



Swansea University
Prifysgol Abertawe



Cronfa - Swansea University Open Access Repository

This is an author produced version of a paper published in :
Computers & Industrial Engineering

Cronfa URL for this paper:
<http://cronfa.swan.ac.uk/Record/cronfa29741>

Paper:

Ransing, R. (in press). A bootstrap method for uncertainty estimation in quality correlation algorithm for risk based tolerance synthesis. *Computers & Industrial Engineering*

<http://dx.doi.org/10.1016/j.cie.2016.09.002>

This article is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Authors are personally responsible for adhering to publisher restrictions or conditions. When uploading content they are required to comply with their publisher agreement and the SHERPA RoMEO database to judge whether or not it is copyright safe to add this version of the paper to this repository.

<http://www.swansea.ac.uk/iss/researchsupport/cronfa-support/>

Accepted Manuscript

A bootstrap method for uncertainty estimation in quality correlation algorithm for risk based tolerance synthesis

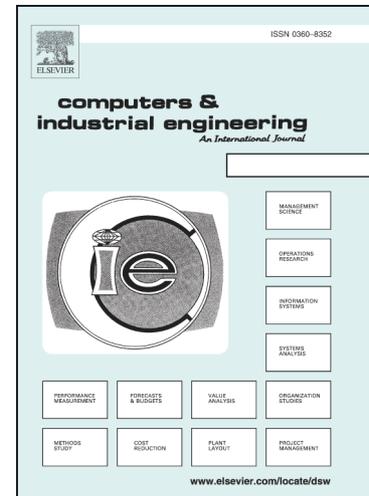
Raed S. Batbooti, R.S. Ransing, M.R. Ransing

PII: S0360-8352(16)30334-5

DOI: <http://dx.doi.org/10.1016/j.cie.2016.09.002>

Reference: CAIE 4458

To appear in: *Computers & Industrial Engineering*



Please cite this article as: Batbooti, R.S., Ransing, R.S., Ransing, M.R., A bootstrap method for uncertainty estimation in quality correlation algorithm for risk based tolerance synthesis, *Computers & Industrial Engineering* (2016), doi: <http://dx.doi.org/10.1016/j.cie.2016.09.002>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A bootstrap method for uncertainty estimation in quality correlation algorithm for risk based tolerance synthesis

Raed S. Batbooti^a, R. S. Ransing^{a,*} and M.R. Ransing^b

^a College of Engineering, Swansea University, Swansea SA1 8EN, United Kingdom

^b p-matrix Ltd, Swansea, United Kingdom

Corresponding author

R. S. Ransing

College of Engineering, Swansea University, Swansea SA1 8EN, United Kingdom

Tel: 0044 1792 295219

Email: r.s.ransing@swansea.ac.uk

ACCEPTED MANUSCRIPT

A bootstrap method for uncertainty estimation in quality correlation algorithm for risk based tolerance synthesis

Abstract

A risk based tolerance synthesis approach is based on ISO9001:2015 quality standard's risk based thinking. It analyses in-process data to discover correlations among regions of input data scatter and desired or undesired process outputs. Recently, Ransing et al. (2016) proposed a quality correlation algorithm (QCA) for risk based tolerance synthesis. The quality correlation algorithm is based on the principal component analysis (PCA) and a co-linearity index concept (Ransing et al. 2013). The uncertainty in QCA results on mixed data sets is quantified and analysed in this paper.

The uncertainty is quantified using a bootstrap sampling method with bias-corrected and accelerated confidence intervals. The co-linearity indices use the length and cosine angles of loading vectors in a p -dimensional space. The uncertainty for all p -loading vectors is shown in a single co-linearity index plot and is used to quantify the uncertainty in predicting optimal tolerance limits. The effects of re-sampling distributions are analysed. The QCA tolerance limits are revised after estimating the uncertainty in limits via bootstrap sampling. The proposed approach has been demonstrated by analysing in-process data from a previously published case study.

Keywords: 7Epsilon, Six Sigma, No-Fault-Found product failures, bootstrapping, in-tolerance faults and in-process quality improvement.

1. Introduction

1.1 Risk Based Tolerance Synthesis

The clause 6.1 of ISO9001:2015 quality standard requires organisations to continually improve the process by enhancing the occurrence of desired process outputs and reducing or preventing the instances when the process has produced undesired results. Except for robust processes, the variation in process inputs may lead to deviation from expected or desired results. The relationship between process inputs and outputs is normally too complex for the tolerance synthesis problem to be modelled by the underlying physics alone. Firstly, the governing equations used to model the physics may not describe the real model accurately and secondly, we may not even know the underlying physics sufficiently (Lewis and Ransing, 2000; Lewis et al. 2004; Pao et al. 2004; Postek et al. 2005). In a continuously monitored manufacturing environment, the synthesis of in-process data can help process engineers to discover subtle relationships among process inputs and outputs (Lewis and Ransing, 1997). Tolerance synthesis is the process of determining allowable variation in products and processes in order to meet the quality requirements (Li, Kokkolaras, Papalambros and Hu, 2008). For

a multi-process manufacturing system, it is essential that the variability in all process inputs (including interactions among process inputs) is analysed to study the variation in one or more process outputs. The tolerance synthesis is usually based on a mathematical model that describes the variation of the process inputs (Ding et al. 2005). A penalty matrix approach is used to estimate the deviation from expected results (Ransing et al. 2013). To embed the risk based thinking in a tolerance synthesis problem, process responses are categorised into three categories (i) desired, (ii) unacceptable and, (iii) a middle region between the two categories. A zero penalty value is assigned to the desired response and a 100 penalty value is given for the unacceptable process response. A process response in the middle region is assigned a penalty value between zero and hundred. A correlation between factor values and penalty values for a given response is discovered using a principal component analysis (PCA) based co-linearity index (CLI) plot (Ransing et al. 2013). The length and angle of each loading vector is calculated in a reduced p -dimensional subspace and is used in a CLI plot. The quality correlation algorithm (QCA) (Ransing et al. 2016) discovers optimal tolerance limits by projecting scores on correlated factors and responses. The scores, bounded between a factor direction and the corresponding response direction in a CLI plot, are used to calculate new tolerance limits for the corresponding factor. For quantitative variables, the range of factor values corresponding to chosen scores defines a new tolerance limit. On the other hand, the optimal and avoid categories for categorical variables are determined by calculating the percentage of occurrences for a corresponding category in the collected scores vector. The obtained projected scores vector leads to an optimal percentage of occurrences if the variable is correlated with low penalty values. For an avoid range, the variable correlates with high penalty values. It is suggested that these new tolerance limits be included in a modified process failure modes effect analysis (PFMEA) table in order to create a reusable organisational knowledgebase (Batbooti et al. 2015).

The number of in-process observations, available for undertaking a tolerance synthesis project, is normally very small (~50-100). The number of input and output variables are of the similar size of number of observations (~50-100) (Ransing et al. 2013). The small sample size can affect the reliability of model predictions, and hence a measure needs to be developed to quantify uncertainty in the model. In the tolerance synthesis context, bootstrapping is based on the notion that the in-process data is representative of the entire population of the data set as the sample size increases to infinity. The QCA estimates population parameters such as the upper and lower tolerance limits for each factors. The novelty and originality of this work is in extending the algorithm proposed by Timmerman et al. (2007) for quantifying the uncertainty in the QCA. The bootstrap parameters used by Timmerman et al. (2007) are different to those used in the QCA. Resampling from the in-process data set is used to imitate the sampling process from the population. The bootstrap parameter distribution is used to quantify the uncertainty by calculating standard errors and confidence intervals. The revised upper and lower tolerance limits of the QCA are derived from the bootstrap parameter values using the weighted mean formulation proposed by Grela (2013) and Finch (2009).

1.2 Uncertainty Estimation with a bootstrap resampling.

Timmerman et al. (2007) have compared the estimation of confidence intervals using the bootstrap approach as well as the asymptotic approach. It was shown that the bootstrap approach was better suited for predicting uncertainty and hence the confidence intervals. The methodology for calculating standard errors and confidence intervals in bootstrap procedures is discussed widely in the literature (Efron, 1977; Efron, & Tibshirani, 1993; Wehrens, Putter, and Buydens, 2000, Hastie, Tibshirani, and Friedman, 2009).

A PCA bootstrap method has been used to find the variances of PCA loadings (Chatterjee, 1984; Lambert, Wildt, and Durand, 1990; Lambert, Wildt, and Durand, 1991). A thousand bootstrapped samples were generated with replacement to determine stopping criteria for choosing number of PC's (Jackson, 1993). Further studies on the finding of the number of retained PCs have been discussed in the literature (Daudin, Duby, and Trecourt, 1988; Besse, 1992; Peres-Neto, Jackson, and Somers, 2005).

Smith and Gemperline (2002) compared two parametric bootstrap methods for analysing small data sets in order to improve the estimation of misclassification rates of microcrystalline cellulose.

A non-parametric bootstrap method was used in an exploratory factor analysis to estimate results of Procrustes rotation to a target pattern matrix (Raykov, and Little, 1999). Bootstrap confidence intervals were estimated for scores and loading values, as well as the global clusters in PCA, to assess the uncertainty (Babamoradi, Van den, and Rinnan, 2013). The study was conducted on two small datasets.

A bootstrap based method is proposed to enhance QCA results by estimating uncertainty in the algorithm for solving risk based tolerance synthesis problems. Table 1 illustrates the symbols used in the paper. The proposed uncertainty estimation method is described in Section 2. A study of one thousand bootstrapped samples is discussed in Section 3 and Section 4 concludes the paper.

Table 1: Nomenclature.

<p>$\hat{\sigma}_e$: Boot strap standard error.</p> <p>\hat{a} : The acceleration or skewness constant</p> <p>\hat{z}_0 : The bias correction.</p> <p>θ : The static of interest</p> <p>B: Number of bootstraps.</p> <p>D_e: Diagonal matrix containing the square roots of eigenvalues.</p> <p>D_s : Diagonal matrix containing the standard deviations of the columns of XT.</p> <p>E: Error matrix.</p> <p>HB: higher the better penalty value settings.</p> <p>L: Loading matrix.</p> <p>LB: lower the better penalty value settings.</p> <p>LL^j: Lower (minimum) value of x_o^j.</p> <p>n_c: number of correlated parameters resulted from applying CLI.</p> <p>n_x^j: length of vector x_o^j.</p> <p>Q: Is the number of original categorical variables.</p>

R : Response.

s_j : Standard deviation of factor j .

$t^{#j}$: The scores vector containing scored bounded between the loading vector j direction and the response direction.

$T^{#}$: Projected score matrix.

T : Score Matrix.

TL^j : Tolerance limit for a factor j .

UL^j : Upper (maximum) value of x_o^j .

V : Matrix of eigenvectors containing eigenvectors as column vectors ordered by greatest eigenvalues.

X : Original data matrix with $(m \times n)$ dimensions, with m is the number of observations and n number of variables.

x_o^j : vector with (o: either optimal or avoid) depending on direction of variable.

X_{Recon} : Reconstruct matrix X from PCA model parameters.

XT : Data matrix after pre-treatment.

z_j : Dummy variable taking binary value one if the categorical variable has been observed and zero otherwise.

2. Bootstrap uncertainty estimation based on the QCA

2.1 The Quality Correlation Algorithm (QCA)

In a given timeframe, each occurrence of process result is recorded and assessed as desired or undesired process outcome. The deviation from the expected results quantified with a penalty value. For a continuous monitoring environment, during the same timeframe the factor values are normally measured at a much higher frequency rate. The median, the average of top the 5% of values and the average of the bottom 5% of values is determined for each factor using values collected in the timeframe and uniquely associated with the penalty value for the given timeframe. In the 7Epsilon context, this dataset is referred to as an equal frequency rate data set. This dataset is stored in matrix X and is referred to as the in-process data matrix with m number of observations and n process variables. The process variables include categorical and quantitative factors and one or more process responses. The data pre-treatment proposed by Giannetti et al. (2014), and shown in Table 2, is applied to the in-process data matrix X to transform this matrix to XT . As shown in Figure 1, this is the first step of the QCA. The second step applies the PCA to the transformed in-process data matrix XT and calculates the corresponding loading (L) and scores (T) matrices in a reduced p -dimensional subspace. This information is used to create a CLI plot and correlated inputs (n_c) with a given response are determined (Ransing et al. 2013). Scores values are projected on n_c loading vectors to create a project scores matrix $T^{#}$. For each factor j , the scores bounded between the loading vector direction and the response direction are stored in vector $t^{#j}$. The corresponding original factor values related to the scores are stored in a vector x_o^j . This variable determines new tolerance limits (TL^j) for factor j . The QCA is used for both categorical and quantitative factors, and for main effects as well as interactions (Ransing et al. 2016).

2.2 Bootstrap QCA

There are three methods for performing bootstrap re-sampling: non-parametric, semi-parametric, and parametric (Babamoradi et al., 2013). The non-parametric bootstrap method is based on re-sampling by a random selection of rows of the data matrix X . In the semi-parametric bootstrap method, the re-sampling is done by random replacement of rows of an error matrix E of the PCA model. In this case, a b^{th} bootstrap sample X_b is generated from $TL^T + E_b$. The parametric bootstrap re-sampling procedure is similar to the semi-parametric method; however, it constructs the error matrix E_b by assuming a suitable distribution.

In the present paper, the semi-bootstrap sampling method is used. The bootstrapping method requires a continuous data matrix hence the bootstrap samples (XT_b with ($b=1 \dots B$) where B is the number of generated bootstraps) are generated from XT rather than the original mixed data set X . In order to calculate UL_b^j and LL_b^j , the original b^{th} data vector $x_{o,b}^j$ corresponding to the projected score vector $t^{\#,j}$ is needed. It is observed that bootstrapping samples are based on the XT_b data matrix rather than the XT matrix. As a result, a reverse data treatment procedure, as shown in Table 3, is defined. Figure 2 summarises the steps for the proposed algorithm.

Table2. Data transformations applied to raw data (Giannetti et al., 2014)

Pre-Treatment steps	Responses	Quantitative Variables	Categorical Variables
Step1: Transformation	$x_{ij} = \begin{cases} 0 & \text{if } x_{ij} \leq Th_{min} \\ 1 & \text{if } x_{ij} \geq Th_{max} \\ (LB) \frac{(x_{ij}^{(k)} - T_{min})}{(T_{max} - T_{min})} & \text{otherwise} \end{cases}$ $x_{ij} = \begin{cases} 0 & \text{if } x_{ij} \geq Th_{max} \\ 1 & \text{if } x_{ij} \leq Th_{min} \\ (HB) \frac{(x_{ij}^{(k)} - T_{max})}{(T_{min} - T_{max})} & \text{otherwise} \end{cases}$	$x_{ij} = \frac{x_{ij} - med_j}{iqr_j}$ $med_j = \text{median of variable } j.$ $iqr_j = \text{interquartile range of variable } j.$	$x_{ij} = \frac{z_{ij} - w_j}{w_j}$ $w_j = \sum_{i=0}^m \frac{z_{ij}}{m}$
Step2: Scaling	$x_{ij}^* = x_{ij} \sqrt{\frac{1}{s_j^2}}$	$x_{ij}^* = x_{ij} \sqrt{\frac{1}{s_j^2}}$	$x_{ij}^* = x_{ij} \sqrt{\frac{w_j}{Q s_j^2}}$

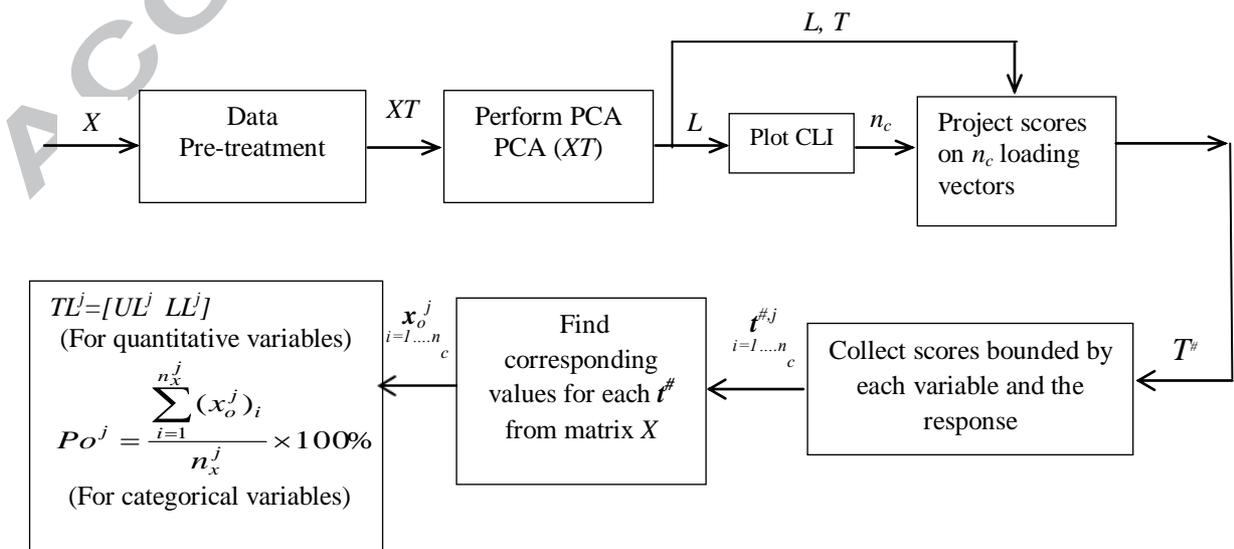


Fig.1 Main steps of quality correlation algorithm

Table3. Steps for a reverse data treatment procedure.

Inverse Weighting steps	Quantitative Variables	Categorical Variables
Step1: Inverse Scaling	$x_{ij}^* = \frac{x_{ij}^*}{\sqrt{s_j^2}}$	$x_{ij}^* = \frac{x_{ij}^*}{\sqrt{Q s_j^2}}$
Step2: Inverse Transformation	$x_{ij} = x_{ij}^* iqr_j + med_j$	$z_{ij} = x_{ij}^* w_j + w_j$

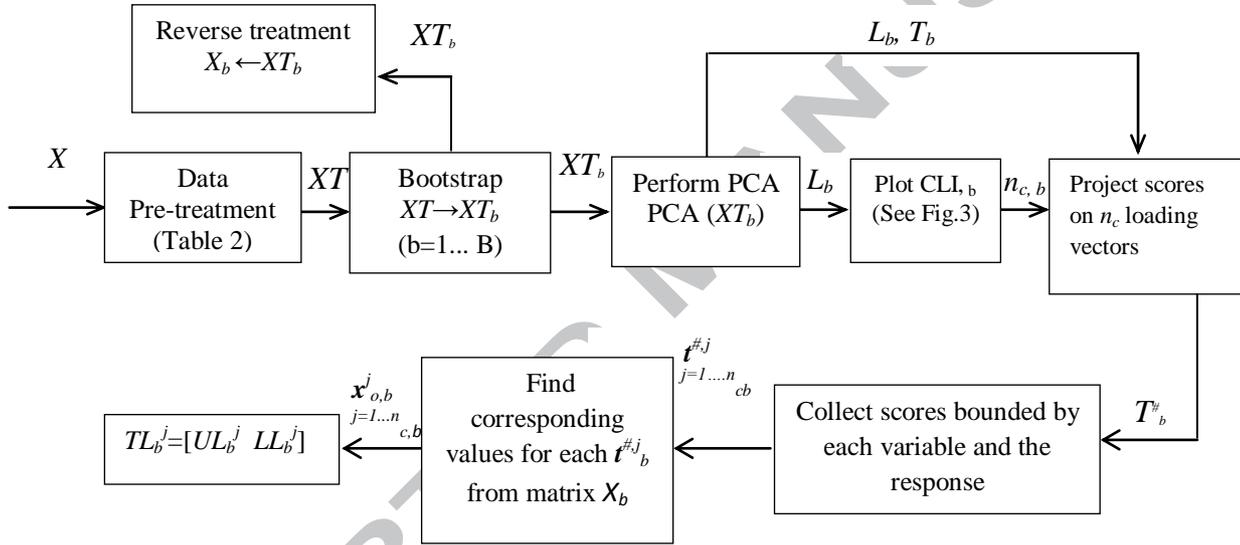


Fig.2 Main steps of the bootstrap uncertainty estimation method for the QCA

2.3 Bootstrap Standard Errors and Confidence Intervals

Bootstrap standard errors are used to measure the uncertainty of bootstrap estimates. The root mean square error is calculated between the bootstrap estimate and empirical estimation of the static of interest θ (e.g. any model parameter such as scores or loading vectors). In the proposed bootstrap method, θ represents (i) the length and cosine of angles of loading vectors corresponding to a factor in a CLI plot, as well as (ii) upper and lower tolerance limits (UL^j, LL^j) corresponding to a factor j .

According to (Efron, & Tibshirani, 1993; Wehrens, Putter, and Buydens, 2000), the bootstrap standard error is:

$$\hat{s}e_B = \left[\frac{\sum_{b=1}^B (\hat{\theta}_b^* - \hat{\theta}^*)^2}{B-1} \right]^{1/2} \dots (1)$$

Where $\hat{\theta}_b^*$: is the b^{th} bootstrap parameter estimate, $\hat{\theta}^*$ is the mean of distributions of $\hat{\theta}^*$ (re-sampled parameter estimations) and B is the total number of bootstrap re-samplings.

Three approaches are normally used to estimate confidence intervals (CI) from the bootstrap samples. The approaches are (i) percentile, (ii) student-t distribution interval (bootstrap-t interval) and, (iii) bias-corrected and accelerated (BCa) (Efron and Tibshirani, 1993; Efron, 1987). BCa is generally considered accurate as it a second order error method as compared with the first order percentile method (Timmerman, Kiers, and Smilde, 2007).

BCa calculates CIs by using modified percentile bootstrap confidence intervals. The BCa method is based on two parameters. The bias correction (\hat{z}_0) parameter corrects the difference between $\hat{\theta}^*$ (the mean of distributions of $\hat{\theta}^*$) and $\hat{\theta}$.

The second parameter is the acceleration constant \hat{a} and is used to correct skewness of the distribution of $\hat{\theta}$. The lower and upper points of the $100(1-2\alpha)\%$ CI can be estimated by $\hat{\theta}^*(\alpha_1)$ and $\hat{\theta}^*(\alpha_2)$, where:

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^\alpha}{1 - \hat{a}(\hat{z}_0 + z^\alpha)}\right) \quad \dots (2-a)$$

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\alpha)})}\right) \quad \dots (2-b)$$

Where:

$\Phi(\cdot)$: is the standard normal cumulative distribution function.

z^α : is the $100\alpha^{\text{th}}$ percentile point of a standard normal distribution.

α : is the significant level, for 95% CI, $\alpha = (1-0.95)/2$.

$\hat{z}_0 = \Phi^{-1}\left(\frac{[\hat{\theta}_b^* < \hat{\theta}]}{B}\right)$, $\Phi^{-1}(\cdot)$: is the inverse of the $\Phi(\cdot)$, and $\hat{\theta}$: is the empirical estimate.

The acceleration or skewness constant \hat{a} estimation based on doing jackknifing re-sampling algorithm on data matrix (see Appendix 1) (Efron and Tibshirani, 1993):

$$\hat{a} = \frac{\sum_{mo=1}^m (\hat{\theta}_{mo} - \hat{\theta})^3}{6 \left[\sum_{mo=1}^m (\hat{\theta}_{mo} - \hat{\theta})^2 \right]^{3/2}} \quad \dots (3)$$

Where $\hat{\theta}_{mo}$ is the estimated parameter from the original data matrix with the mo^{th} row removed and $\hat{\theta}$ is the mean value of the $\hat{\theta}_{mo}$ estimates. The detailed methodology is described below in Table 4:

Table 4: The proposed method for bootstrap uncertainty estimation based on the QCA

<p>Step 1- $X \rightarrow XT$ (From Table 3)</p> <p>Step 2- $XT \rightarrow XT_b$ ($b=1, \dots, B$)</p> <p>(2-1): PCA(XT)</p> <p>(2-2): $XT_{\text{Recon}} = TL^T + E$</p> <p>(2-3): Boot strap E_b from $E \rightarrow XT_b = X_{\text{Recon}} + E_b$</p> <p>Step 3-For $b=1:B$</p> <p>(i) -Reconstruct X_b from XT_b (From Table 4)</p> <p>(ii) -PCA (XT_b)</p> <p>(iii) -Draw CLI:</p> <p>(iii-a): Calculate loading matrix for p principal components from: $L^b = (D^b_s)^{-1} V^b D^b_e$</p> <p>(iii-b): Calculate the length of each loading vector and find the cosine between each factor with response.</p> <p>(iii-c): Draw each loading vector from its length and its angle.</p> <p>(iv) -Discover Tolerance Limits:</p> <p>(iv-a): Project scores on all correlated variables and responses, for score i and variable or response</p> $j: t_i^{b,\#} = \left\ t_i^{b,i} \right\ \frac{L_j^b \cdot t_i^{b,i}}{\left\ L_j^b \right\ \left\ t_i^{b,i} \right\ }$ <p>(iv-b): If variable $j \sim R^b$ Find the projected scores bounded by positive direction of variable j and positive direction of response. Else Find the projected scores bounded by positive direction of variable j and negative direction of response. End If</p> <p>(iv-c): Find the original values from matrix (X_b) corresponding to the scores in the last step for variable j and store it in the vector $x_{o,b}^j$</p> <p>(iv-d): Find tolerance limits: $TL_b^j = [UL_b^j \quad LL_b^j]$ (if j is quantitative variable), $UL_b^j = \max(x_{o,b}^j)$ and $LL_b^j = \min(x_{o,b}^j)$.</p> <p>End For</p> <p>Step 4 - Estimate CI</p> <p>(4-1): Calculate z_0, a.</p> <p>(4-2): Calculate α_1, α_2</p> <p>(4-3): Estimate CI as: $[\hat{\theta}^*(\alpha_1) \quad \hat{\theta}^*(\alpha_2)]$</p>

3. Results from 1000 Bootstrap samples for a Nickel Based Alloy Data Set

The QCA was tested on a Nickel based alloy (Ransing et al., 2016). The data set has 16 factors (chemical compositions) that affect the process response % defective components produced in a batch due to shrinkage defect. The input factors are quantitative variables and the in-process data was available for 60 batches or observations. Interactions among process inputs (various chemical compositions) are considered. The interactions variables combined two continuous factors to produce a categorical interactions factor. The combined matrix of original factors and the interactions variables are considered as matrix X . The optimal tolerance limits determined by the QCA are published in the literature (Ransing et al. 2016).

The number of bootstrap samples (B) used in this simulation is 1000. The QCA depends on two steps: (i) CLI plots and, (ii) tolerance limits estimation. The CIs for CLI parameters and tolerance limit parameters are estimated in order to quantify the uncertainty of QCA. The CLI parameters used in this study are the length and cosine of angles of loading vectors for correlated factors. As an example

the length and the angle for loading vector for factor %Co (%cobalt) is shown in Figure 3. All loading vectors will have unit length if they are plotted in the space defined by all principal components. The co-linearity indices are discovered on few but dominant principal component space and hence, the length of the loading vectors can be less than one. This reduction in the magnitude represents that amount of contribution attributed to the reduced principal component space (Ransing et al. 2013). The angle of the loading vector with respect to the response variable (%Shrinkage Penalty) is a measure of the correlation. The length and the angle are captured by the co-linearity index as a measure of the positive and negative correlation of the factor with respect to the penalty values of the response. The circles around the %Co loading vector show the end points of the loading vector for each bootstrap resample. Thus the uncertainty in the loading vector length and the angle is measured by the spread of corresponding circles. The corresponding distributions of lengths and cosine of angles for %Co factors are shown in Figure 4.

The upper and lower tolerance limits for a given factor j (UL LL) are considered as the tolerance limit parameters. The uncertainty in these parameters is shown in Figures 5 and 6. The bootstrap error, calculated using Equation 1 for few correlated factors, is highlighted in Figs. 7 and 8 for quantitative and categorical factors respectively. The standard error values in the cosine angle are highest and tend to vary significantly for different factors (Figs 7 and 8) however, the high magnitudes of standard errors in cosine angle and the length of factors have not contributed to the standard errors in the upper and lower tolerance limits. This may be because the calculation procedure for the tolerance limits in the QCA (Ransing et al. 2016) is less sensitive to the corresponding changes in angle and length values.

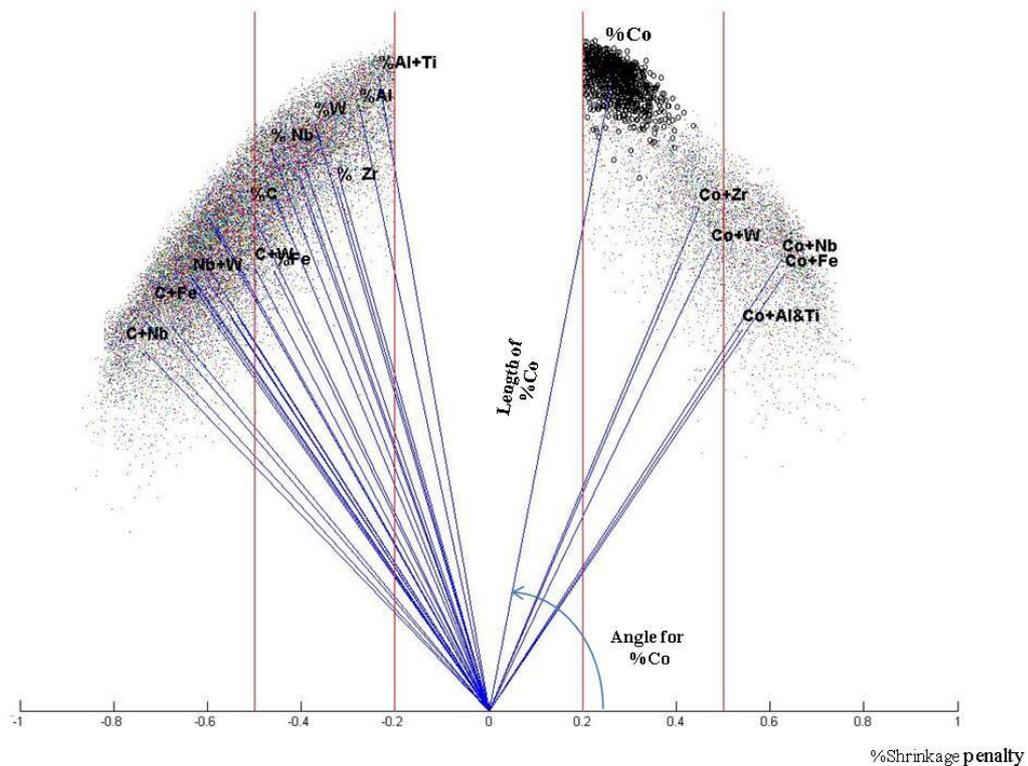


Fig.3. Co-linearity Index (CLI) plot for original and bootstrap resampled data

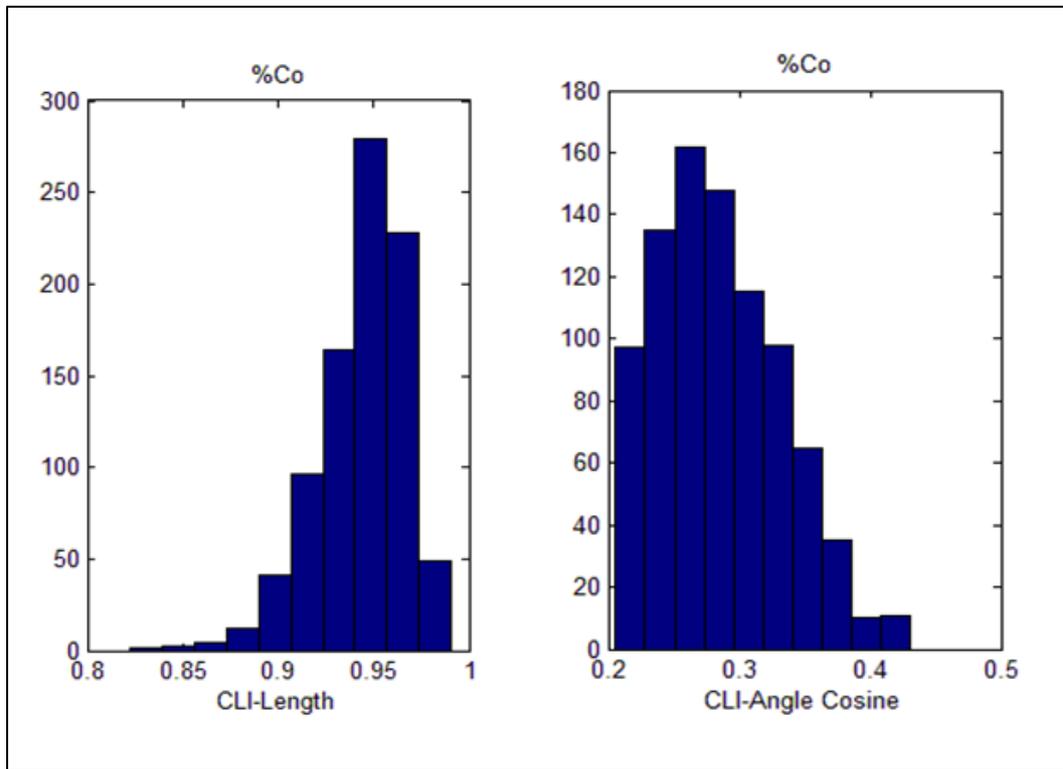


Fig.4. Distributions of CLI length and cosine angle of the factor %Co.

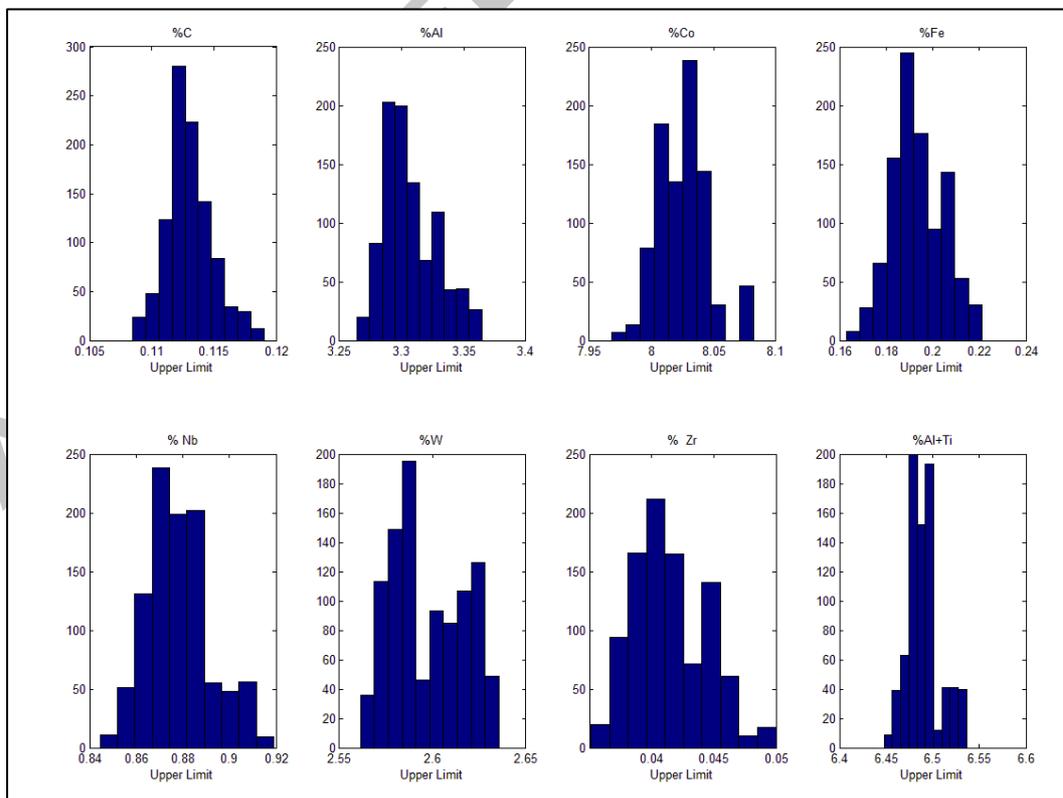


Fig.5. Distributions of upper tolerance limit values

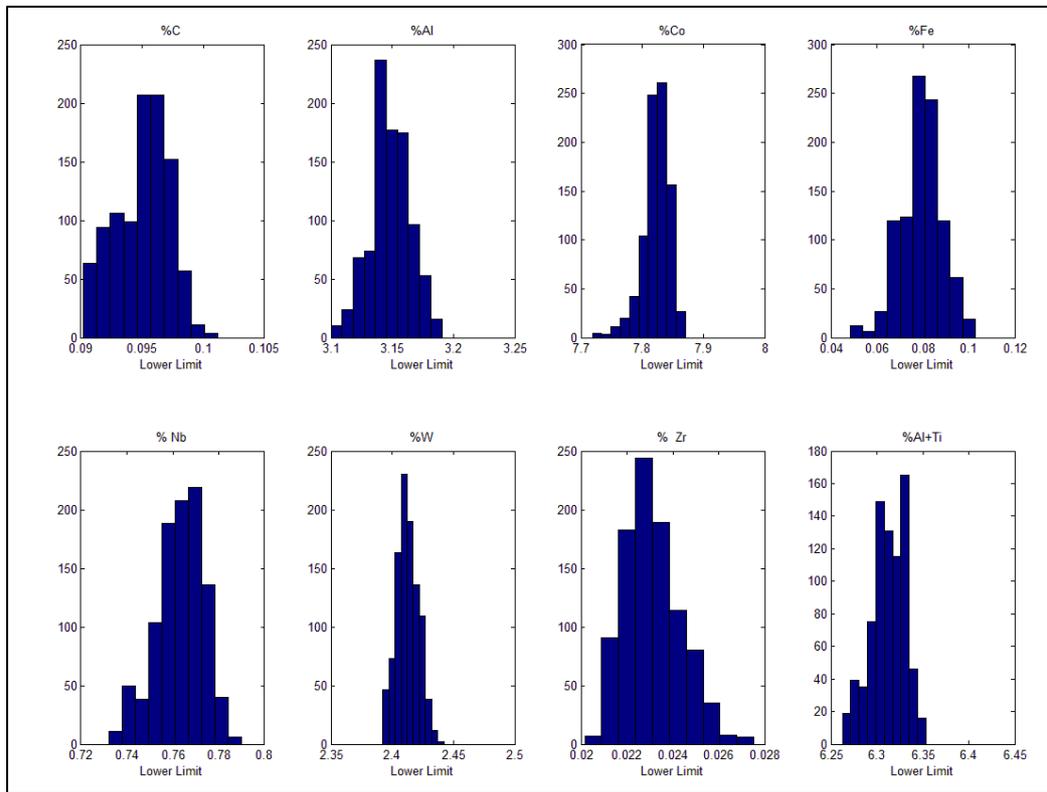
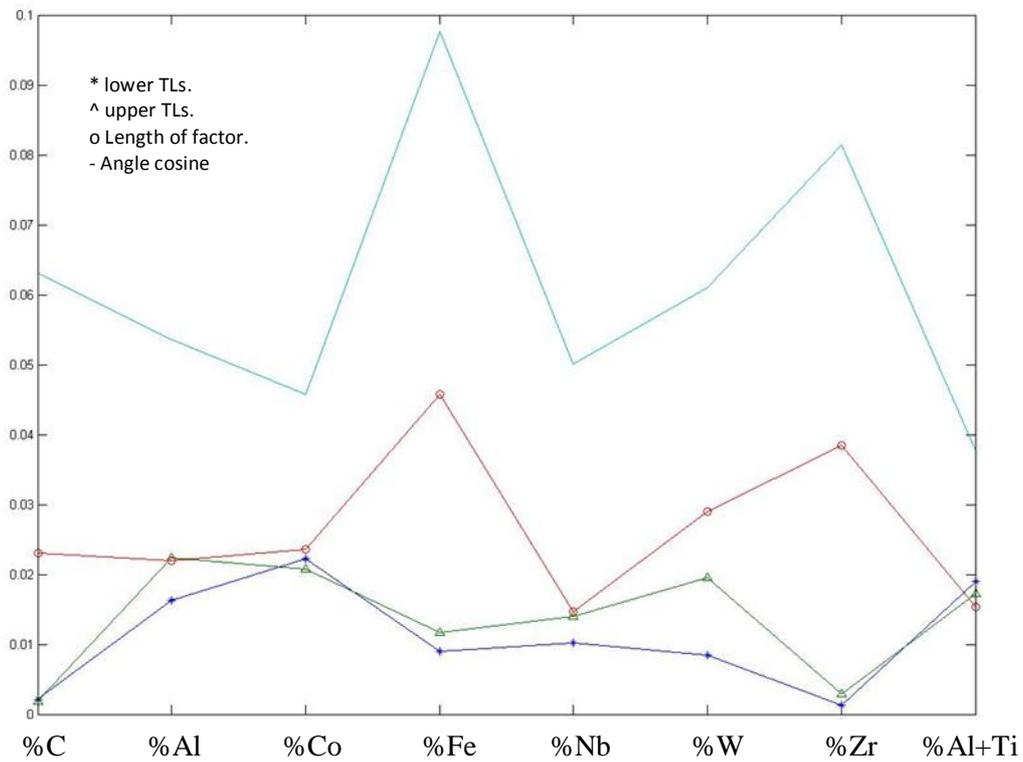


Fig.6. Distributions of lower tolerance limits values



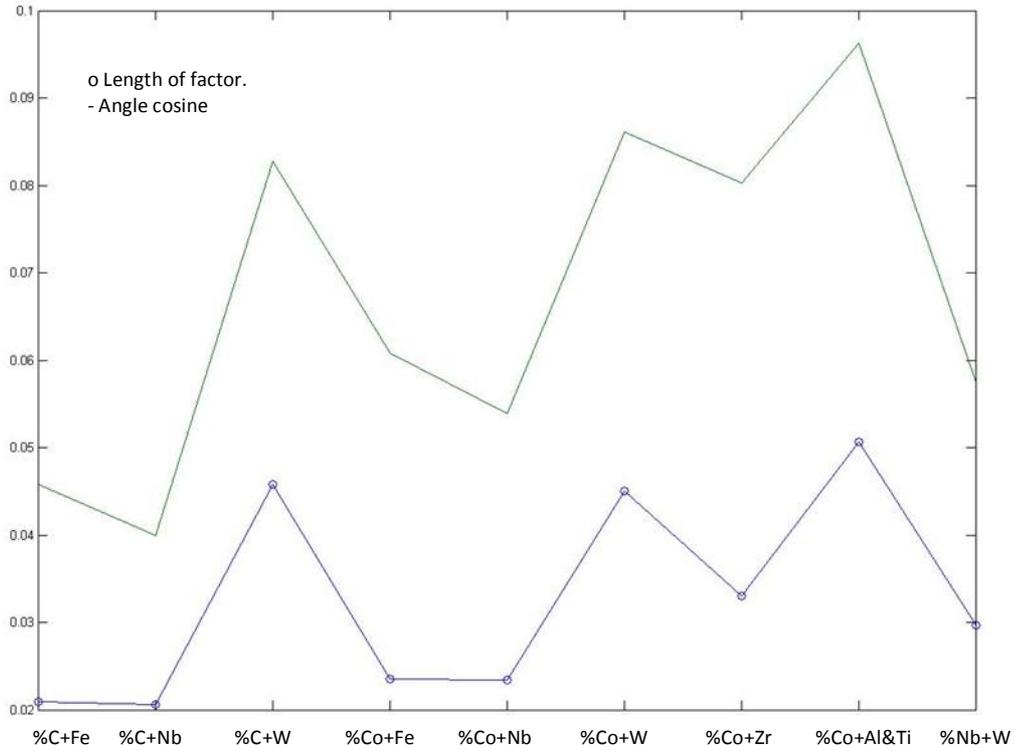


Fig. 8. Bootstrap standard error for categorical factors

3.1. Discussion of results

The CIs are calculated for all factors and system parameters. The distributions of samples were analysed for all system parameters. It was observed that some distributions skewed to the right (e.g. CLI length in Fig.4) or left (CLI cosine angle in Fig.4) and others showed nearly normal distribution (%Fe factor in upper and lower tolerance limits distribution plots in Fig.5 and Fig.6). A few factors demonstrated bimodal distributions (%W and %Al+Ti factors for upper tolerance limits distributions in Fig. 5).

Some distributions clearly identify a most frequent value generated during bootstrap resampling, but other distributions have no clear winning value. In statistics a weighted mean value is normally used to quantify a distribution centre. It is often used to identify the most significant value from a set of similar values (Grela, 2013). The sample frequencies are used as weights to calculate probabilities (Finch, 2009). For example, the weighted mean estimate for a factor j is given by the following equation:

$$Lim^j = \frac{\sum_{i=1}^I f_i^j \times Lim_i^j}{\sum_{i=1}^I f_i^j} \quad \dots(4)$$

Where: Lim^j : weighted mean of tolerance limit for factor j corresponding to either upper limits as shown in Fig.5 or lower limits as in Fig.6, I : the number of all possible values that appear in the distribution for factor j and f_i^j : the frequency of the tolerance limit value i for the factor j .

It is proposed that the uncertainty in the QCA results, as described by frequency distributions shown in Figs 5 and 6, be considered during any risk based tolerance synthesis project. The tolerance limits proposed by the QCA algorithm (Ransing et al. 2016) were compared with the corresponding weighted mean values and the revised tolerance limits were calculated. These tolerance limits are shown on the left or the first bars in Fig 9. The QCA tolerance limits are shown in the middle tolerance limit bar. The original tolerance limits in the in-process data cover the entire rectangle in Fig. 9 as shown on the rightmost bar.

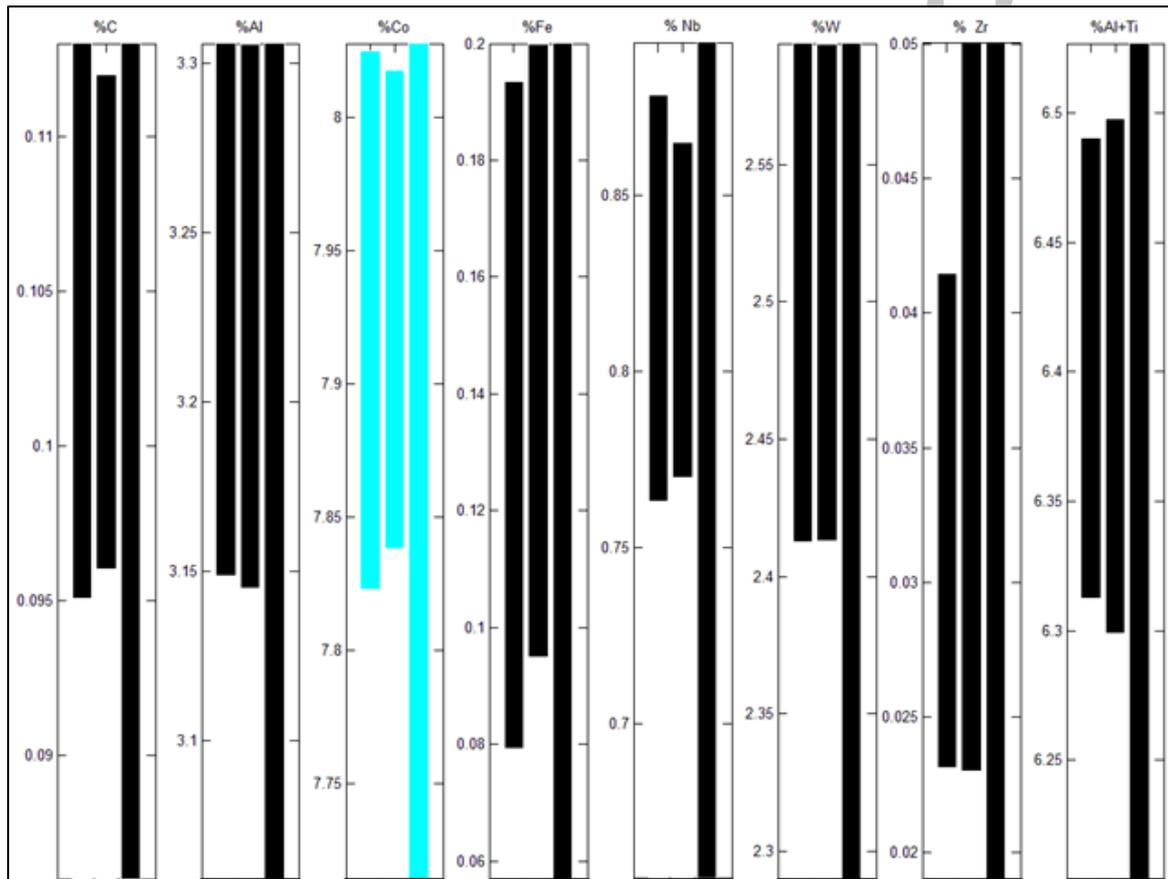


Fig.9. Identification of revised tolerance limits after quantifying uncertainty in the quality correlation algorithm (QCA)

4. Conclusions

A new algorithm to quantify uncertainty in a quality correlation algorithm (QCA) has been proposed. The algorithm is based on bootstrap resampling from the error matrix of a mixed data principal component analysis (PCA) model used in the QCA. A bias corrected and accelerated (BCa) technique is used to estimate confidence intervals (CIs). The traditional approach for quantifying uncertainty in a PCA model generates CIs for score and loading vectors for every principal component. The proposed algorithm has extended this approach to calculate CIs in a p -dimensional principal component subspace used to calculate co-linearity indices (Ransing et al. 2016). The uncertainty in

the tolerance limit for each factor was visualised using frequency distributions and the revised tolerance limits were compared with original limits as well as the QCA limits.

The proposed approach has been illustrated by analysing in-process chemistry data for a nickel based alloy for manufacturing cast components. The proposed uncertainty simulation based on the most repeated estimated values in the resampling distributions can assist process engineers to develop a robust tolerance synthesis approach.

References

Babamoradi, H., van den Berg, F. and Rinnan, Å. (2013). Bootstrap Based Confidence Limits In Principal Component Analysis — A Case Study. *Chemometrics and Intelligent Laboratory Systems*, 120(97-105).

Batbooti, R. S., Ransing R. S. and Ransing, M. (2015). Discover and embed product specific process knowledge in line with ISO9001:2015 requirements, *CIE45 Proceedings*, 28-30 October 2015, Metz/France(1143-1150).

Besse, P. (1992). Pca stability and choice of dimensionality, *Statistics & Probability Letters*, 13(405–410).

Chatterjee, S. (1984). Variance-estimation in factor analysis an application of the bootstrap, *British Journal of Mathematical and Statistical Psychology*, 37 (252–262).

Daudin, J.J., Duby, C. and Trecourt, P. (1988-a). Stability of principal component analysis studied by the bootstrap method, *Statistics*, 19(241–58).

Efron, B. & Tibshirani, R. An Introduction to the Bootstrap, Chapman & Hall, NewYork, 1993.

Efron, B. (1977). Bootstrap methods: another look at the Jackknife, *The Annals of Statistics*, 7 (1-26).

Efron, B. (1987). Better bootstrap confidence-intervals, *Journal of the American Statistical Association*, 82(171–185).

Finch, T. (2009). Incremental calculation of weighted mean and variance. University of Cambridge

Giannetti, C., Ransing, R.S., Ransing, M.R., Bould, D.C., Gethin, D.T., & Sienz, J. (2014). A novel variable selection approach based on co-linearity index to discover optimal process settings by analysing mixed data, *Computers & Industrial Engineering*, 72 (217-229).

Grela G., (2013). Does Weighted Average Really Work?. *Management Knowledge and Learning Proceedings*, 19-21 June 2013, Croatia. (1219-1227).

Hastie, T., Tibshirani, R. and Friedman, J. H. (2009). The Elements Of Statistical Learning. New York: Springer.

Jackson, D. A. (1993). Stopping Rules In Principal Components Analysis: A Comparison Of Heuristical And Statistical Approaches". *Ecology*, 74(2204–2214).

Lambert, Z., Wildt, A. and Durand, R. (1990). Assessing sampling variation relative to number of factors criteria, *Educational and Psychological Measurement*, 50(33–48).

Lambert, Z., Wildt, A. and Durand, R. (1991). Approximating confidence intervals for factor loadings, *Multivariate Behavioral Research*, 26(421–434).

Lewis, R.W., & Ransing, R.S. (1997). A semantically constrained Bayesian network for manufacturing diagnosis, *International Journal of Production Research*, 35 (2171-2188).

Lewis, R.W., & Ransing, R.S. (2000). The optimal design of interfacial heat transfer coefficients via a thermal stress model. *Finite Elements in Analysis and Design*, 34 (193-209).

Lewis, R.W., Ransing, R.S., Pao, W.K.S., Kulasegaram, K., & Bonet, J. (2004). Alternative techniques for casting process simulation. *International Journal of Numerical Methods for Heat & Fluid Flow*, 14(145 – 166).

Pao, W.K.S., Ransing, R.S., Lewis, R.W., & Lin, C. (2004). A medial-axes-based interpolation method for solidification simulation, *Finite Elements in Analysis and Design*, 40 (5-6).

Peres-Neto, P., Jackson, D. and Somers, K. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited, *Computational Statistics and Data Analysis*, 49(974–997).

Postek, E.W., Lewis, R.W., Gethin, D.T., & Ransing R. S. (2005). Influence of initial stresses on the cast behaviour during squeeze forming processes, *Journal of Materials Processing Technology*, 159 (338–346).

Ransing, R., Batbooti, R., Giannetti, C. and Ransing, M. (2016). A quality correlation algorithm for tolerance synthesis in manufacturing operations, *Computers and Industrial Engineering*, 93(1–11).

Ransing, R.S., Giannetti, C., Ransing, M.R., James, M.W. (2013). A coupled penalty matrix approach and principal component based co-linearity index technique to discover product specific foundry process knowledge from in-process data in order to reduce defects., *Computers in Industry*, 64 (514–523).

Raykov, T. and Little, T. (1999). A note on procrustean rotation in exploratory factor analysis: a computer intensive approach to goodness of fit evaluation, *Educational and Psychological Measurement*, 59(47–57).

Smith, B. and Gemperline, P. (2002). Bootstrap methods for assessing the performance of near-infrared pattern classification techniques, *Journal of Chemometrics*, 16(241–246).

Timmerman, M.E. Kiers, H.A.L. and Smilde, A.K. (2007). Estimating confidence intervals for principal component loadings: a comparison between the bootstrap and asymptotic results, *British Journal of Mathematical and Statistical Psychology*, 60(295-314).

Wehrens, R., Putter, H., and Buydens, L. (2000). The bootstrap: a tutorial, *Chemometrics and Intelligent Laboratory Systems*,. 54(35–52).

Ding Y., Jin J., Ceglarek D., and Shi J.. (2005). Process-Oriented Tolerancing for Multi-station Assembly Systems, *IIE Trans.*, 376(493-508).

Li Z., Kokkolaras M., Papalambros P. and Hu, S. J., (2008). Product and process tolerance allocation in multistation compliant assembly using analytical target cascading, *Journal of Mechanical Design*, 130(1-9)

Appendix 1: Leave-one-out jackknife algorithm (Efron and Tibshirani, 1993):

1. Estimate $\hat{\theta}$ from the empirical dataset (X). (By using presented algorithm)
2. Remove object mo ($mo = 1, 2, \dots, m$) from the empirical dataset.
3. Apply data treatment (Table 1) on the reduced data set resulted from (step2).
4. Estimate $\hat{\theta}_{mo}$ from the reduced dataset (with mo th object removed).
5. Go back to step 2 and continue until $mo = m$.
6. Estimate bias for θ : $\hat{bias} = (m-1)(\hat{\theta} - \theta)$, and standard error for θ :

$$\hat{se}_B = \left[\frac{m-1}{m} \sum_{mo=1}^m (\hat{\theta}_{mo} - \hat{\theta})^2 \right]^{1/2}.$$

Where $\hat{\theta}$: is the mean value of all jackknife estimated parameter values.

ACCEPTED MANUSCRIPT

A bootstrap method for uncertainty estimation in quality correlation algorithm for risk based tolerance synthesis

Acknowledgements

The first author would like to thank the financial support from the Ministry of Higher Education and Scientific Research in Iraq, and its representative in the UK, Iraqi cultural attaché in London. He also wants to thank College of Engineering, University of Basra, Basra, Iraq for giving him study leave to pursue PhD research at Swansea University.

ACCEPTED MANUSCRIPT

Highlights

- An approach to embed ISO 9001:2015's risk based thinking for in-process quality improvement is proposed.
- The uncertainty in the quality correlation algorithm has quantified using an enhanced bootstrap method.
- The algorithm determines robust optimal and avoid ranges within the process variation including process interactions.

ACCEPTED MANUSCRIPT