also meant improvements in the quality KPIs related to customer satisfaction.

5.1. Conclusions on quality predictability

Conclusions on the aspect of the predictability of the QMS can be summarised as follows:

- The stable (structural) and powerful relationship between *FTT* and *R1000 0MIS* was confirmed regardless of the study period and even when using data from different model years.
- Such a strong correlation implies an excellent calibration between the internal quality controls and the VoC.

- Every 2% improvement in *FTT* equals approx. 0.4 *R1000 0MIS*. With *FTT* = 78.94% it is possible to reach the ideal zero *R1000* at *0MIS* (assuming the existence of a linear model).
- There was another strong and stable correlation between *R1000 1MIS* and *R1000 3MIS*. Since $\rho>0.9$, both indicators can be considered as different measures of almost the same thing. Therefore, it would make sense to use only one KPI for the balanced scorecard. The best option is to maintain *R1000 1MIS* and eliminate *R1000 3MIS*, since the KPIs of *R1000 1MIS* are obtained two months previously and the reaction to a deterioration of the metric would be faster.
- The general leakage of defects can be quantified as between 0.8% and 0.9%, which is much better than what is considered a good leak, namely 10% for a Type II error (β Risk).
- This study proved that statistical analyses of KPIs can be used to diagnose the predictability of quality systems in a manufacturing environment.
- Since this method uses statistical tools with real data, it has the limitation of needing a sufficiently sized sample. Future research may focus on changing the data period (measure more frequently) to overcome or minimise this limitation.
- Future research can focus on the generalisation of the method by applying it to the other six management systems.

5.2. Conclusions on quality feedback

Conclusions about the feedback ability of the quality system can be summarised as follows:

- It took three months to provide feedback to the product audits (60 days for data maturity plus 30 additional days for the feedback process itself).
- The strength of the relationships and their stability weakened as we increased MIS. Only the relationship between *PA* and *R1000 0MIS* remained independent of the study period. Therefore, the capacity and stability to capture warranties in product audits was reduced as MIS increased
- Product audits were working as a calibrator of the internal quality system but not as a predictor.
- It was recommended that *R1000 1MIS* appear in the balanced scorecard instead of *R1000 3MIS*. The reaction would be two months faster as *R1000 1MIS* and *R1000 3MIS* were strongly correlated.
- This study proved that the statistical analysis of KPIs can be used to diagnose how the quality management system works in terms of feedback.
- Future research may focus on the generalisation of the method by applying it to other sectors beyond the manufacturing environment.
- Since this method uses statistical tools with real data, it has the limitation of needing enough sample. Future research may focus on changing the data period (measuring more frequently) to overcome or minimise this limitation.

6. Abbreviations

- ACF: autocorrelation function
- ANN: artificial neural network
- ANP: analytical network process
- BSC: balanced scorecard
- CB-SEM: covariance-based structural equation modelling
- EL: end of line
- FA: factor analysis
- KPI: key performance indicator
- MIS: months in service
- MLR: multiple linear regression
- PA: product audit
- PACF: partial autocorrelation function
- PCA: principal component analysis
- PLS: partial least squares
- PLS-SEM: partial least squares structural equation modelling
- PMS: performance management system
- QMS: quality management system
- SLR: simple linear regression
- SME: subject matter expert
- VIF: variance inflation factor
- VoC: voice of customer

7. References

[1] Amaratunga, D., & Baldry, D. (2002). Moving from performance measurement to performance management. Facilities, 20(5/6), 217-223.

[2] Bititci, U. S., Turner, U., & Begemann, C. (2000). Dynamics of performance measurement systems. International Journal of Operations & Production Management, 20(6), 692-704.

[3] Robert S. Kaplan and David P. Norton (2001). Transforming the Balanced Scorecard from Performance Measurement to Strategic Management: Part I. Accounting Horizons: March 2001, Vol. 15, No. 1, pp. 87-104.

[4] Dennis P (2006). Getting the right things done: A learner's guide to planning and execution. The Lean Enterprise Institute, Cambridge, MA, USA.

[5] Kaplan, R. S. (2009). Conceptual foundations of the balanced scorecard. Handbooks of management accounting research, 3, 1253-1269.

[6] Hoque, Z. (2014). 20 years of studies on the balanced scorecard: trends, accomplishments, gaps and opportunities for future research. The British accounting review, 46(1), 33-59.

[7] Malbašić, I. & Marimon, F. (2019). A Simplified Balanced 'Balanced Scorecard'. European Accounting and Management Review, 5(2), 38-60.

[8] Malmi, T. (2001). Balanced scorecards in Finnish companies: a research note. Management Accounting Research, 12(2), 207-220.

[9] Anand, M., Sahay, B. S., & Saha, S. (2005). Balanced scorecard in Indian companies. Vikalpa, 30(2), 11-26.

[10] Junior, I. C. A., Marqui, A. C., & Martins, R. A. (2008). MULTIPLE CASE STUDY ON BALANCED SCORECARD IMPLEMENTATION IN SUGARCANE COMPANIES. Accessed 26 Dec 2016. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.572.3364&rep=rep1&type=pdf
[11] Ferenc A (2011). Balanced Scorecard Measurement applications at a car manufacturer supplier company. Accessed 8 May 2017. https://pdfs.semanticscholar.org/f10e/409533c49dd2934ace78405126978302ab96.pdf.
[12] Grillo-Espinoza, H., Campuzano Bolarin, F., & Mula, J. (2018). Modelling performance management measures through statistics and system dynamics-based simulation. Dirección y Organización, 65, 20-35.
[13] Rodriguez-Rodriguez R., Saiz, J. J. A., & Bas, A. O. (2009). Quantitative relationships between key performance indicators for supporting decision-making processes. Computers in Industry, 60(2), 104-113.
[14] Morard, B., Stancu, A., & Jeannette, C. (2013). Time evolution analysis and forecast of key performance indicators in a balanced scorecard. Global Journal of Business Research, 7(2), 9-27.
[15] Sanchez-Marquez, R., Guillem, J. M. A., Vicens-Salort, E., & Vivas, J. J. (2018b). Intellectual Capital and Balanced Scorecard: impact of Learning and Development Programs using Key Performance Indicators in Manufacturing Environment. Dirección y Organización, (66), 34-49.
[16] Sanchez-Marquez, R., Guillem, J. A., Vicens-Salort, E., & Vivas, J. J. (2018a). A statistical system management method to tackle data uncertainty when using key performance indicators of the balanced scorecard. Journal of Manufacturing Systems, 48, 166-179.
[17] Boj, J. J., Rodriguez-Rodriguez, R., & Alfaro-Saiz, J. J. (2014). An ANP-multi-criteria-based methodology to link intangible assets and organizational performance in a Balanced Scorecard context. Decision Support Systems, 68, 98-110.
[18] Gurrea V, Alfaro-Saiz JJ, Rodriguez-Rodriguez R, Verdecho MJ (2014). Application of fuzzy logic in performance management: a literature review. International Journal of Production Management and Engineering, 2(2), 93-100.
[19] Chytas, P., Glykas, M., & Valiris, G. (2011). A proactive balanced scorecard. International Journal of Information Management, 31(5), 460-468.
[20] Marin-Garcia, J., & Alfalla-Luque, R. (2019). Key issues on Partial Least Squares (PLS) in operations management research: A guide to submissions. Journal of Industrial Engineering and Management, 12(2), 219-240.
[21] Peña, D. (2002). Análisis de datos multivariantes. Retrieved July 5th, 2018, from: http://bida.uclv.edu.cu/bitstream/handle/123456789/12092/Daniel%20Pena%20-%20Analisis%20de%20datos%20multivariantes%20.pdf?sequence=1
[22] Coelho, M. T. P., Diniz-Filho, J. A., & Rangel, T. F. (2019). A parsimonious view of the parsimony principle in ecology and evolution. Ecography, 42(5), 968-976.
[23] Nalborczyk, L., Bürkner, P. C., & Williams, D. R. (2019). Pragmatism should not be a substitute for statistical literacy, a commentary on Albers, Kiers, and van Ravenzwaaij (2018). Collabra: Psychology, 5(1).
[24] Gunitsky, S. (2019). Rival Visions of Parsimony. International Studies Quarterly.
[25] Rencher, A. C. (2005). A review of "Methods of Multivariate Analysis". Retrieved June 30th, 2018, from: https://pdfs.semanticscholar.org/a83c/fec9c23390a10e5c215c375480b8cd3a1565.pdf
[26] He, Q. P., & Wang, J. (2018). Statistical process monitoring as a big data analytics tool for smart manufacturing. Journal of Process Control, 67, 35-43.
[27] Neely, A., Gregory, M., & Platts, K. (1995). Performance measurement system design: a literature review and research agenda. International journal of operations & production management, 15(4), 80-116.
[28] Akkerman, R., Farahani, P., & Grunow, M. (2010). Quality, safety and sustainability in food distribution: a review of quantitative operations management approaches and challenges. Or Spectrum, 32(4), 863-904.
[29] Goetsch, D. L., & Davis, S. B. (2014). Quality management for organizational excellence. Upper Saddle River, NJ: pearson.
[30] Molina-Azorín, J. F., Tarí, J. J., Claver-Cortés, E., & López-Gamero, M. D. (2009). Quality management, environmental management and firm performance: a review of empirical studies and issues of integration. International Journal of Management Reviews, 11(2), 197-222.
[31] Norreklit, H. (2000). The balance on the balanced scorecard a critical analysis of some of its assumptions. Management accounting research, 11(1), 65-88.
[32] Wu, J. P., & Wei, S. (1989). Time series analysis. Hunan Science and Technology Press, ChangSha. Retrieved January 20th, 2018, from: http://www2.geog.ucl.ac.uk/~mdisney/teaching/GEOGG121/time_series/GEOGG121_5_TimeSeries_Wu.pdf
[33] Box GE, Jenkins GC, Reinsel GC (2008). Time Series Analysis Forecasting and Control. New York: John Wiley and Sons.
[34] Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association, 74(366a), 427-431.
[35] Becketti, S. (2013). Introduction to time series using Stata (pp. 176-182). College Station, TX: Stata Press.
[36] Joliffe, I. T. & Morgan, B. J. T. (1992). Principal component analysis and exploratory factor analysis. Statistical methods in medical research, 1(1), 69-95.
[37] Akoglu, H. (2018). User's guide to correlation coefficients. Turkish journal of emergency medicine, 18(3), 91-93.