CORE Metadata, citation and similar papers at core.ac.uk
Provided by Cronfa at Swansea University

Swansea University
Prifysgol Abertawe

Cronfa
Setting Research Free

Cronfa - Swansea University Open Access Repository

This is an author produced version of a paper published in:
CIE44 & IMSS'14 Proceedings

Cronfa URL for this paper:
http://cronfa.swan.ac.uk/Record/cronfa43764

Conference contribution :
Giannetti, C. (2014). *Product Specific Process Knowledge Discovery Using Co-Linearity Index and Penalty Functions to Support Process FMEA in the Steel Industry.* CIE44 & IMSS'14 Proceedings, Istanb: CIE44 & IMSS'14.

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder.

Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

http://www.swansea.ac.uk/library/researchsupport/ris-support/

PRODUCT SPECIFIC PROCESS KNOWLEDGE DISCOVERY USING CO-LINEARITY INDEX AND PENALTY FUNCTIONS TO SUPPORT PROCESS FMEA IN THE STEEL INDUSTRY.

C. Giannetti[1*], M. R. Ransing[2], R.S. Ransing[1], D.C. Bould[1], D.T. Gethin[1], J. Sienz[1]

[1]College of Engineering, Swansea University, Swansea SA2 8PP, United Kingdom
c.giannetti@swansea.ac.uk

[2] *p-matrix Ltd., Swansea SA2 7PD, United Kingdom*

ABSTRACT

Process FMEA is a well-established technique for failure analysis widely used to systematically improve manufacturing processes. Despite its widespread adoption process FMEA effectiveness is hindered by the fact that often root causes are not correctly identified. For complex industrial processes, such as casting processes, root cause analysis is challenging because defects are caused by complex interactions between several process inputs. This paper extends previous work and improves the co-linearity index methodology to provide automatic selection of principal components to discover correlations in a reduced dimensional space. It also describes the steps to generate hypotheses about root causes and discover actionable process knowledge that can be used in FMEA studies or broadly as part of process improvement activities. The described 7Epsilon methodology provides a pathway to leverage Product Lifecyle Management in foundries by integrating design and manufacturing knowledge and contributes to the realisation of more sustainable manufacturing processes. The proposed concepts are illustrated using a realistic but anonymised case study from a steel casting foundry.

Keywords: Process Optimisation, Principal Component Analysis, Penalty Matrix Approach, Data Mining, Manufacturing Informatics

1 INTRODUCTION

Quality control and process improvement activities are nowadays standard practices in many organisations. A number of approaches have been developed to support continual process improvement. ISO9001 is the most well-known and adopted Quality Management Standard. Alongside this set of standards foundries also use methodologies such as Six Sigma [1] or 7Epsilon [2] to control process variations and reduce defects. Regardless of the chosen methodology, root cause analysis is at the heart of any process improvement activity. Identification of root causes of defects is the first step to devise preventive and corrective action plans as required by ISO9001.

Process Failure Mode and Effect Analysis (PFMEA) is an engineering technique used to define, identify, and eliminate known and/or potential failures from processes [3]. In addition to processes, FMEA can also be applied to systems, designs and services. FMEA was firstly introduced in 1940s for military usage and subsequently was formalised as a risk assessment tool by NASA in the mid-1960s to satisfy their stringent safety requirements. Nowadays, the use of FMEA has gone beyond the aerospace industry and it is now a very powerful tool widely adopted in the automotive industry, in medical device manufacturing and chemical

---

* Corresponding Author

processing. From 2002, the implementation of FMEA for quality assurance has become a requirement of the ISO/TS 16949 automotive standard, hence it is a widely adopted in foundries who are suppliers of automotive industry.

The FMEA process itself is straightforward, but carrying out an FMEA is far from being a trivial task. An FMEA is a time consuming and tedious activity that requires team effort and extensive domain knowledge [4]. An important step of an FMEA is to identify root causes of failures so that appropriate actions can be taken. Identification of root causes can be performed using a range of approaches some of which rely on team subjective knowledge such as 5 Whys and Cause and Effect Diagrams. Alternatively root cause reasoning can be supported by data analysis techniques that can be used to discover new process knowledge.

Foundry processes are complex manufacturing process with several sub-processes such as patternmaking, molding, coremaking, melting and pouring, heat treatment, welding and finishing. In a foundry situation the effectiveness of FMEA is hindered by the fact that it is difficult to perform root cause analysis because casting defects are often caused by complex interactions of process variables. Furthermore the lack of process knowledge and adequate personnel trained in process control, as described in [2, 5], is another obstacle for the successful implementation of FMEA. As a result of this, when FMEA activities are based on subjective team knowledge, only generic root causes are identified.

In recent years several attempts have been made to discover new process knowledge by analysing patterns in data. In the literature a variety of data mining and statistical techniques have been successfully employed in foundries to discover new process knowledge. These methods include techniques rooted in Statistics such as analysis of variance (ANOVA) [6-8] and methods based on Artificial Intelligence such as Genetic Algorithms and Artificial Neural Networks [9-15]. For foundry processes the influence of design and process parameters has also been quantified via numerical simulation methods and results stored in a database [16-19] or using Bayesian Networks [20]. Typically these methods attempt to model the complex relationships between process inputs and outputs either to characterise or sometimes predict process behaviour and find improvement opportunities. Unfortunately, in foundries, the practical implementation of these methods is hindered by the fact that process knowledge is foundry, product and process specific [21] and data sets are noisy and heterogeneous [22]. Recently the 7Epsilon approach has been proposed to promote discovery, retention and reuse of product specific process knowledge in foundries [2, 23]. In a recent application Principal Component Analysis has been used as part of the co-linearity index methodology to quantify noise free correlations among process factors and responses and discover product specific process knowledge in foundries [21]. In the calculation of the co-linearity index PCA is employed to filter out noise and display correlations in a reduced dimensional space that accounts for most of the variance. The method has also been recently extended to a mixture of categorical and continuous variable by using Multiple Factor Analysis (MFA) [22].

In this paper a novel approach for the discovery of root causes of defects in the foundry industry is presented. It overcomes current limitations of FMEA practices by promoting the discovery of new product specific process knowledge that can support root cause reasoning during FMEA meetings. Hypotheses about root causes are discovered from in-process data and then validated during confirmation trials. The knowledge loop is closed by ensuring that FMEA knowledge is re-used for future purposes.

This paper is structured as follows. A general literature review of FMEA is presented in Section 2 which also describes the motivation of this work and limitations of current FMEA approaches in the foundry industry. In Section 3 the proposed methodology is illustrated through an industrial case study and the paper is concluded in Section 4.

2 FAILURE MODE AND EFFECT ANALYSIS (FMEA) AND RELATED WORK

The main objective of FMEA is to identify potential failure modes of a systems or process, analyse the effects and identify actions to mitigate risks associated with the failure modes. While doing an FMEA risks are prioritised according to defined criteria and actions are taken starting from the higher priority failure modes. The process of conducting an FMEA consists of 10 steps summarized in Figure 1.

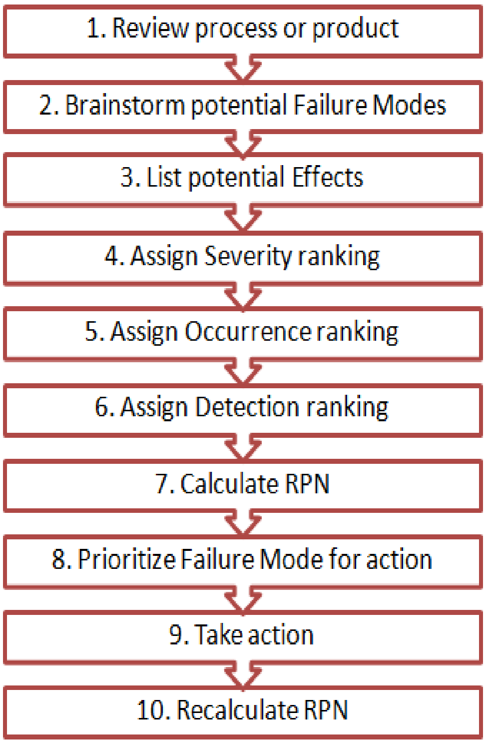

Figure 1: The 10 steps of FMEA process.

The risk priority number (RPN) is a key concept to FMEA analysis. This is calculated as $RPN = S \times O \times D$, where S is the severity of the consequences of the failure, O is the likelihood of occurrence of the causes of failure and D is the likelihood of detection. In general these three factors, usually ranked from 1 to 10, are estimated by experts according to predetermined criteria. The FMEA is a cyclic process and does not terminate when RPNs are calculated, but actions must be taken to reduce the values of RPNs. Once actions have been taken, RPNs are computed again. The FMEA cyclic process is represented in Figure 2.

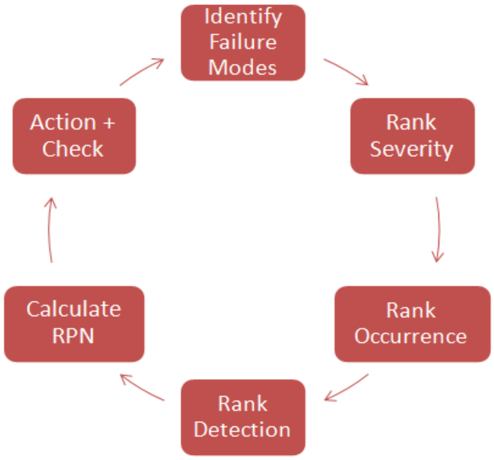

Figure 2: The FMEA process is cyclic loop to support continual process improvement.

Despite the fact that the use of FMEA is very widespread in the industry, a great deal of research has highlighted and addressed various shortcomings. A major one is that the ranking system based on the calculation of the Risk Priority Number (RPN) relies on subjective judgment of the FMEA team members and it is often difficult for a team to agree on the ratings [24]. It is also argued that the mathematical formula for calculating the risk priority number is questionable and debatable [24, 25]. Also the effort to maintain an FMEA is not negligible, since a new analysis needs to be carried out each time changes are made to the design, process or service. The reuse of past FMEAs is made more difficult because the analysis is usually performed using spreadsheets in textual natural language. This poses limitations on computer-based extraction of knowledge for future use.

Researchers have addressed the above issues in different ways and suggested improvements to classical FMEA. The intrinsic difficulties in evaluating and ranking failure modes have been addressed by several studies where researchers have used fuzzy logic to account for the uncertainty in evaluating the risk priority number. For instance, Braglia et al. [24] have developed a technique that uses fuzzy set theory to quantify and rank critical failures. In another paper the difficulty of predicting the occurrence score (O) is overcome by using a framework equipped with a fuzzy inference system based occurrence model [26]. Other researchers have suggested improvements to RPN calculations using a technique that combines Ordered Weighted Average (OWA) and Decision Making Trial and Evaluation Laboratory (DEMATEL) methodologies [27]. An improvement to calculations is also proposed by an approach that uses fuzzy rule base and grey relation theory to rank potential causes which would have identical RPN but different implications [28]. In addition Evidential Reasoning (ER) techniques are also adopted to capture FMEA team members diversity of opinions and prioritise failure modes under different types of uncertainty [29]. In order to take into account the relative importance of risk factors, researchers have suggested the use of Data Envelopment Analysis (DEA) to determine risk priorities of failure modes. In a recent paper the calculation of RPN using fuzzy variables is further refined to integrate the weighted least square method, the method of imprecision and the partial ranking method to improve risk priority calculations and better deal with uncertainty [30]. Finally to address economical aspects related to the FMEA, Hassan et al. [31] have suggested a cost based FMEA aimed at improving quality/cost ratio.

In many manufacturing environments it is important to be able to carry out failure analysis earlier on during the development stage. To perform an FMEA during the conceptual design requires a great deal of expertise and knowledge. Sometimes this knowledge is not readily available within the company. To overcome this limitation Stone et al. [32] have proposed a novel knowledge-base design methodology called Function Failure Design Method that allows designers to perform failure analysis during conceptual phase. A knowledge base of historical failures is used to link failure modes to functionalities and guide the designer to an improved design by predicting failures on the basis of product's functionalities.

The knowledge and information in the form of spreadsheets or tables accumulated while performing an FMEA may become vast and it is often unpractical to try to reuse this information. There is hence the need to develop strategies to reuse the FMEA knowledge so that the FMEA process can be automated. Automatic FMEA generation is particularly useful during conceptual design to overcome the issue of keeping the FMEA knowledge up to date. Researchers have demonstrated that FMEA knowledge can be reused through a knowledge modelling approach assisted by functional reasoning techniques to enable automated FMEA generation from historical data [33]. An alternative approach is to reuse FMEA knowledge for continuous process improvement. For instance, Xiuxu et al. [34] have demonstrated how the knowledge accumulated during FMEAs can be stored and retrieved to create an FMEA knowledge-base repository to assist operators in decision making to improve the reliability and quality of products.

2.1 Motivation of present work

Process FMEA is widely used in foundries to systematically identify ways in which a process can fail and, once root causes have been established, devise corrective action. In a foundry situation typically casting defects are caused by multiple root causes which are the results of complex interactions of different process settings. Because of this, FMEA activities based solely on team member subjective knowledge are often unable to capture multiple root causes. In the case study presented in this paper a steel foundry had the requirement of achieving 0% fractured surface area with conchoidal nature during facture test. However fracture tests were failing on conchoidal fracture. As shown in the scatter diagram in Figure 3, conchoidal fracture shows variability of values across heats.

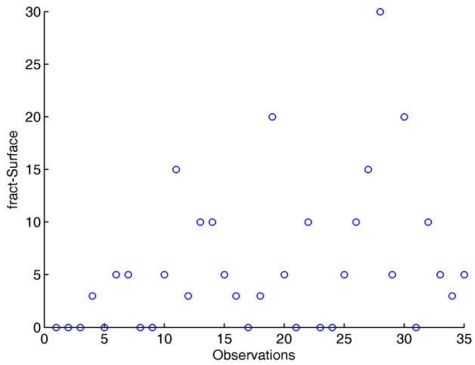

Figure 3: The scatter plot of response, namely conchoidal fracture, shows variability of response values.

An FMEA applied to the melting process was conducted as part of the process improvement activities and it is shown in Figure 4. Improper settings of chemistry parameters were identified as potential failure modes and lack of training in various melting practices was considered the only type of root cause. Although lack of training is one possible root cause, in order to achieve continual process improvement, process engineers need to be able to develop hypotheses about root causes that are specific for a given process and product. In the context of foundries product specific process knowledge is defined as actionable information, in terms of the optimal tolerance limits and target values for continuous factors and optimal levels for discrete factors, in order to achieve desired process response(s) [21]. Process knowledge can be obtained by developing a sound understanding of the relationships between process factors and responses for a specific casting. Knowing which ranges of process factors are associated with desired/undesired response values can help process engineers to develop hypotheses about root causes of defects. Such knowledge can be gained by analysing sometimes weak patterns in noisy in-process data and re-using existing product specific process knowledge.

In the next section a novel approach for dynamic update and re-use of FMEA knowledge during process improvement activities is presented. This approach builds on existing research carried out as part of the 7Epsilon methodology to promote discovery and re-use of product specific process knowledge [2]. The product specific process knowledge discovered by analyzing in-process data is re-used in the context of FMEA to quantify potential causes of defects and find new tolerance limits of process parameters. The new tolerance limits can be adopted to devise preventive and corrective actions specific for a certain process and part. The newly discovered product specific process knowledge is then used to dynamically update FMEA tables stored in a knowledge base.

FMEA Number- MELT-1
Page 2 of 2
Prepared By:

FAILURE MODE AND EFFECTS ANALYSIS
(PROCESS FMEA)

Name of your company | Process Resposiblity: Name: | FMEA Date : Oct10-2013
Core Team :

Process Function / Requirements	Potential Failure Mode	Potential Effect(s) of Failure	SEV	Class	Potential Cause(s) Mechanism(s) of Failure	Occur	Current Process Controls	Detect	RPN	Recommended Action(s)	Responsibility & Target Completion Date	Action Results: Actions Taken	Sev	Occ	Det	RPN
Melting - Chemistry	Incorrect percentage of Cr (X11)	% Conchoidal fracture(Y1)	7	FF	Melter Error - Inadequate Training of Melter in calculation of alloys	7	Periodic Auditing by Supervisor	6	294							
					Incorrect Melt practice											
Melting - Chemistry	Incorrect percentage of Mo (X12)	% Conchoidal fracture(Y1)	7	FF	Melter Error - Inadequate Training of Melter in calculation of alloys	7	Periodic Auditing by Supervisor	6	294							
Melting - Chemistry	Incorrect percentage of Cu (X13)	% Conchoidal fracture(Y1)	7	FF	Inadequate Returns Segregation System	7	Periodic Auditing by Supervisor	7	343							
Melting - Chemistry	Incorrect percentage of Al (X14)	% Conchoidal fracture(Y1)	7	FF	Melter Error - Inadequate Training of Melter in Deoxidation Practice	8	Periodic Auditing by Supervisor	8	448							
Melting - Chemistry	Incorrect percentage of Ti (X15)	% Conchoidal fracture(Y1)	7	FF	Melter Error - Inadequate Training of Melter in Deoxidation Practice	8	Periodic Auditing by Supervisor	8	448							
Melting - Chemistry	Incorrect Ratio of Mn/S (X16)	% Conchoidal fracture(Y1)	7	FF	Melter Error - Inadequate Training of Melter in melting practice	7	Periodic Auditing by Supervisor	6	294							
Melting - Chemistry	Incorrect percentage of Zr (X17)	% Conchoidal fracture(Y1)	7	FF	Melter Error - Inadequate Training of Melter in Deoxidation Practice	8	Periodic Auditing by Supervisor	8	448							
Melting - Chemistry	Incorrect percentage of Ca (X18)	% Conchoidal fracture(Y1)	7	FF	Melter Error - Inadequate Training of Melter in Deoxidation Practice	8	Periodic Auditing by Supervisor	8	448							
Melting - Chemistry	Incorrect Ratio of Ca/Al (X19)	% Conchoidal fracture(Y1)	7	FF	Melter Error - Inadequate Training of Melter in Deoxidation Practice	8	Periodic Auditing by Supervisor	8	448							

Figure 4 - FMEA applied to continual process improvement activities aimed at reduction percentage of conchoidal fractured surface of steel casting. When FMEA is based on subjective knowledge there is the risk that the identified causes are too generic.

3 A SYSTEMATIC APPROACH TO ROOT CAUSE ANALYSIS BASED ON CO-LINEARITY INDEX AND PENALTY MATRICES

In the next sections a novel approach for root cause analysis in the foundry industry is presented. It combines evidence obtained by analysing in-process data with systematic re-use of process knowledge via a knowledge repository and consists of several steps:

- Visualisation of noise free correlation using the co-linearity index methodology
- Product specific process knowledge discovery with penalty matrices
- Hypotheses validation and confirmation trials
- Update and re-use of FMEA knowledge

The steps are illustrated through an industrial process improvement case study aimed at reduction of percentage of conchoidal fractured surface area during melting sub-process in steel casting.

3.1 Mathematical formulation and notation

In steel foundries data are routinely being collected as part of the requirement of ISO9001 standard. A typical dataset contains about 20-40 process variables and a number of observations between 50 and 100. In-process multivariate data can be mathematically represented as a matrix X of dimension $m \times n$ where m is the number of process observations is and n is the total number of process variables, including process inputs and process outputs (also referred as responses).

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \tag{1}$$

The matrix X may contain a mixture of categorical variables such as operator name, shift, and days of the week as well as continuous variables like chemistry concentration, values of temperatures and moisture. Responses are typically occurrences of defects like shrinkage or inclusion as well as material properties such as yield strength, toughness or ductility. An example of mathematical representation of data sets is shown in Figure 5.

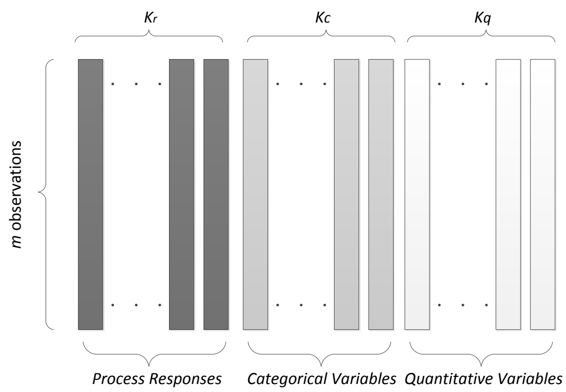

Figure 5: Multivariate in-process data consists of responses, categorical and continuous factors.

Given in-process data where each column has been centred about the mean, the covariance matrix is defined as

$$Cov = \frac{1}{m-1} X^t X. \tag{2}$$

The pair wise correlation of any two variables i and j can be estimated by calculating the sample correlation coefficient defined as

$$r_{i,j} = \frac{Cov(i,j)}{\sigma_i \sigma_j} \tag{3}$$

where σ_i and σ_j are the standard deviations calculated from the set of observations (i.e. columns of the matrix X). The correlation coefficient is a number between zero and one and indicates the degree of linear dependency of the two variables. Geometrically, for centred data, the correlation coefficient is the cosine of the angle formed by the two vectors representing the variables in the m dimensional space of the observations.

3.2 Visualisation of noise free correlations via the co-linearity index

When performing root cause analysis, hypotheses about potential root causes of defects can be generated by examining the correlations between process factors and responses from in-process data. In particular the co-linearity methodology index has been developed to simultaneously visualise, in a simple two dimensional plot, noise free correlations between factors and responses [21, 22]. Although a causal relationship cannot be inferred by just looking at correlation, the co-linearity plots can support root cause reasoning during FMEA meetings. One advantage of the co-linearity index methodology is the ability to perform simultaneous analysis of categorical and continuous data as shown in [22]. In order to mix categorical and continuous data, pre-processing data transformations are applied as explained in detail in [22]. Also responses are transformed using a penalty matrix approach that penalises deviation from desired responses.

3.2.1 Response transformation using penalty functions

Responses are scaled using penalty functions which penalise deviation from desired responses [21]. If lower values of response correspond to a desirable outcome, a penalty value of 1 is given to response values above a certain threshold T_{max} and penalty value 0 to response values below a certain threshold T_{min}. Vice versa applies if higher values correspond to desirable outcomes. The full set of transformations is given in Table 1. Thresholds can be chosen based on subjective knowledge or some heuristic rules have been suggested [22]. The use of penalty functions help to highlight patterns related to desirable/undesirable responses.

Table 1: Penalty functions scaling for responses

Response	Transformation
Lower the better	$x_{ij} = 0$, if $x_{ij} \leq T_{min}$ $x_{ij} = 1$, if $x_{ij} \geq T_{max}$ $x_{ij} = \frac{x_{ij} - T_{min}}{T_{max} - T_{min}}$ otherwise.
Higher the better	$x_{ij} = 0$, if $x_{ij} \geq T_{max}$ $x_{ij} = 1$, if $x_{ij} \leq T_{min}$ $x_{ij} = \frac{x_{ij} - T_{max}}{T_{min} - T_{max}}$ otherwise.

In the current example, an upper threshold of 10% and lower threshold of 0% were chosen that correspond respectively to undesired and desired values of responses.

3.2.2 Categorical and continuous data transformations

If mixed data are present, a set of data transformations is applied to both categorical and continuous factors so that they can be analysed simultaneously. Detailed description of these transformations can be found in [22]. For continuous variables a data transformation based on median and quartile has been proved to be more effective compared to standardisation using mean and standard deviation. Categorical variables are transformed so that PCA can be carried out simultaneously with continuous variables. In the case study presented in this paper it was decided to include only continuous variables but the proposed approach can be applied to the general case of heterogeneous data containing a mixture of categorical and continuous variables.

3.2.3 Balancing group of variables

The final step of data pre-treatment is to balance the influence of different group variables (responses, categorical and continuous variables). This is achieved by performing a separate PCA analysis of the groups and then dividing data columns of each set by the respective square root of their maximum eigenvalue [22].

3.2.4 Calculation of the co-linearity index

Let us denote with X_T the transformed and balanced in-process data matrix whose columns have been centered about the mean. The co-linearity index is an approximation of the correlation between two variables calculated in a reduced dimensional space obtained by means of PCA. The co-linearity index is calculated by applying PCA to the covariance matrix. In order to reduce noise PCA seeks to find a new orthogonal basis (i.e. principal axis) where it is possible to filter out the noise [35]. The new basis is chosen so that the first axis maximises the variance of the projections of the observations to the axis, which is equivalent to finding a unit vector $v^1 = (v_{1,} \cdots \cdots, v_n)$ that maximises the quadratic form $l(v) = v X^T X v^T$ [21]. By using simple linear algebra properties, it can be shown that the eigenvector of $X^T X$ corresponding to the maximum eigenvalue is the direction that maximises $l(v)$. The second axis is then given by the eigenvector corresponding to the second highest eigenvalue and so on. PCA finds the eigenvectors of Cov which are arranged in a matrix denoted as V containing the eigenvectors as column vectors ordered by greatest eigenvalues. Let us denote with L the loading matrix in full dimension as:

$$L = D_s^{-1} V D_e \tag{4}$$

where D_s is the diagonal matrix containing the standard deviations of the columns of X_T and D_e is the diagonal matrix containing the square roots of eigenvalues. The multiplication by the inverse of D_s is due to the fact that the global analysis is performed on the covariance matrix rather than the correlation matrix. Using simple linear algebra properties it can be shown that LL^t is the correlation matrix of the dataset. In the lower dimensional space spanned by the first p principal components the loadings are calculated by removing the last $n-p$ coordinates of each row vector of L. The resulting $n \times p$ matrix, denoted with L_p, is the loading matrix in the reduced space. The cosine of the angle between the reduced loadings (rows of L_p) is an approximation of the correlation. This is calculated from the inner product of the reduced loadings as:

$$cosine(i,j) = \frac{<l_i, l_i>}{|l_i||l_j|} \tag{5}$$

For each variable j, the co-linearity index with respect to variable i can be plotted in a 2D space by drawing a bi-dimensional vector with the same magnitude as the reduced loading vector of j and angle equal to the cosine of the angle between the reduced loadings of i and j. By means of co-linearity plots process engineers are able to view correlations between process variables and response in a lower dimensional space where the noise has been filtered out by means of PCA.

3.2.5 How many principal components?

The choice of the right number of principal components is critical to the successful implementation of the co-linearity index. If too many components are retained, noise is included causing a distortion of the correlations, while if too few variables are chosen important information may be lost. A great deal of literature has discussed the choice of principal components and several methods have been suggested [36-39]. These include "ad hoc" rules such as Kaiser's and Joliffe's rules suggesting to retain components whose eigenvalues are greater or equal than a given threshold (1 for Kaiser's rule and 0.7 for Joliffe's rule) and methods that use a more rigorous statistical approach, either based on resampling methods such as bootstrap and cross validation or on distributional assumptions. Although "ad hoc" methods seem to work well in most cases, they do not have any rigorous mathematical justification. Usually the analyst will decide which rules are more appropriated based on domain knowledge and the structure of the dataset. Stopping rules based on statistical foundations are often computationally expensive and more difficult to implement. Furthermore it is argued that the performance of different stopping rules may also be dependent on variable correlations and the number of observations and variables [37].

When applying the co-linearity index to mixed and balanced data it has been suggested to use the rule of thumb of including all the components whose respective eigenvalues are greater or equal to 0.7 [22]. This rule was appropriate for data sets analysed in [22] but, for the dataset used in this case study, it seems to underestimate the number of principal components. In fact the suggested rule would only retain one component to account for 48% of the variance. Following a simulation study it was found that the best criteria for the choice of principal components, in the case of mixed data, is to take as a cut off the average eigenvalue $\bar{\lambda}$. According to this rule, a component is retained if its corresponding eigenvalue is greater or equal to $\bar{\lambda}$. For the data set discussed in this paper this equates to retain three components to account for 75.7% of the variance. Further investigation is needed to find out whether rules based on bootstrap or cross-validation can provide more general purpose criteria to be applied to foundry in-process data.

3.2.6 Co-linearity index plots

An example of co-linearity index plot is displayed in Figure 6. It refers to data collected as part of process improving activities aimed at reducing the incidence of conchoidal fractured surface are in a steel casting process. Five regions are identified according to strength of correlation:

- No correlation: one central region with co-linearity index between -0.2 and 0.2
- Weak positive correlation: intermediate region between 0.2 and 0.5
- Weak negative correlation: intermediate region between -0.5 and -0.2
- Strong positive correlation: extreme region between 0.5 and 1
- Strong negative correlation: extreme region between -1 and -0.5

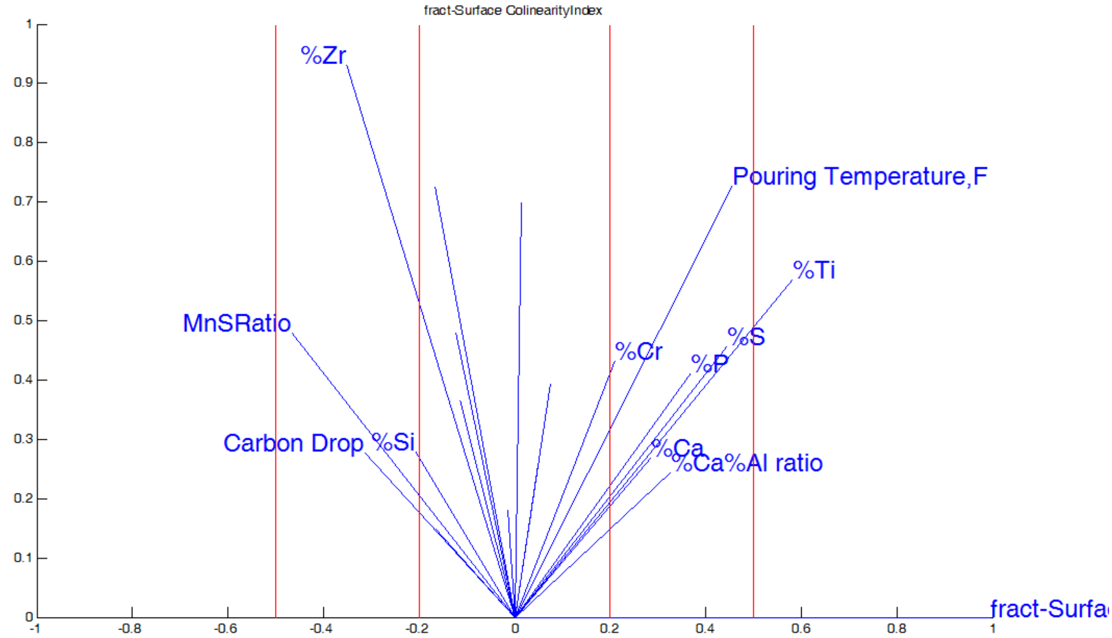

Figure 6: The co-linearity plot is used during FMEA meetings to investigate noise free correlation between process factors and penalty values of response (conchoidal fracture).

From the plot it can be noted that %Ti is identified as having strong positive correlation with conchoidal fracture penalty values. An hypothesis can then be formulated of the existence of a potential causal relationship between high values of %Ti and high values of conchoidal fracture. The co-linearity plots can be examined during root cause analysis meetings to identify potential roots causes. In the context of foundry processes, the latter are measurable factors being in a range associated with undesired responses. In order to associate ranges of factors with desired/undesired responses penalty matrices are subsequently used.

3.3 Product Specific process knowledge discovery with penalty matrices

Penalty matrices are a way to visually discover ranges of measurable factors that are associated with desired/undesired responses. The penalty matrix algorithm described in [21] bins data according to quartile of factor values and levels of responses. Responses are penalised using the approach described in Section 3.2.1. When performing root cause analysis, penalty matrices can be used to associate ranges of factors to desired/undesired responses. As shown in Figure 7, observations corresponding to Bottom 50% of Ti (Q1 and Q2) are associated with low values of penalty functions (desired response) while Top 50% Ti (Q3 and Q4) are associated with high values of penalty function (undesired response). In this case evidence from data suggests that a possible root cause for high conchoidal fracture is %Ti being in the Top 50% range, namely $0.011 < \%Ti \leq 0.016$.

Q1	Q2		Q3	Q4
Minimum		Median		Maximum
0.0009	0.0075	0.011	0.0135	0.016

Q1 & Q2: Optimal; Range: Bottom 50%, [>=0.0009 & <=0.011];
Q1: Optimal; Range: Bottom 25%, {>=0.0009 & <=0.0075};
Q3 & Q4: Avoid; Range: Top 50%, [>0.011 & <=0.016];

Penalty	Q1	Q2	Q3	Q4
0.8-1.0	1		4	5
0.6-0.8				
0.4-0.6		4	2	3
0.2-0.4	2	2		1
0-0.2	6	3		2

Figure 7: ***The Penalty Matrix of Conchoidal Fracture for %Ti shows regions of desired and undesired response.***

3.4 Hypotheses validation and confirmation trials

Correlations discovered with the co-linearity index and penalty matrices are not sufficient to infer causal relationships between ranges of factors and desired/undesired responses. Findings of the analysis need to be discussed during process improvement meetings so that hypotheses can be validated by re-using domain knowledge. Optimal and avoid ranges discovered by analysing patterns in data are compared with trends found during past process improvement activities and available literature review accessed via a knowledge repository. Causation is then inferred if the results of the analysis are supported by the knowledge base, otherwise it is suggested that correlations should be dropped. Following a brainstorming session a few hypotheses are selected to be verified during confirmation trials.

3.5 Update and re-use of FMEA knowledge

Upon successful completion of confirmation trials the newly discovered root causes can be used to update the FMEA spreadsheet. Potential root causes are ranges of factors that are associated with undesired responses. A corrective action can be devised by changing the specification range of this factor to being outside the range associated with undesired response values. An updated FMEA table is displayed in Figure 8. The new root cause and corrective action have been included in the table.

Name of your company — Process Resposiblity: Name: — FMEA Date : Oct10-2013

Core Team :

Process / Requirements	Potential Failure Mode	Potential Effect(s) of Failure	SEV	Class	Potential Cause(s) Mechanism(s) of Failure	Occur	Current Process Controls	Detect	RPN	Recommended Action(s)	Responsibility & Target Completion Date	Action Results: Actions Taken	Sev	Occ	Det	RPN
Melting - Chemistry	Incorrect percentage of Ti (X15)	% Conchoidal fracture(Y1)	7	FF	Melter Error - Inadequate Training of Melter in Deoxidation Practice	8	Periodic Auditing by Supervisor	8	448	Train operator						
Melting - Chemistry	Incorrect percentage of Ti (X15)	% Conchoidal fracture(Y1)	7	FF	High values of %Ti **(from 0.011 to 0.016)**	10	Periodic Auditing by Supervisor	8	560	Keep %Ti in Optimal Range **0.0009 to 0.011**						

Figure 8: Improved FMEA table: %Ti in Top 50% range has been identified as a root cause for high incidence of conchoidal fracture. A new tolerance limit is suggested to prevent the occurrence of the failure.

It is recommended that the new knowledge embedded in the FMEA table is stored in a knowledge repository for future re-use so that the knowledge loop is closed and existing FMEA knowledge can be used to continually improve the process. As part of the 7Epsilon initiative a knowledge repository for re-use of product specific process knowledge is currently under development [2].

4 CONCLUSION

In this paper a systematic approach to discover root causes of defects during process FMEA in the foundry industry has been illustrated. This work builds on previous research focused at developing data driven methodologies to discover product specific process knowledge in foundries. It overcomes limitations of current FMEA approaches in foundries because it facilitates the discovery of product and process specific root causes that are quantifiable in terms of measurable factors being in an unsuitable range. This approach enables quality engineers to find improvement opportunities and refine current tolerance limits of process factors to achieve continual process improvement.

5 ACKNOWLEDGEMENT

The first author would like to thank the financial support from the ASTUTE project (http://www.astutewales.com/) that is funded via the European Regional Development Fund through the Welsh Government.

6 REFERENCES

[1] **Raisinghani, M.S., Ette, H., Pierce, R., Cannon, G., Daripaly, P.** 2005. Six Sigma: concepts, tools, and applications, *Industrial Management & Data Systems*, 105 (4), pp. 491-505.

[2] **Roshan, H.M., Giannetti, C., Ransing, M.R., Ransing, R.S.** 2014. "If only my foundry knew what it knows …": A 7Epsilon perspective on root cause analysis and corrective action plans for ISO9001:2008, Proceedings of the 71st World Foundry Congress, Bilbao, Spain 19th - 21st May 2014.

[3] **Stamatis, D.H.** 2003. *Failure mode and effect analysis: FMEA from theory to execution*, Asq Press.

[4] **Price, C., Hunt, J., Lee, M., Ormsby, A.** 1992. A model-based approach to the automation of failure mode effects analysis for design, *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 206 (4), pp. 285-291.

[5] **Roshan, H.M.** 2013. Start gaining process knowledge, *Modern Casting May 2013, American Foundry Society*, pp. 24-29.

[6] **Cvijovic, Z., Radenkovic, G., Maksimovic, V., Dimcic, B.** 2005. Application of ANOVA Method to Precipitation Behaviour Studies, *Materials Science and Engineering*, 297(1), pp. (195-203).

[7] **Sun, Z., Hu, H., Chen, X.** 2008. Numerical optimization of gating system parameters for a magnesium alloy casting with multiple performance characteristics, *Journal of Materials Processing Technology*, 199 (1), pp. 256-264.

[8] **Senthilkumar, B., Ponnambalam, S., Jawahar, N.** 2009. Process factor optimization for controlling pull-down defects in iron castings, *Journal of Materials Processing Technology*, 209 (1), pp. 554-560.

[9] **Verlinden, B., Duflou, J., Collin, P., Cattrysse, D.** 2008. Cost estimation for sheet metal parts using multiple regression and artificial neural networks: A case study, *International Journal of Production Economics*, 111 (2), pp. 484-492.

[10] **Mohanty, I., Bhattacharjee, D., Datta, S.** 2011. Designing cold rolled IF steel sheets with optimized tensile properties using ANN and GA, *Computational Materials Science*, 50 (8), pp. 2331-2337.

[11] **Zárate, L.E., Dias, S.M.** 2009. Qualitative behavior rules for the cold rolling process extracted from trained ANN via the FCANN method, *Engineering Applications of Artificial Intelligence*, 22 (4), pp. 718-731.

[12] **Shafyei, A., Anijdan, S., Bahrami, A.** 2006. Prediction of porosity percent in Al-Si casting alloys using ANN, *Materials Science and Engineering: A*, 431 (1), pp. 206-210.

[13] **Shabani, M.O., Mazahery, A.** 2011. The ANN application in FEM modeling of mechanical properties of Al–Si alloy, *Applied Mathematical Modelling*, 35 (12), pp. 5707-5713.

[14] **Esfahani, M.B., Toroghinejad, M.R., Abbasi, S.** 2009. Artificial neural network modeling the tensile strength of hot strip mill products, *ISIJ international*, 49 (10), pp. 1583-1587.

[15] **Ransing, R.S., Lewis, R.W.** 1997. A semantically constrained neural network for manufacturing diagnosis, *International Journal of Production Research*, 35 (9), pp. 2639-2660.

[16] **Lewis, R.W., Ransing, R.S., Pao, W.K.S., al., e.** 2004. Alternative techniques for casting process simulation *International Journal of Numerical Methods for Heat & Fluid Flow*, 14 (2), pp. 145-166.

[17] **Postek, E.W., Lewis, R.W., Gethin, D.T., al., e.** 2005. Influence of initial stresses on the cast behaviour during squeeze forming processes, *JOURNAL OF MATERIALS PROCESSING TECHNOLOGY* 159 (3), pp. 338-346.

[18] **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 (2), pp. 193-209.

[19] **Pao, W.K.S., Ransing, R.S., Lewis, R.W., al., e.** 2004. A medial-axes-based interpolation method for solidification simulation, *Finite Elements in Analysis and Design*, 40 (5-6), pp. 577-593.

[20] **Lewis, R.W., Ransing, R.S.** 1997. A semantically constrained Bayesian network for manufacturing diagnosis, *International Journal of Production Research*, 35 (8), pp. 2171-2178.

[21] **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 (5), pp. 514-523.

[22] **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), pp.

[23] **Giannetti, C., Ransing, R.S., Ransing, M.R., Bould, D.C., Gethin, D.T., Sienz, J.** 2014. Knowledge management and knowledge discovery for process improvement and sustainable manufacturing: a foundry case study, Proceedings of the Sustainable Design and Manufacturing 2014, Cardiff, UK, 28th-30th April 2014, pp. 537-548.

[24] **Braglia, M., Frosolini, M., Montanari, R.** 2003. Fuzzy criticality assessment model for failure modes and effects analysis, *International Journal of Quality & Reliability Management*, 20 (4), pp. 503-524.

[25] **Wang, Y.-M., Chin, K.-S., Poon, G.K.K., Yang, J.-B.** 2009. Risk evaluation in failure mode and effects analysis using fuzzy weighted geometric mean, *Expert Systems with Applications*, 36 (2), pp. 1195-1207.

[26] **Tay, K.M., Teh, C.S., Bong, D.** 2008. Development of a Fuzzy-logic-based Occurrence Updating model for Process FMEA, Proceedings of the Computer and Communication Engineering, 2008., Kuala Lumpur, Malaysia May 13-15, pp. 796-800.

[27] **Chang, K.-H., Cheng, C.-H.** 2011. Evaluating the risk of failure using the fuzzy OWA and DEMATEL method, *Journal of Intelligent Manufacturing*, 22 (2), pp. 113-129.

[28] **Pillay, A., Wang, J.** 2003. Modified failure mode and effects analysis using approximate reasoning, *Reliability Engineering & System Safety*, 79 (1), pp. 69-85.

[29] **Chin, K.-S., Wang, Y.-M., Ka Kwai Poon, G., Yang, J.-B.** 2009. Failure mode and effects analysis using a group-based evidential reasoning approach, *Computers & Operations Research*, 36 (6), pp. 1768-1779.

[30] **Zhang, Z., Chu, X.** 2011. Risk prioritization in failure mode and effects analysis under uncertainty, *Expert Systems with Applications*, 38 (1), pp. 206-214.

[31] **Hassan, A., Siadat, A., Dantan, J.-Y., Martin, P.** 2010. Conceptual process planning-an improvement approach using QFD, FMEA, and ABC methods, *Robotics and Computer-Integrated Manufacturing*, 26 (4), pp. 392-401.

[32] **Stone, R.B., Tumer, I.Y., Stock, M.E.** 2005. Linking product functionality to historic failures to improve failure analysis in design, *Research in Engineering Design*, 16 (1-2), pp. 96-108.

[33] **Teoh, P., Case, K.** 2004. Failure modes and effects analysis through knowledge modelling, *Journal of Materials Processing Technology*, 153 pp. 253-260.

[34] **Xiuxu, Z., Yu-bao, M., Rui, C., Xiao-li, B., Lin-yan, N.** 2009. Research and Application of Intelligent Quality Control System Based on FMEA Repository, Proceedings of the Information Technology and Computer Science, 2009. ITCS 2009, pp. 514-517.

[35] **Shlens, J.** 2005. *A Tutorial on Principal Component Analysis*, Accessed June 2014 http://www.brainmapping.org/NITP/PNA/Readings/pca.pdf.

[36] **Joliffe, I.T.** 2002. *Principal Component Analysis*, Springer.

[37] **Peres-Neto, P.R., Jackson, D.A., Somers, K.M.** 2005. How many principal components? stopping rules for determining the number of non-trivial axes revisited, *Computational Statistics & Data Analysis*, 49 (4), pp. 974-997.

[38] **Eastment, H.T., Krzanowski, W.J.** 1982. Cross-Validatory Choice of the Number of Components from a Principal Component Analysis, *Technometrics*, 24 (1), pp. 73-77.

[39] **Josse, J., Husson, F.** 2012. Selecting the number of components in principal component analysis using cross-validation approximations, *Computational Statistics & Data Analysis*, 56 (6), pp. 1869-1879.