A Framework for Improving Process Robustness with Quantification of Uncertainties in Industry 4.0

Cinzia Giannetti
College of Engineering
Swansea University
Bay Campus, Swansea, SA1 8EN
Email: c.giannetti@swansea.ac.uk

***Abstract*—Digitalisation of industrial processes, also called the fourth industrial revolution, is leading to availability of large volume of data containing measurements of many process variables. This offers new opportunities to gain deeper insights on process variability and its effects on quality and performance. Manufacturing facilities already use data driven approaches to study process variability and find improvement opportunities through methodologies such as Design of Experiment (DOE) and Six Sigma. However, current approaches are not adequate to model the complexity of modern manufacturing systems, especially when these systems exhibit non-linear interactions between high numbers of variables. In this paper a methodology to improve process robustness is proposed. This methodology uses non-parametric estimation of quantiles of response to discover new tolerance limits of factors. This method does not make any stringent assumption of linearity and works well in finding the interactions effects of covariates on response quantiles. Process robustness, which is defined as the ability of a process to have acceptable quality whilst tolerating variability of the input, is measured through calculation of Likelihood Ratios (LR) associated to the new tolerance limits. Uncertainty of this estimation is quantified via simulations using the bootstrapping method. The novel contribution of this paper is the application of quantile regression and likelihood ratios to the tolerance synthesis problem applied to a low alloy foundry. It shows the validity of the methodology in modelling behaviours of complex manufacturing processes using data driven approaches to gain new insights on causes of process variabilities and discover new product specific process knowledge. This work contributes to bridging the gap between theory and application towards implementing Industry 4.0 predictive analytics.**

I. Introduction

The new generation of manufacturing systems, also called Industry 4.0, will generate additional streams of in-process data with the potential of transforming the way products are designed, manufactured and serviced [1]. The availability of large data sets provides new opportunities to enhance quality improvement activities through better monitoring and control of processes and real-time adjustments based on the analysis of continuous data streams [2], [3], [4]. ISO9001:2015 standard defines continual improvement as a "recurring activity to enhance performance", and the one that generally leads to a corrective or preventive action [5, p. 16]. Continual improvement activities typically focus on reducing variation in production processes to satisfy customer requirements. For multi-stage manufacturing processes, which consist of several sub-process, keeping variability of process inputs within customers tolerance limits may not be enough to achieve quality objectives. Typically these processes may exhibit variability from unknown sources, which only exist for a specific process and at a given time. This may lead to situations where, despite the process being within the agreed tolerance limits of its input factors, large variations in the process output are still present, leading to sub-optimal operations.

9978-1-5090-5795-5/17/$31.00 2017 IEEE

In the literature process robustness is defined as "the ability of a process to demonstrate acceptable quality and performance while tolerating variability in inputs" [6, p. 2]. According to Striker and Lanza [7], a robust production system needs to deal with disturbances to keep a high level of performance. Giannetti et al. [2] argue that processes robustness can be achieved by making the necessary adjustments to the tolerance limits to enhance process performance. In multi-stages manufacturing achieving process robustness is a challenging activity because the quality of the final product is often influenced by hundreds of factors as well as part specific quality constraints [8], [9], [10], [11]. Recently a novel methodology, called 7Epsilon [12], has been developed which promotes the use of risk based analysis of in-process data to create new product specific process knowledge to achieve robustness of manufacturing processes [8], [10], [13], [2]. As part of the 7Epsilon methodology several data-driven techniques to solve the tolerance synthesis problem have been proposed. This is the study of variability in all process inputs (including interactions among process inputs) in order to discover optimal regions that correlate with the occurrences of expected process outputs (results) [13]. In [13] new tolerance limits are found by visualising data in a reduced dimensional space which accounts for most of the variance. Another approach to discover optimal tolerance limits and evaluate uncertainty of results is proposed by [8]. Process robustness is achieved by selecting optimal tolerance limits of process variables that reduce variation of responses by solving the tolerance synthesis problem [13], [2].

In this paper the methodology developed in [2] is presented and it is applied to a real case study from a low alloy foundry to demonstrate its effectiveness as a general framework for improving robustness of manufacturing processes. The novel contribution of this paper is the application of quantile regression and estimation of likelihood ratios to discover root causes of defects due to interactions of factors.The paper is structured as follows. Section II describes data mining approaches for process improvement found in the literature. Section III describes the methodology and its application to process robustness, while Section IV presents the results on a real case study. The paper is concluded in Section V.

II. Data Mining for Process Improvement

Data mining is the process of discovering new knowledge from data by identifying pattens and trends. For many years data mining methods have found widespread application to support the development of better products across many industrial sectors, including manufacture of fabricated metal products, computers and electronic goods [14]. In a data mining scenario, new knowledge can be created by aggregating and analysing data stored in heterogeneous databases using mathematical methods and visualisation techniques. These methods attempt to characterise process behaviour by modelling the complex relationships between process inputs and outputs . Predictive models can then be used to optimise or tune process parameters to obtain the desired quality output. Popular techniques used in manufacturing scenarios are Genetic Algorithms (GA), Artificial Neural Networks (ANN) [15], [16], [17], [18], [19], [20] as well as statistical methods such as analysis of variance (ANOVA) [21], [22], [23] and Principal Component Analysis (PCA) [24], [25]. Despite the availability of many techniques, the application of data mining to study complex industrial processes is hindered by the lack of appropriate method to model complex interactions and overfitting problems due to the presence of noise. Unless some prior knowledge about the underlying model is available, fitting the data with simple models, such as a linear model, would fail to capture the complex interactions [26]. On the other hand, using more complex models (e.g. polynomials) would lead to overfitting because of the presence of noise and the small amount of observations. Overfitting will then produce a model that performs very well on the available data but has very poor predictive performance. Issues such as noisy and sparse data can also affect the statistical analysis leading to situation where process engineers may not have entire trust in the results, hence missing out in process improvement opportunities. Product specific process knowledge is often necessary to discover and validate hypotheses for improvements and make the necessary adjustments to enhance process performance [25], [8], [24].

III. Process Robustness and Tolerance Synthesis

Tolerance synthesis seeks to adjust ranges of process parameters in order to reduce the variability of a given response, so that process variation is kept "close" enough to the optimal values. In order to achieve tolerance synthesis, hypotheses regarding how variability of factors affects process responses needs to be developed. The new hypotheses can be learnt from analysing steam of in-process data collected during manufacturing operations. The proposed methodology includes the following steps:

- Data pre-processing and variable selection
- Tolerance Synthesis
- Probabilistic Estimation of Process Robustness
- Quantification of Uncertainties

A. Data Pre-Processing and Variable Selection

Data collected during multi-stages processes often involve large number of variables being measured. However, typically, only a few of these variables and their interactions are those responsible for quality problems. Due to the large number of variables and noise in the data, selecting the most important variables is challenging. Variable selection for multi-stages manufacturing processes can be achieved using co-linearity index introduced by [24] and further developed by [25]. The co-linearity index is a measure of correlation and approximates the correlation coefficient in a reduced dimensional space that accounts for most of the variance. A full detailed description of the technique can be found in [24], [25]. An example of co-linearity plot can be seen in Figure 1. The factors that appear in the extreme regions delimited by the 0.2 and -0.2 lines are the most influential, hence selected for further analysis.

Mathematically, a quality objective can be represented by a bound response variable (i.e. a response variable being within a specific range). In a manufacturing environment a quality objective is, for instance, keeping the percentage of defects below a nominal threshold. A quality attribute (such as percentage of defects) can be expressed as a penalty function that represents deviation of the response value from expected or optimal result [24]. 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. Heuristic rules for the choice of the thresholds are discussed in [25]. After this data transformation, a quality output can be quantified using a penalty value Y, with $0 \leq Y \leq 1$, which indicates deviation from desired response (0 being desirable response and 1 undesirable response). Optimal and avoid outcomes of the response variables are then defined as being $Y = 0$ and $Y = 1$. The robustness of a process can be increased by discovering optimal ranges (or regions) that correlate with the occurrences of expected process outputs (results).

B. Tolerance Synthesis

An optimal range can be mathematically described in terms of quantiles of response variables. Let's assume that there are n covariates (i.e. process factors) $X_1, X_2, ..., X_n$ and an initial quality level $\alpha_0 = P(Y \leq 0) = P(Y = 0)$. Let α a given

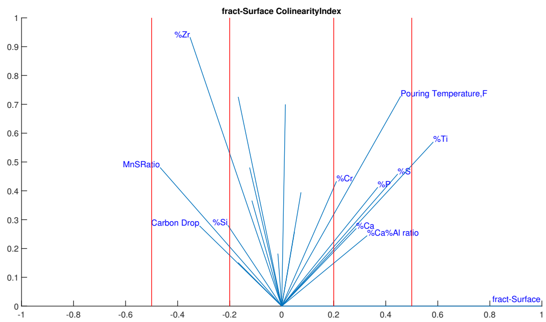

Fig. 1. The Co-linearity index is used to select the most important variables. Influential variables are those highlighted at the two extreme regions of the plot delimited by the vertical lines at -0.2 and 0.2, with the ones to the right being positively correlated to response and those to the left negatively correlated to response.

expected quality output $\alpha > \alpha_0$, an optimal range is found if the conditional α-th quantile for that range $Q_\alpha(R_{ij})$ is zero. In such case the following condition is satisfied:

$$\begin{aligned} P(Y = 0|X_i \in R_{ij}) &= P(Y \leq 0|X_i \in R_{ij}) \\ &\geq \alpha > \alpha_0 = P(Y = 0) \end{aligned} \tag{1}$$

In practical terms, this means that a tolerance limit is optimal if the probability of obtaining good responses will increase when the predictor variable is bound to that range. Optimal ranges can be found by applying quantile regression to in-process data [2]. Quantile regression is a type of regression analysis used to predict the conditional quantiles of a dependent variable Y, given values of predictor variables $X_1, X_2, \ldots X_n$ [27] and it is a robust alternative to least squares estimator in the presence of outliers or non-Gaussian distributions [27]. Quantile regression methods are appropriate to study how covariates affect, not only the centre of the distribution, but also lower and upper tail, becoming very useful when covariates have different effects on different parts of the distribution. Typically several quantile levels will be studied through quantile regression trees to achieve tolerance limit optimisation. When studying pairwise interactions between variables, this framework extends naturally by considering the conditional quantiles given two variables being simultaneously in both ranges. By using the quantile regression method described in [2] it is possible to identify those tolerance limits that are associated with optimal values of response.

C. Probabilistic Estimation of Process Robustness

A probabilistic estimation of process robustness is achieved by calculating Likelihood Ratios (LRs) associated with the new tolerance limits found at the previous step. The LR is the ratio between the conditional probabilities of obtaining optimal values and avoid values when the new tolerance limit is chosen. It is defined as follows:

$$LR = \frac{P(X_i \in Range_{ij}|Optimal)}{P(X_i \in Range_{ij}|Avoid)} \tag{2}$$

By definition, the LR measures the effectiveness of proposed tolerance limits to reduce variability of the output, hence it estimates process robustness. This definition can be extended to the combination of two variable ranges

$$LR = \frac{P(X_i \in Range_{ij} \cap X_k \in Range_{km}|Optimal)}{P(X_i \in Range_{ij} \cap X_k \in Range_{km}|Avoid)} \tag{3}$$

If $LR > 1$, the new tolerance limit (or the combination of the new tolerance limits) is associated with optimal process outcomes, while $LR < 1$ indicates association with avoid process outcomes. Another important property of the LR is that it also measures the ratio between pre-intervention (i.e. original tolerance limit) and post-intervention (i.e. new tolerance limits) odds:

$$LR = \frac{Odds(Optimal|X_{ij} \in Range_{ij})}{Odds(Optimal)}. \tag{4}$$

The likelihood ratio involves calculation of conditional probabilities, namely $P(X_i \in Range_{ij}|Optimal)$ and $P(X_i \in Range_{ij}|Avoid)$ which can be estimated from available data using the maximum-likelihood estimate as follow:

$$P(Range|c) = \frac{n_{rc}}{n_c} \tag{5}$$

where c is a class (either $Optimal$ or $Avoid$), n_{rc} is the number of observations in the given range and belonging to class c and n_c is the number of observations belonging to class c. For sparse data the calculation of the conditional probability may become problematic when only few observations for a given class exist. This often happens when analysing in-process data where the number of avoid observations might be limited. In this case estimation of the conditional probability using frequency counts might not me appropriate. In order to overcome this problem, it is suggested in the literature to use a smoothed estimation of probability[28], [29], defined as:

$$P(Range|c) = \frac{n_{rc} + m * p}{n_c + m}. \tag{6}$$

The term $m * p$ is a smoothing parameter that takes into account the prior probability p and a weighting factor m. The weighting factor is chosen by the user based on the importance given to the initial probability. The m-estimator is equivalent to Laplace estimator when $m = C$ and $p = 1/C$, where C is the number of classes. The Laplace estimators assumes that classes are distributed with uniform probability. In the current paper, the prior probability is chosen as the maximum likelihood estimation of $P(Range)$. The smoothing correction enables to change the probability of each node taking into account the prior probability of the range. Since m gives a weight to the prior probability of the class, the choice of m is determined by how much trust is placed on this initial estimation. Higher values of m indicate more confidence in the prior probability and typically lead to more conservative estimation of LR. In the industrial case study presented in this paper the value of m is chosen via numerical simulations. The approach is similar to the one described in [2].

D. Quantification of Uncertainties

Uncertainty of LR estimations obtained from in-process data and associated risks are also further evaluated with the bootstrap method to overcome the lack of process knowledge and give confidence in the results before implementing costly confirmation trials. The bootstrap method is a well know sampling technique that has been widely applied in the literature to evaluate uncertainty of different statistical parameters including correlation coefficient and regression models parameters [30] or to estimating confidence interval in diagnostic testing [31].

In this work, bootstrap with replacement is carried out to simulate the manufacturing process and quantify the uncertainty of the LR ratio estimation. Simulations with bootstrap will produce the LR empirical distribution. The method is particularly useful in the absence of additional data to test the hypotheses since bootstrapping simulates new experimental data by randomly choosing observations (with re-placement) from the available data to create a fixed number n of samples. By using bootstrap it is possible to evaluate uncertainty of the likelihood ratio estimation as shown in Figure 2.

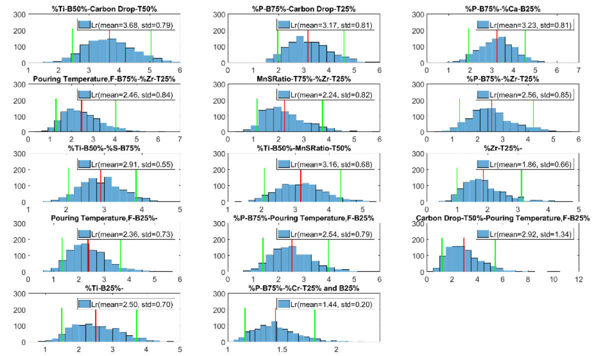

Fig. 2. Through simulation of experiments with bootstrapping, the empirical distribution of LR of optimal tolerance limits is calculated. The mean LR is represented by the red line. The green lines indicate the 95% percentiles. The interaction between Ti-B50% and Carbon Drop-T50% has a LR greater than 2 in 95% of the experiments, so Ti-B50% and Carbon Drop-T50% is considered an optimal range.

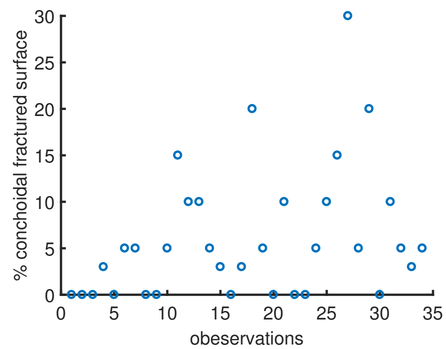

Fig. 3. The scatter of response (% of conchoidal fracture defects) shows large variability of response.

IV. Industrial Case Study

The tolerance synthesis methodology is applied to previously used dataset to improve process robustness in a low alloy steel foundry and reduce the occurrences of conchoidal fractured surface defects [10]. The dataset consists of 19 continuous process variables and one response variable, namely the percentage of conchoidal fractured. A full description of the dataset can be found in [10]. As discussed in [10], despite the fact that chemistry of the heats for the products was within the material specification, fracture tests were failing in conchoidal fracture. The plot below shows variability of values across the process (Figure 3). This variability indicates that the process is not robust. The methodology described in the previous section is applied to find ranges of factors that increase robustness.

After scaling the response using the same penalty matrix transformation as [10] (i.e. thresholds $T_{min} = 0$ and $T_{max} = 10$), the most important variables are selected with the co-linearity index methodology as shown in Figure 1. The original estimated probability of optimal responses from the data is $P(Optimal) = 0.3235$, while the target value of probability was set to $\alpha = 0.65$. Optimal ranges associated with low variability of response are then discovered through non-parametric estimation of 0.65-quantiles using the regression tree algorithm described in [2]. These ranges are displayed in Table I. Optimal tolerance limits are those whose predicted quantile is greater than a given tolerance,in this case a tolerance value was set to 0.2. The optimal tolerance limits are tested using the LR approach described in Section III-C and they are shown in Figure 2. The estimation of LR from the data allows process engineer to identify robust tolerance limits associated with reduction in variation of response. As it can be seen from Figure 2, the most significant interaction is between Ti-B50% and Carbon Drop-T50% with a mean LR of 3.68. This hypothesis is aligned with results of a previous case study presented in [10] where both Ti-B50% and Carbon Drop-B50% where found to be optimal tolerance limits. Other interactions are P-B75% and Carbon Drop-B25% (mean LR 3.16) and Ti-B50% and Mn/S-ratio T50% (mean LR 3.17). The interaction between P-B75% and Carbon Drop-B25% was not discovered in [10] but this interaction is supported by domain knowledge which confirms that %P in the melt plays a significant role in the incidence of brittle fracture in steel castings [10].

The calculation of LR allows process engineer to quantify the strength of interaction and estimate process robustness without the need of confirmation trials. This method is a further enhancement to methods developed as part of the 7Epsilon methodology [25], [24], [13] because it enables to calculate strengh of interactions and their uncertainty via simulations. This allows process engineers to make informed decision before implementing changes to their processes. Following this

TABLE I
OPTIMAL TOLERANCE LIMITS OF FACTORS DISCOVERED THROUGH QUANTILE REGRESSION.

Range Values	Range Values
Ti-B50% [0.0009 0.0102]	Carbon Drop-T50% [49 73]
P-B75% [0.009 0.013]	Carbon Drop-T25% [60 84]
P-B75% [0.01 0.013]	Ca-B25% [0.0002 0.0012]
Pouring Temp-B75% [2820 2850]	Zr-T25% [0.0044 0.0224]
MnSRatio-T75% [96 134]	Zr-T25% [0.0044 0.0224]
P-B75% [0.01 0.013]	Zr-T25% [0.0044 0.0224]
Ti-B50% [0.0009 0.011]	S-B75% [0.007 0.01]
Ti-B50% [0.0009 0.011]	MnSRatio-T50% [100 134]
Zr-T25% [0.0044 0.0224]	Main Effect
Pouring Temp-B25% [2818 2846]	Main Effect
P-B75% [0.009 0.013]	Pouring Temp-B25% [2818 2846]
Carbon Drop-T50% [51 84]	Pouring Temp-B25% [2820 2840]
Ti-B25% [0.0009 0.009]	Main Effect
P-B75% [0.009 0.013]	Cr-T25% and B25% [0.97 1.17]

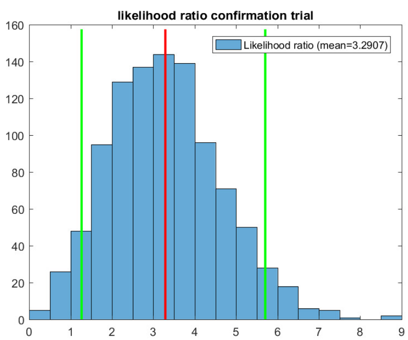

Fig. 4. The LR of the confirmation trial is calculated via simulation with bootstrapping.

study the following new tolerance limits that improve process robustness were identified as follows:

- Ti-B50% [0.0009 0.0102]
- Carbon Drop-T50% [49 73]
- P-B75% [0.01 0.013]
- Ms/S ratio-T50% [100 134]

Using the same methodology is also possible to calculate the overall LR of the confirmation trial as show in Figure 4. This measure the combined effects of the chosen new tolerance limits on the variability of response. The fact that the lower bound of the 95 percentile is greater than 1 gives confidence that the new specification will improve process robustness.

V. CONCLUSION

In this paper a methodology to increase process robustness is described. This methodology allows to discover optimal tolerance limits due to main effects and interactions between process variables. New tolerance limits are discovered with quantile regression trees without making any assumption on distributions of the variables and linearity of the relationship. The strength of improvement is quantified by calculating the Likelihood Ratio (LR) which represents the ratio between conditional probabilities of optimal and avoid events. In order to reduce bias when limited number of observations are available, a smoothing technique is used for the calculation of conditional probabilities. Uncertainty of the LR is calculated through simulations using bootstrapping, hence providing quantification of the LR estimation and its uncertainty. The applicability of the method has been shown on a data set from a previously published case study [10]. It is shown that the method is effective to isolate a small number of root causes of defects due to interactions and helps process engineers to find new tolerance limits that reduce variation in the process, hence improve its robustness.This method extends previous work [13], [8], [24] and solve the tolerance synthesis problem under uncertainty, making a contribution towards the development of novel data analytics to extend Six Sigma methodologies to Industry 4.0 scenarios.

ACKNOWLEDGEMENT

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) CHERISH-DE Centre (EP/ M022722/1); and the European Regional Development Fund (ERDF) ASTUTE project. The author would like to thank Dr Rajesh S. Ransing for proving the dataset and helpful discussion.

REFERENCES

[1] Foresight, *The future of manufacturing: a new era of opportunity and challenge for the UK*. The Government Office for Science, 2013, retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/255922/13-809-future-manufacturing-project-report.pdf (accessed 13/01/16).

[2] C. Giannetti and R. R.S., "Risk based uncertainty quantification to improve robustness of manufacturing operations," *Computers & Industrial Engineering*, vol. 101, pp. 70 – 80, 2016.

[3] Deloitte AG, *Industry 4.0 Challenges and solutions for the digital transformation and use of exponential technologies*. McKinsey Global Institute, 2015, retrieved from http://www2.deloitte.com/content/dam/Deloitte/ch/Documents/manufacturing/ch-en-manufacturing-industry-4-0-24102014.pdf (accessed 13/01/16).

[4] J. Manyika, *Manufacturing the future: the next era of global growth and innovation*. McKinsey Global Institute, 2012, retrieved from http://www.mckinsey.com/insights/manufacturing/the_future_of_manufacturing (accessed 13/01/16).

[5] International Standard Organisation ISO, *Draft BS EN ISO 9001 Quality Management Systems - Requirements*. International Standard Organization, Geneva, Switzerland, 2014.

[6] M. e. a. Glodek, "Process robustness a pqri white paper," *Pharmaceutical Engineering*, vol. 26, no. 6, pp. 1–11, 2006.

[7] N. Stricker and G. Lanza, "The concept of robustness in production systems and its correlation to disturbances," *Procedia CIRP*, vol. 19, pp. 87–92, 2014, 2nd CIRP Robust Manufacturing Conference (RoMac 2014).

[8] C. Giannetti, M. R. Ransing, R. S. Ransing, D. C. Bould, D. T. Gethin, and J. Sienz, "Organisational knowledge management for defect reduction and sustainable development in foundries," *International Journal of Knowledge and Systems Science (IJKSS)*, vol. 6, no. 3, pp. 18–37, 2015.

[9] C. Giannetti, R. S. Ransing, M. R. Ransing, D. C. Bould, D. T. Gethin, and J. Sienz, "Knowledge management and knowledge discovery for process improvement and sustainable manufacturing: a foundry case study." in *Proceedings of the 1st Sustainable Design and Manufacturing Conference, SDM14*, Cardiff, UK, 28th-30th April 2014 2014.

[10] H. Roshan, C. Giannetti, M. Ransing, and R. Ransing, "If only my foundry knew what it knows : A 7Epsilon perspective on root cause analysis and corrective action plans for ISO9001:2008," in *Proceedings of the 71st World Foundry Congress (WFC14)*, Bilbao, Spain, 19th-21st May, 2014.

[11] M. Ransing and R. Ransing, "If only my foundry knew what it knows : A 7Epsilon perspective on root cause analysis and corrective action plans for ISO9001:2008," in *61st Annual Technical Conference on Investment Casting*, Covington, KY, USA, 5-8 October 2014.

[12] 7Epsilon Website, "What is 7Epsilon?" 2015, retrieved from http://www.7epsilon.org/what-is-7Epsilon.html (accessed 20/02/16).

[13] R. Ransing, R. Batbooti, C. Giannetti, and M. Ransing, "A quality correlation algorithm for tolerance synthesis in manufacturing operations," *Computers & Industrial Engineering*, vol. 93, pp. 1–11, 2016.

[14] G. Köksal, İ. Batmaz, and M. C. Testik, "A review of data mining applications for quality improvement in manufacturing industry," *Expert systems with Applications*, vol. 38, no. 10, pp. 13 448–13 467, 2011.

[15] B. Verlinden, J. Duflou, P. Collin, and D. Cattrysse, "Cost estimation for sheet metal parts using multiple regression and artificial neural networks: A case study," *International Journal of Production Economics*, vol. 111, no. 2, pp. 484–492, 2008.

[16] I. Mohanty, D. Bhattacharjee, and S. Datta, "Designing cold rolled IF steel sheets with optimized tensile properties using ANN and GA," *Computational Materials Science*, vol. 50, no. 8, pp. 2331–2337, 2011.

[17] L. Zárate and S. Dias, "Qualitative behavior rules for the cold rolling process extracted from trained ANN via the FCANN method," *Engineering Applications of Artificial Intelligence*, vol. 22, no. 4, pp. 718–731, 2011.

[18] A. Shafyei, S. Anijdan, and A. Bahrami, "Prediction of porosity percent in AlSi casting alloys using ANN," *Materials Science and Engineering: A*, vol. 431, no. 1, pp. 206–210, 2006.

[19] M. Shabani and A. Mazahery, "The ANN application in FEM modeling of mechanical properties of Al-Si alloy," *Applied Mathematical Modelling*, vol. 35, no. 12, pp. 5707–5713, 2011.

[20] M. Esfahani, M. Toroghinejad, and S. Abbasi, "Artificial neural network modeling the tensile strength of hot strip mill products," *ISIJ international*, vol. 49, no. 10, pp. 1583–1587, 2009.

[21] Z. Cvijovic, G. Radenkovic, V. Maksimovic, and B. Dimcic, "Application of ANOVA method to precipitation behaviour studies," *Materials Science and Engineering*, vol. 297, no. 1, pp. 195–203, 2005.

[22] Z. Sun, H. Hu, and X. Chen, "Numerical optimization of gating system parameters for a magnesium alloy casting with multiple performance characteristics," *Journal of Materials Processing Technology*, vol. 199, no. 1, pp. 256–264, 2008.

[23] B. Senthilkumar, S. Ponnambalam, and N. Jawahar, "Process factor optimization for controlling pull-down defects in iron castings," *Journal of Materials Processing Technology*, vol. 209, no. 1, pp. 554–560, 2008.

[24] R. S. Ransing, C. Giannetti, M. R. Ransing, and M. W. James, "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*, vol. 64, no. 5, pp. 514–523, 2013.

[25] C. Giannetti, R. Ransing, M. Ransing, D. Bould, D. Gethin, and J. Sienz, "A novel variable selection approach based on co-linearity index to discover optimal process settings by analysing mixed data," *Computers & Industrial Engineering*, vol. 72, no. 0, pp. 217 – 229, 2014.

[26] B. Bakır, I. Batmaz, F. Güntürkün, İ. İpekçi, G. Köksal, and N. Özdemirel, "Defect cause modeling with decision tree and regression analysis," *World Acad Sci Eng Technoly*, vol. 24, pp. 1–4, 2006.

[27] R. Koenker and G. Bassett Jr, "Regression quantiles," *Econometrica: journal of the Econometric Society*, pp. 33–50, 1978.

[28] B. Cestnik and I. Bratko, "On estimating probabilities in tree pruning," in *Machine LearningEWSL-91*. Springer, 1991, pp. 138–150.

[29] L. Jiang and C. Li, "An empirical study on class probability estimates in decision tree learning," *Journal of Software*, vol. 6, no. 7, pp. 1368–1373, 2011.

[30] B. Efron and R. Tibshirani, "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy," *Statistical science*, pp. 54–75, 1986.

[31] R. W. Platt, J. A. Hanley, and H. Yang, "Bootstrap confidence intervals for the sensitivity of a quantitative diagnostic test," *Statistics in Medicine*, vol. 19, no. 3, pp. 313–322, 2000.