On the Use of Johnson's Distribution in Quality and Process Improvement

M. Aichouni, N. Ait Messaoudene, A. Al-Ghonamy

***Abstract*—** Modern Quality Improvement methodologies such as Total Quality Management (TQM) and Six Sigma use statistical process control (SPC) techniques to improve processes and assure high quality delivery to customers. Statistical Process Control charts used in industry and services by quality professionals require that the quality characteristic of interest follows a normal distribution. However, in real business situations such as in the construction industry, process distributions show departure from normality. Any conclusions drawn from standards control charts on the stability of the process and its capability to meet customer requirements may be misleading and erroneous. In the present paper, an alternative approach, based on the identification of the best distribution that would fit the quality characteristic data, is proposed. In particular, Johnson distribution is used as a model for normalizing real field data showing departure from normality. Through such an approach, false alarms of variability can be detected and unnecessary corrective actions can be avoided. An illustrative example from the construction industry is used to demonstrate the validity of the analysis.

***Keywords*—** Quality Improvement, Statistical Process Control, control charts, non-normal data, Johnson System of distributions, Construction Industry

I. Introduction

STATISTICAL Process Control is a process improvement methodology widely used by modern manufacturing and service organizations. This methodology is mainly based on the use of control charts and frequency distributions of process and quality characteristics data [1]. Common and well established control charts include the Shewhart control chart ($\overline{X}_{-R}$ and $\overline{X}_{-S}$ charts), the cumulative sum control chart (CUSUM) and the exponentially weighted moving average control chart (EWMA). In process improvement strategies, these control charts are used to monitor product quality and detect special events that may cause out-of-control situations that would lead to an unstable process with unpredictable outcomes. Such processes deliver poor quality products and services to customers who expect suppliers to provide proof of process control and process capability. Control charts help organizations management to continuously improve processes, by making them more stable and capable of meeting customer specifications, thus achieving business excellence.

Standards (Shewhart) control charts are designed on the assumption that the process being monitored produces a quality characteristic that can be approximated by a symmetrical normal distribution, when only the innate sources and common causes of variability are present in the process. The central limit theorem can be used to approximate distributions to the normal distribution provided that the samples being measured and monitored would be large enough, with sample size > 30 [1,2]. However, in many industrial situations, this cannot be assured and the process output is not normally distributed and heavy tailed and skewed. Experience has showed that in some manufacturing processes, such as chemical processes parameters, cutting tool wear processes and some concrete production processes, the distribution are usually skewed. In this case, standard control charts based on normality assumptions can lead to erroneous conclusions regarding the stability and the capability of the process. Moreover, even interpretations based on Shewhart control chart can yield to misleading results when the underlying data is not normal [2]. Such wrong conclusions would cost manufacturing and service organization big financial losses and lost customers to competitors.

With the advents in statistical theories and computing facilities, this can be easily solved, by understanding of distributions that provide good model for most non-normal quality characteristics. Such an approach has been reported in the technical literature [3]-[6]. Derya and Canan [6] developed standards control charts based on Weibull, Gamma and lognormal distributions. Sherill and Johnson [5] showed the possibility to use exponential, Weibull, Lognormal, Box-Cox and the Johnson distributions for transforming non-normal data process control and capability calculations for a chemical process. Kan and Yazici [7] proposed a method for designing control charts with an appropriate correction for skewness.

The objective of the present paper is to examine the use of the Johnson's family of distributions to model control charts that can be used for process improvement purposes. A real

This research work has been undertaken within the Binladin Research Chair on Quality and Productivity Improvement in the Construction Industry, funded by the Saudi Binladin Constructions Group; this is gratefully acknowledged. The opinions and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the sponsoring organization.

M. Aichouni is a Principal Investigator of "*Binladin Research Chair on Quality and Productivity Improvement in the Construction Industry*", College of Engineering., Hail University, Saudi Arabia, phone:+966 165358276; e-mail: m.aichouni@uoh.edu.sa.

N. Ait Messaoudene is a Principal Investigator of "*Binladin Research Chair on Quality and Productivity Improvement in the Construction Industry*", College of Engineering., Hail University, Saudi Arabia; e-mail: n.messaoudene@uoh.edu.sa.

A. Al-Ghonamy is the Chairperson of the "*Binladin Research Chair on Quality and Productivity Improvement in the Construction Industry*", College of Engineering., Hail University, Saudi Arabia; e-mail: a.alghonamy@uoh.edu.sa.

field case study is presented for ready mixed concrete production plants where process distribution showed a skewed non-normal distribution.

II. The Johnson's Frequency Distribution in Quality Improvement

Quality professionals together with statisticians are often faced with the problem of summarizing a set of data by means of a mathematical function which fits the data and allow obtaining estimates of percentiles, quartiles and other statistics for process improvement purposes. Frequently, quality professionals have insufficient theoretical grounds for selecting a model like normal, gamma or extreme-value distributions for a "real world" data set from manufacturing or services processes [8]. Usually data are obtained and empirical methods are used to draw conclusions and make decisions on process and quality improvement in real business situations. The fitting of empirical distributions to data has a long history, and many different procedures have been advocated. The most common of these is the use of normal distribution. The central limit theorem leads one to expect this distribution to provide a reasonable representation for many, but not all, real world phenomena [9].

Although models like gamma, log-normal and beta distributions do lead to a wide diversity of distribution shapes, they still do not provide the degree of generality that is frequently desirable. In 1949, Johnson derived a system of curves that has the flexibility of covering a wide variety of shapes [8]. This system has the practical and theoretical advantages of being able to transform these curves to the normal distribution. The Johnson system is able to closely approximate many of the standard continuous distributions through one of three functional forms and is thus highly flexible. The Johnson system provides one distribution corresponding to each pair of mathematically possible values of skewness and kurtosis. Any data set can be fitted by a member of the Johnson families such as S_U , S_L , and S_B. This family of distributions [8] is perhaps the most versatile choice. It is based on a transformation of the standard normal variable, and includes four forms:

1. Unbounded: the set of distributions that go to infinity in both the upper or lower tail.
2. Bounded: the set of distributions that have a fixed boundary on either the upper or lower tail, or both.
3. Log Normal: a border between the unbounded and bounded distribution forms.
4. Normal: a special case of the unbounded form.

The fact that the Johnson system involves a transformation of the raw variable to a normal variable allows estimating of the percentiles of the fitted distribution to be calculated from the Normal distribution percentiles, for use in control limits calculations (on the Individual $\overline{X}$-chart or the $\overline{X}_{-R}$ charts) or for Capability Analysis [5], [10]. Thus, although capability indices and control limits are generally only defined for normal variables, this approach allows their calculation for all distribution types [10]. In the present study, the Johnson system, which includes the S_U , S_L , and S_B distributions, is considered and applied for processing data from construction industry as it is able to accommodate all theoretically feasible skewness (β_1)-kurtosis(β_2) combinations (Fig. 1).

The standard process capability analysis is one of many statistical process control widely used in manufacturing and services engineering. It is based on the assumption that process data are normally distributed. When this condition cannot be guaranteed, either capability indices should be computed based on distributions other than normal, or the data should be transformed so that it conforms better to the normal distribution [3]. Sherill and Johnson [5] and many others have shown that the use of Box-Cox and Johnson transformations would help quality professionals to perform correct process analysis using both control charts for process stability and capability indices for process capability for meeting customer specifications. In addition, it is worth mentioning that in a recent study, Kilink et al [11] have showed that compressive strength data of concrete elements in buildings are best modeled using log-normal and the Johnson S_B distributions.

III. Johnson's Distribution Mathematical Model

As stated earlier, when process data exhibit non-normal distribution, it is erroneous to draw standards control charts for process improvement and perform process stability and capability analyses. The practical solution is to transform the data and drive them towards normality, using common and well established probability distributions, such as Box-Cox, log-normal or the Johnson distribution. Such an approach has been used in the open literature ([8], [9], [10], [11]). Basically, the Johnson transformation computes an optimal transformation function from three flexible distribution families (S_U, S_B, and S_L). This makes this transformation more powerful than other distributions [5].

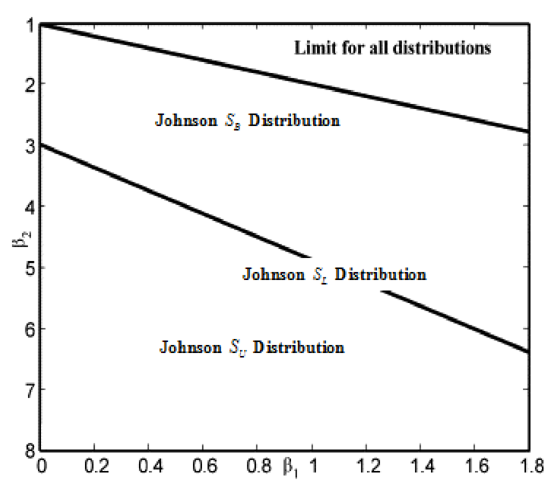

Fig. 1 Chart for Johnson subsystem identification

These translations transform any continuous random variable X into a standard normal variable Z using general form:

$$Z = a + bg(\frac{X-\mu}{\sigma}) \qquad (1)$$

Where: a and b are shape parameters, μ is a location parameter, and $g(x)$ is a function defining the Johnson system of families, determined as:

$$g(x) = \begin{cases} \ln(x), & \text{for the lognormal family,} \\ \ln\left(x + \sqrt{x^2+1}\right), & \text{for the unbounded family,} \\ \ln\left(\frac{x}{1-x}\right), & \text{for the bounded family,} \\ x, & \text{for the normal family.} \end{cases}$$

As discussed in [8], the above system has the flexibility to match any feasible set of values for the mean, variance, skewness, and kurtosis coefficients. With this system, the skewness and kurtosis also uniquely identify the appropriate form for the (g) function.

A. Johnson's Translation System

Johnson proposed three normalizing transformations having the general form:

$$Z = \gamma + \sigma\ f\left(\frac{X-\mu}{\lambda}\right) \qquad (2)$$

Where $f(x)$ denotes the transformation function, Z is a standard normal random variable γ and σ are shape parameters, λ is a scale parameter and μ is a location parameter. Without loss of generality, it is assumed that $\sigma > 0$ and $\lambda > 0$.

The first transformation proposed by Johnson defines the lognormal system of distributions denoted by S_L:

$$Z = \gamma + \sigma \ln\left(\frac{X-\mu}{\lambda}\right) = \overset{*}{\gamma} + \sigma \ln(X-\mu), \quad X > \mu, \qquad (3)$$

The bounded system of distributions S_B is defined by:

$$Z = \gamma + \sigma \ln\left(\frac{X-\mu}{\mu+\lambda-X}\right) = \quad \mu < X < \mu + \lambda, \qquad (4)$$

S_B curves cover bounded distributions. The distributions can be bounded on either the lower end, upper end, or both. This family covers gamma distributions, beta distributions and many others.

The unbounded system of distributions S_U is defined by:

$$Z = \gamma + \sigma \ln\left[\left(\frac{X-\mu}{\lambda}\right) + \left\{\left(\frac{X-\mu}{\lambda}\right)^2 + 1\right\}^{1/2}\right] \qquad (5)$$

$$= \gamma + \sigma \sinh^{-1}\left(\frac{X-\mu}{\lambda}\right), \quad -\infty \prec X \prec +\infty$$

The S_U curves are unbounded and cover the t and normal distributions, among others.

B. Johnson's Family of Distributions

The Johnson family of distributions is made up of three distributions, Johnson S_U, Johnson S_B and lognormal. It covers any specified average, standard deviation, skewness and kurtosis. Together, they form 4-parameter family distributions that cover the entire skewness-kurtosis region other than the impossible region. The Johnson S_U distribution covers the area above the lognormal curve and the Johnson S_B covers the area below the normal curve. A family of distributions is constituted of several distributions combined so that they cover a well defined region in a skewness and kurtosis plot (lognormal family of distributions, negative lognormal and normal distributions). Detailed developments about the Johnson family of distributions can be found in reference books [9].

This family of distributions is usually parameterized as a function of skewness and kurtosis. Skewness is a measure of non symmetry in the data; so for a normal distribution it takes the value of zero. Negative values of skewness indicate that data are skewed left, and positive values indicate that data are skewed right. On the other hand, kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. The kurtosis for a normal distribution is 3.0. A kurtosis value larger than 3.0 indicate a “peaked” distribution and a kurtosis value less than 3.0 indicates a “flat” distribution. Thus, both can be seen as measures of the shape of distributions.

IV. Application of Johnson's System of Distributions for Real Field Data In The Construction Industry

In order to illustrate the above analysis, real field data from the construction industry is chosen as a case study. There is no need to demonstrate the importance of such business for emerging economies such as Saudi Arabia. The Kingdom of Saudi Arabia has been rated as the 13th most economically competitive country in the world, according to the International Finance Corporation (IFC)-World Bank annual "Doing Business" report issued for 2010 [12]. The report highlights the rapid rate of economic growth among Middle Eastern countries, especially in the construction industry. According to a recent study conducted by the national research institution KACST [13], the Saudi construction industry counts for 8 percent of the national GDP.

Data from Ready mix concrete plants are gathered and analyzed using Minitab 16 statistical software. The observed quality characteristic is the compressive strength (kg/cm^2) of concrete as defined by international quality standards (ACI-214) [14]. The gathered data consist of 22 samples of concrete with a nominal specification for the compressive strength equal to 350 kg/cm^2. The sampling process consists of a sample size of 3 spanning over a period of 22 days. These data are presented in table I.

Initial analysis of the concrete data using standard $\overline{X}$ chart (Fig. 2) shows that the process is out of statistical control.

This means that the process is affected by special causes of variations. The plant's quality professional would need to perform root cause analysis to determine the causes of such out of control situations. These causes would in turn require some managerial and engineering corrective actions on the different components of the process.

Table I – Data for compressive strength for Ready Mixed Concrete (Kg/cm^2)

Sample	Cylinder 1	Cylinder 2	Cylinder 3
1	353.8	363	360.6
2	357.8	358.7	370.9
3	365.2	360	356.6
4	340.4	335.2	330.1
5	359.6	358.1	351.2
6	368.1	366.7	369.3
7	357.9	355.0	350.6
8	337.8	352.6	361.6
9	359.1	349.2	363.7
10	361.1	358.2	358.3
11	358.3	345.7	341.7
12	357.3	359.2	356.9
13	352.6	363.1	374.6
14	360.8	356.2	352.7
15	347.5	339.8	354.3
16	358.2	359.5	353.9
17	375.2	372.5	370.2
18	357.5	359.5	348.9
19	343.2	355.8	362.4
20	362.1	356.6	359.1
21	365.2	362	359.4
22	361.3	346.8	339.0

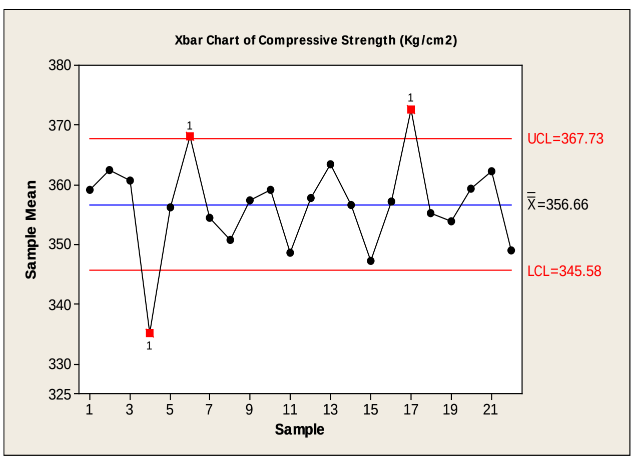

Fig. 2 Standards $\overline{X}$ chart for the concrete compressive strength

The out of control situation shown in Fig. 2 is based on an $\overline{X}$ control chart assuming normally distributed concrete data. But the question arises about the validity of such assumption and about the best distribution that fits real field data in case it is not valid. To answer this question, distribution identification is carried out for the data, and the outcome is presented in Fig. 3 in the form of probability plots. From this figure, it can be seen that concrete compressive strength does not fit any of the normal, exponential, Weibull, or lognormal distributions since the calculated P-value for each distribution is less than 0.05. It is obvious that the exponential distribution is a poor model for concrete compressive strength test results. Consequently, the Johnson distribution is considered an alternative in this case, as suggested in recent studies [5], [11] and [15].

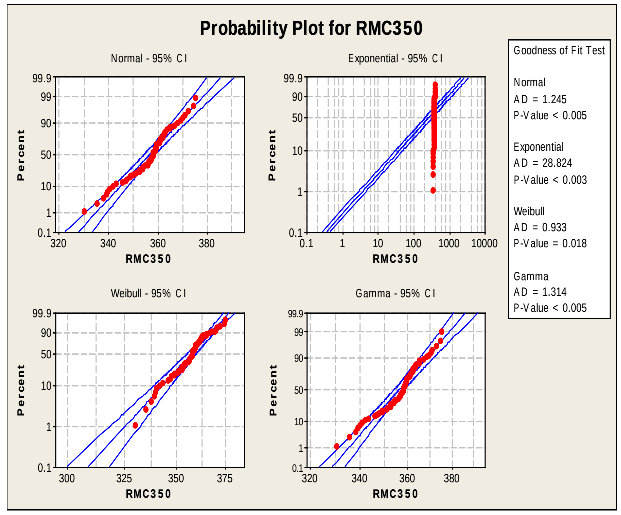

Fig. 3 Probability Plots for the Concrete Compressive Strength

The transformed data using the Johnson system are illustrated in Fig. 4. It can be seen that the S_U Johnson distribution would be the best model for the considered concrete data. Moreover, it can be seen that the best fit of the data is within the percentile interval ranging from 1.054 to 98.94 percent. In other worlds, normality can be guaranteed within this interval. This allows determining both the lower and upper control limits for the normalized data; their values are UCL=375.2 (kg/cm^2) and LCL=330.1 (kg/cm^2). These control limits are used as new control limits for the $\overline{X}$ chart as shown in Fig. (5). It is clear that the control chart with new control limits totally contradicts the previous conclusion drawn from the standard control chart which is based on the assumption of normally distributed process data. The actual process is shown to be in statistical control and does not require any intervention from the production personnel or the quality professional. The early out of control situations indicated by the standard control chart can therefore be considered as a false alarm for process.

Similar observations have been reported in the literature for manufacturing processes and for pavement and concrete production plants. Robert and Sherrill [5] show the importance for quality professionals and statisticians to know how to use transformations for creating control charts in the case of quality characteristics that are not normally distributed in order to avoid false alarms about process stability and capability. In a recent paper, Uddin et al [15] show that skewness and kurtosis originating from non normal distributions in highways pavement projects result in pay factors that are very different from those calculated based upon the assumption of normality. Therefore, corrective measures should be taken when construction data are not normally distributed. These measures may include data cleaning, use of median instead of mean, and data transformation using mathematical models such as Johnson transformations. This is shown to significantly reduce non-normality errors and produce less bias in pay factors in highways construction projects [5].

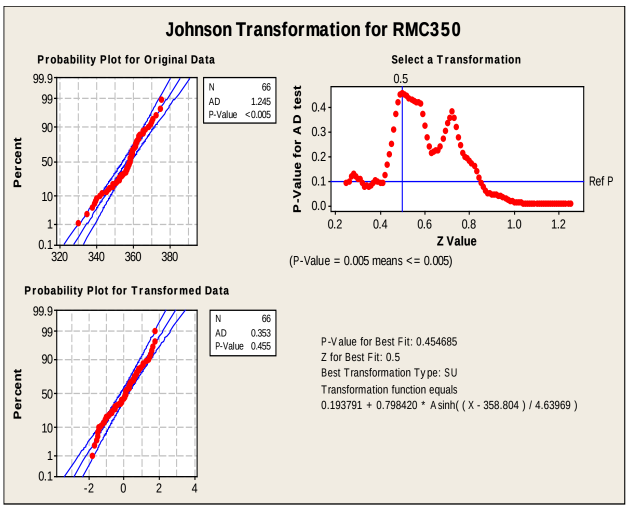

Fig. 4 Probability Plots for the Johnson Transformed data of Concrete Strength

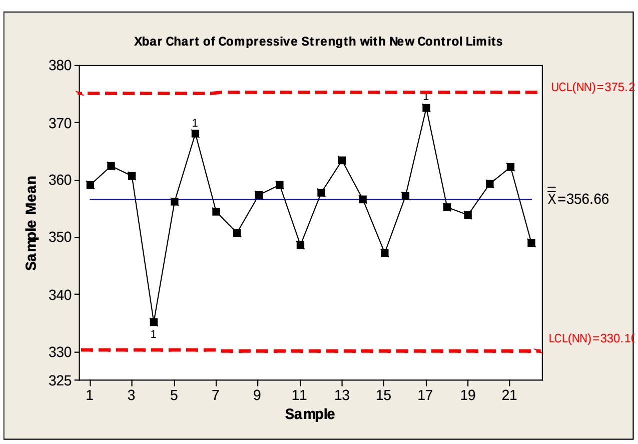

Fig. 5 $\overline{X}$ -chart with the new Control Limits Calculated using Johnson transformations

V. Conclusion

Quality improvement methodologies such as Total Quality Management (TQM) and Six Sigma use statistical process control (SPC) techniques for improving processes and yielding high quality delivery to customers. SPC, which is concerned with process stability and process capability to deliver products and services meeting customer specifications, is mainly based on control charts. Standard control charts, known as Shewhart control charts, require monitored quality characteristic to follow a normal distribution. If, in reality, the distribution of the quality characteristic of interest is not normal, conclusions about the stability of the process drawn from control charts may be misleading and highly erroneous. In this study, an alternative approach to standard control charts based on the identification of the best distribution fitting the data is suggested. In particular, the Johnson distribution is used as a model for normalizing real field data departing from normality.

Real field data from the construction industry is used as a case study for illustrating the validity of the proposed analysis. It is shown that the considered data, namely compressive strength, does not follow a normal distribution, On the other hand, the control chart based on the assumption of normality leads to the conclusion that the monitored process is out of statistical control. This usually indicates that some special causes affect the process, requiring corrective actions from the management in order to eliminate them and prevent them from occurring again. This, in turn, would certainly infer a cost to the organization. However, when the data are recognized not to be normal and treated through Johnson transformations with computation of corrected control limits, the actual control chart no longer exhibits any out of control situation or sign of special causes of variation. Therefore, in such situations of non normal data, the proposed approach can avoid false alarms of process variability, thus avoiding unnecessary and costly corrective actions.

References

[1] H. S. V. Alena, Application of Fuzzy Principles in Evaluating Quality of Manufacturing Process, WSEAS Transactions on Power Systems, Vol. 7 No 2 (2012) pp. 50-59.

[2] A. Jamali and J. Li, False alarm rates for the Shewhart control chart with interpretation rules, WSEAS Transactions on Information Science and Applications Vol.3 No 7 (2006) pp. 238-241.

[3] R. N. Farnum., Using Johnson Curves to Describe Non-Normal Process Data, Quality Engineering, Vol. 9, No. 2, December, 1996, 329-336.

[4] Y. Chou, A. M. Polansky, and R. L. Mason, Transforming Non-normal Data to Normality in Statistical Process Control", Journal of Quality Technology, Vol. 30, April, 1998pp 133-141.

[5] R. W. Sherill and L. A. Johnson, Calculated Decisions, Quality Progress, Vol. 42 (1), 2009, pp. 30-35.

[6] K. Derya, and H. Canan, Control Charts for Skewed Distributions: Weibull, Gamma and Lognormal, Metodoloski zvezki, Vol. 9, N. 2, 2012 pp. 95-106.

[7] B. Kan and B. Yazici, The individuals control charts for Burr distributed and Weibull distributed data, WSEAS Transactions on Mathematics. 2006.

[8] N. L. Johnson, Systems of frequency curves generated by methods of translation, Biometrika, Vol. 36, 1949, 149-176.

[9] J. G. Hahn and S. S. Shapiro, Statistical models in Engineering, John Wiley and Sons, 1967.

[10] N. L. Johnson, S. Kotz and N. Balakrishnan, Continuous Univariate Distributions, Second Edition, New York: John Wiley & Sons., 1994.

[11] K. Kilink, A. Celik, M. Tuncan, G. Arslan and O. Arioz, Statistical distributions of in situ microcore concrete strength, Construction & Building Materials, Vol. 26 Issue 1, Jan2012, p393-403.

[12] International Finance Corporation (IFC)-World Bank annual "Doing Business" report issued for 2010, Available : https://openknowledge.worldbank.org/bitstream/handle/10986/11857/DB13_Full%20report.pdf?sequence=1

[13] KASCT , Strategic Priorities for Building and Construction Technology, 2009

[14] Report ACI Committee 214, Evaluation of Strength Test Results of Concrete (ACI 214R-02), 2005.

[15] M. Uddin, K. C. Mahboub and P. M. Goodrum, Effects of Nonnormal Distributions on Highway Construction Acceptance Pay Factor Calculation, Journal of Construction Engineering and Management, Vol. 137, No. 2, 2011.

M. Aichouni is currently Professor of Quality Engineering at the Engineering College, University of Hail (Saudi Arabia). He is Principal Investigator of the "The Binladin Research Chair on Quality and Productivity Improvement in the Construction Industry". His research interests include quality control, statistical process control, metrology, fluid flow metering systems, experimental and computational fluid dynamics, engineering education, curriculum development, and web-based education. Professor Mohamed earned his PhD degree in Mechanical and Aeronautical Engineering from the University of Salford, Manchester, the United Kingdom in 1992 and a Bachelor degree in Marine Engineering (Mechanical Machinery) from the University of Sciences and Technology of Oran (Algeria) in 1987. He authored six textbooks in the areas of quality control, Statistical Process Control and dimensional metrology, and more than fifty papers in refereed journals and international conferences proceedings.

N. Ait Messaoudene is currently Professor of Mechanical Engineering at the College of Engineering, University of Hail (Since September 2010). He is a Principal Investigator at the "Binladin Research Chair on Quality and Productivity Improvement in the Construction Industry". He earned a PhD from the Department of Mechanical and Aerospace Engineering of Case Western Reserve University (Cleveland, Ohio, USA) in 1989, a Master degree from the same university in 1985 and a BSc from the National Polytechnic Institute of Algiers (Algiers, Algeria) in 1982. Prior to joining Hail University, Professor Ait Messaoudene acted as full Professor at the Department of Mechanical Engineering, University of Blida in Algeria (1989-2010) and Head of the Laboratory of Energetic Applications of Hydrogen which he initiated in 2003.

A. Al-Ghonamy is Associate Professor of Civil and Architecture Engineering at the Engineering College, University of Hail. He the Chair person of "The Binladin Research Chair on Quality and Productivity Improvement in the Construction Industry". His research interests include quality in construction, construction materials and constructions projects management. Dr Al-Ghonamy is the Dean of the Engineering College at the Univesity of Hail, Saudi Arabia.