Statistical Variation Analysis Using Pearson Distribution Family Based on Jacobian-Torsor Model

Siyi Ding[1,2], *Sun* Jin[1,2, *], *Zhimin* Li[1,2], *Fuyong* Yang[1], *Jia* Lin[2]

[1]State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai, 200240, P.R. China
[2]Shanghai Key Laboratory of Digital Manufacture for Thin-walled Structures, Shanghai Jiao Tong University, Shanghai, 200240, P.R. China

Abstract. Assembly variations are unavoidable due to parts' geometrical errors. Statistical variation analysis is an effective method to quantitatively predict product quality in the original design stage. However, traditional methods can't handle the problem of abnormal distribution of the actual variation variables. Meanwhile, they are underdeveloped in regard to the complex geometrical errors in spatial 3D state. To overcome this problem, firstly, Jacobian-Torsor model is used to build the variation propagation, which is well suited to a complex assembly that contains large numbers of joints and geometric tolerances; secondly, Pearson distribution family is adopted to determine probability distribution pattern and build probability density function. By comparing results of the suggested method to the Monte Carlo method, it is observed that this novel method has the same accuracy, but much higher efficiency. The results also demonstrate that probability distribution types of the parts variations have a significant impact on the final assembling variation.

1 Introduction

Variation is inherent during any manufacturing and assembly process owing to the variability of the part, datum, positioner, fixture, etc. They cause minor variations in components from the nominal geometry [1]. Statistical variation analysis is an effective way to predict product quality quantitatively in the design stage [2]. Traditional methods are still based on engineering experience, extremum method [3, 4], or the mean square root method [5, 6]. The results obtained by extremum method are overly pessimistic which will cause high fabricating cost, while the mean square root method only gives results for the mean-square variation which is primarily used to solve the plane size-chain problem. These methods are underdeveloped in regard to the geometrical errors in spatial 3D state. Meanwhile, they widely adopt normality assumption which is convenient for statistical evaluation.

As is known to all, most variables of parts variations disobey normal distribution in the actual machining process. For instance, if tool wearing plays a leading role in the machining process, the actual variation variables will satisfy uniform distribution; if one manufacturing process is accomplished using two machine tools, the mixing variables of parts variations will follow bimodal distribution; when turning shaft journals or apertures, operators usually keep a certain allowance, and this will lead to skewed distribution. Monte Carlo (MC) simulation is suitable for the data of abnormal distributions. Currently, MC method has been developed into a core of some commercial softwares such as CETOL, VSA, 3DCS, etc. A big disadvantage to this approach, however, is that large amounts of data samples are required to achieve accurate predictions [7], which inevitably result in time-consuming computation. Thus there is a clear need to adopt a simple and rapid probabilistic approach for statistical variation analysis.

Yang et al. [8] studied the probability distribution of table-axis error in aero-engine assembly. Probability density functions (PDF) were derived based on Rayleigh distribution. Ghie [9] applied uncertain measurements into Jacobian-Torsor model. But the PDF deduced could not consider skewed distribution of variables. Ahsanullah and Hamedani [10] attempted to investigate various characterizations of certain univariate continuous distributions, providing a base of further research of multivariable continuous distributions. Kharoufehy and Chandraz [11] adopted a multivariate kernel density estimate to approximate the probability distribution. But it was difficult to determine the smoothing parameter (h). In this paper, we will introduce an innovation of Pearson distribution family [12] to model PDF that covers polymorphic probability distributions.

In addition, Jacobian-Torsor theory [13] will be used to model the variation propagation, which is well suited to a complex assembly that contains large numbers of joints and geometric tolerances. It combines the advantages of the torsor model which is suitable for tolerance representation and the Jacobian matrix which is suitable for tolerance propagation.

---

* Corresponding author: jinsun@sjtu.edu.cn

2 Related works

2.1 Jacobian–Torsor model

The unified Jacobian-Torsor model is composed of the Jacobian matrix and torsor model. Torsor model uses three translational vectors and three rotational vectors to express the position and orientation of an ideal surface. Supposing that S_0 is the nominal position and S_1 is an arbitrary variational surface relative to S_0 with tolerance t, the torsor model of S_1 which represents position and orientation can be represented as:

$$T_{1-0} = \begin{bmatrix} u & v & w & \alpha & \beta & \gamma \end{bmatrix}^T \qquad (1)$$

where u, v, w are three specific vectors along the axes x, y and z in local reference system respectively; likewise, α, β and γ indicate the vectors around the axes x, y and z respectively.

These vectors are transferred to the objective controlled variable by virtue of the Jacobian matrix. Jacobian matrixes describe relative positions, as well as relative orientations between local reference frames and the global reference frame. The unified Jacobian-Torsor model can be expressed as:

$$\begin{bmatrix} (u,\bar{u}) \\ (v,\bar{v}) \\ (w,\bar{w}) \\ (\alpha,\bar{\alpha}) \\ (\beta,\bar{\beta}) \\ (\gamma,\bar{\gamma}) \end{bmatrix}_{FR} = \begin{bmatrix} [J]_{FE1} & \cdots & [J]_{FEn} \end{bmatrix} \times \begin{bmatrix} \begin{bmatrix} (u,\bar{u}) \\ (v,\bar{v}) \\ (w,\bar{w}) \\ (\alpha,\bar{\alpha}) \\ (\beta,\bar{\beta}) \\ (\gamma,\bar{\gamma}) \end{bmatrix}_{FE1} & \cdots & \begin{bmatrix} (u,\bar{u}) \\ (v,\bar{v}) \\ (w,\bar{w}) \\ (\alpha,\bar{\alpha}) \\ (\beta,\bar{\beta}) \\ (\gamma,\bar{\gamma}) \end{bmatrix}_{FEn} \end{bmatrix}^T \qquad (2)$$

where α, β, γ, u, v and w are the same as Eq. (1); n is the number of functional elements (FEs); $(\alpha,\bar{\alpha})$ is tolerance interval where α should belong to; likewise, other torsors of components follow the similar way as α.

There are two types of FE pairs in assembly, i.e., internal FE (IFE) pairs which are composed of two FEs on the same part and contact FE (CFE) pairs which are made up of two FEs on different parts. Variations of each FE pair can be transferred to functional requirement (FR) with a Jacobian matrix, which of the ith FE can be expressed as:

$$[J]_{FEi} = \begin{bmatrix} [R_0^i]_{3\times3} \cdot [R_{Pti}]_{3\times3} & \vdots & [W_i^n]_{3\times3} \cdot ([R_0^i]_{3\times3} \cdot [R_{Pti}]_{3\times3}) \\ \cdots & \vdots & \cdots \\ [0]_{3\times3} & \vdots & [R_0^i]_{3\times3} \cdot [R_{Pti}]_{3\times3} \end{bmatrix} \qquad (3)$$

where $[R_0^i]$ represents the local orientation of the ith frame with respect to the 0th frame which is global reference system; $[R_{Pti}]$ is a projection matrix designating the unit vectors along local axes respectively for tolerance zone tilted according to the direction of tolerance analysis; $[W_i^n]$ is a skew-symmetric matrix [13] allowing the representation of the vector among the ith and nth frame (end point).

2.2 Pearson distribution family

Pearson system of distributions was firstly proposed by statistician K. Pearson in 1895, which included various types of distributions. Pearson system adopts four parameters to characterize each distribution. The probability density function (PDF) $f(x)$ depends on the following differential equation:

$$\frac{df(x)}{dx} = \frac{(a_0 + a_1 x) f(x)}{b_0 + b_1 x + b_2 x^2} \qquad (4)$$

While $a_1=1$, the parameters a_0, b_0, b_1, and b_2 depend entirely on the first four order central moments of distributions μ_1, μ_2, μ_3 and μ_4. It can also be expressed as moment ratio, i.e. $\beta_1 = \mu_3^2 / \mu_2^3$ and $\beta_2 = \mu_4 / \mu_2^2$, as follows:

$$a_0 = -b_1 \qquad (5)$$

$$b_0 = A\sigma^2(4\sigma^2\mu_4 - 3\mu_3^2) = B\sigma^2(4\beta_2 - 3\beta_1) \qquad (6)$$

$$b_1 = A\mu_3(\mu_4 + 3\sigma^4) = B\sigma\sqrt{\beta_1}(\beta_2 + 3) \qquad (7)$$

$$b_2 = A(2\sigma^2\mu_4 - 3\mu_3^2 - 6\sigma^6) = B(2\beta_2 - 3\beta_1 - 6) \qquad (8)$$

where $\qquad A = -(10\sigma^2\mu_4 - 18\sigma^6 - 12\mu_3^2)^{-1} \qquad (9)$

$$B = -(10\beta_2 - 12\beta_1 - 18)^{-1} \qquad (10)$$

The general solution of $f(x)$ are:

$$f(x) = C\exp\left\{-\int \frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2} dx\right\} \qquad (11)$$

To be specific, Pearson system mainly contains seven distribution patterns, which are elaborated in [14]. These types can be distinguished according to the different distributions of roots of the quadratic equation $b_0 + b_1 x + b_2 x^2 = 0$. Supposing $D = b_1^2 - 4b_0 b_2$, $\lambda = b_1^2 \div (4b_0 b_2)$, the Pearson distribution pattern can be judged by the value of D and λ, as provided in Table.1.

Table 1. Pearson distribution types.

a_1	D	λ	Distribution pattern	Remarks
0	0	0	Exponential distribution	
$\neq 0$	>0	$-\infty$	Pattern VII	$b_0=0$
		$(-\infty,0)$	Pattern I	
		0	Pattern II	
		$(1,\infty)$	Pattern VI	
		∞	Pattern III	
	=0	1	Pattern V	$b_1=b_2=0$
		Uncertain	Normal distribution	
	<0	$(0,1)$	Pattern IV	

3 Statistical analysis

According to the Pearson system's characteristics, distribution pattern can be determined and PDF can be derived while the first four central moments are known. As for arbitrary distributions, the first-order central moment μ_1 is identically equal zero; the second-order central moment μ_2 reflects data dispersion degree, which is equal to variance σ^2; the third-order central moment μ_3 is related to coefficient of skewness S, and $S = \mu_3/\sigma^3$; and the μ_4 is related to coefficient of kurtosis K, and $K = \mu_4/\sigma^4$. Due to the average value, variance, skewness and kurtosis of a probability distribution can visually reflect the shape of the distribution, we will solve these four parameters in this paper to indirectly gain the first four central moments.

Eq. (2) can be represented as $y_{FR} = f(u_i, v_i, w_i, \alpha_i, \beta_i, \gamma_i,)$, $(i=1,2,\cdots,n)$, where y_{FR} is the FR's variation, and (u_i,

vi, wi, αi, βi, γi,) is the *FE*'s shifting variation or rotating variation in all directions. This representation is fit for deterministic variation analysis and beneficial to the error compensation. Supposing all of the *FE*'s variations are randomly distributed and independent of one another, we can solve the four parameters of the *FR* by means of the *FE*'s parameters. Then a_0, b_0, b_1, b_2 can be obtained with the known distribution parameters of *FR*. Using the suggested method (Table.1), distribution pattern can be distinguished and PDF can be solved, after which the distribution curve can be drawn and assembly qualified rate will be calculated. The analysis process is illustrated in Fig.1.
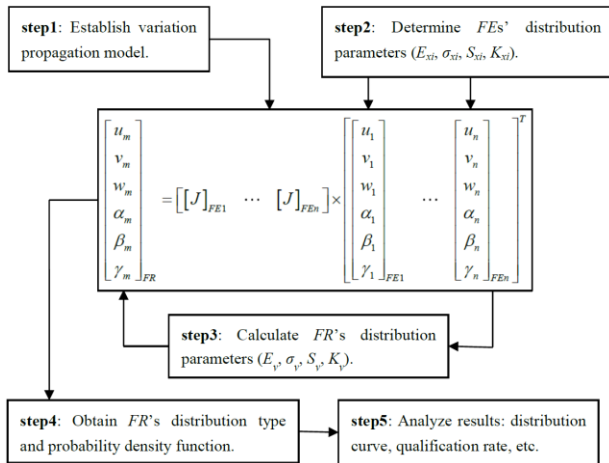
Figure. 1. Statistical variation analysis using Pearson distribution family.

4 Case study

To validate the practicability of the solution mentioned above, a realistic subassembly originating in an aero-engine is adopted for statistical variation analysis. The aero-engine subassembly is composed of four precise revolutionary parts and all the coordinate frames are specified in Fig.2 (a). Reference 0 is the global reference frame and the revolution axis is designated with a line passing through the base centre of the first component and is perpendicular to it.

(a)

(b)
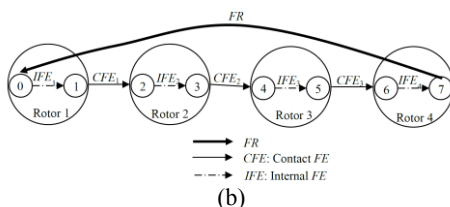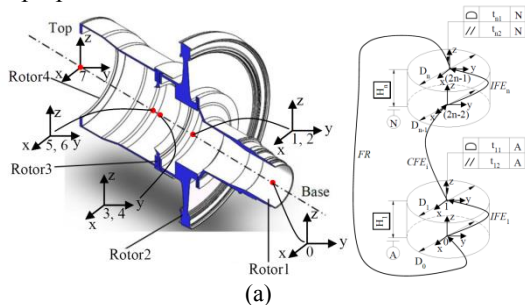
Figure. 2. Four-stage assembly: (a) parts drawing, (b) connection graph.

Each component's dimensions and composite tolerances which have been mentioned in Fig.2 (a) are listed in Table.2. The pose position and orientation of the final rotor which is accumulated by all variations of four components and three joints are very critical to concentricity control in aero-engine assembly technical requirement. Based on the 3D variation analysis method proposed in Section 3, the solving procedures are elaborated below:

Table 2. Dimensions and geometric tolerances of each component(Unit: mm).

Compo nent	Base diameter	Top diame ter	Hei ght	Profile tolerance	Parallel ism toleran ce
1	107	405	315	0.020	0.015
2	405	410	135	0.020	0.015
3	410	400	165	0.020	0.015
4	400	440	365	0.020	0.015

Step 1: Build coordinate system and assembly graph. In this case, the global coordinate system (0) is set in the base centre of the rotor 1, and other local reference frames are built in the centre of related contact areas. The connection graph is shown in Fig.2 (b). *FR* is the pose of the top rotor in relation to the *0th* frame. As can be seen, there is a serial route of variation propagation for *FR*: IFE_1-CFE_1-IFE_2-CFE_2-IFE_3-CFE_3-IFE_4.

Step 2: Establish the variation propagation model based on the above connection graph. Using the unified Jacobian-Torsor theory, the Jacobian matrixes and corresponding tolerancing torsors, as well as their variation ranges can be obtained. Details of FE_1 are presented in Table.3, and other FE_S have similar representations.

Table 3. Jacobian matrixes and tolerancing torsors.

FE_s	Jacobian matrixes	Tolerancing torsors	Variations
IFE_1	$\begin{bmatrix} 1 & 0 & 0 & 0 & 315 & 0 \\ 0 & 1 & 0 & -315 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	0
CFE_1	$\begin{bmatrix} 1 & 0 & 0 & 0 & 0.02 & 0 \\ 0 & 1 & 0 & -0.02 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ w_1 \\ \alpha_1 \\ \beta_1 \\ 0 \end{bmatrix}$	$-0.020/2 \le w_1 \le 0.020/2$ $-0.015/405 \le \alpha_1 \le 0.015/405$ $-0.015/405 \le \beta_1 \le 0.015/405$

Next, on the basis of Eq. (2), the final results can be obtained by multiplying the Jacobian matrix and torsor values. In this example, we only focus on the eccentricity in a certain direction, for instance the x axis direction, which can be extracted from u_m and reformulated as Eq. (12).

$$y_{u_4} = 0.02\beta_1 + 0.04\beta_2 + 0.06\beta_3 + 0.08\beta_4 \qquad (12)$$

In addition, as listed in Table.3, all the variations of effective vectors in the torsor should be satisfied.

Step 3: Calculate distribution parameters of the *FE*s and then determine *FR*'s distribution type and its PDF. The distribution parameters of *FR* can be calculated as follows:

$$E_y = \sum_{i=1}^{n} a_i E_{x_i} \qquad (13)$$

$$\sigma_y^2 = \sum_{i=1}^{n} a_i^2 \sigma_{x_i}^2 \qquad (14)$$

$$S_y = \frac{E\left[y - E(y)\right]^3}{\sigma_y^3} = \frac{\sum_{i=1}^{n} a_i^3 S_{x_i} \sigma_{x_i}^3}{\sigma_y^3} \qquad (15)$$

$$K_y = \frac{E\left[y - E(y)\right]^4}{\sigma_y^4} = (\sum_{i=1}^{n} a_i^4 K_{x_i} \sigma_{x_i}^4 + 6\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_i^2 a_j^2 \sigma_{x_i}^2 \sigma_{x_j}^2) / \sigma_y^4 \quad (16)$$

where a_i is the sensitivity coefficient, E_{xi} is the *i*th *FE*'s mathematical expectation, and E_y is the *FR*'s mathematical expectation. σ_{xi} is the *i*th *FE*'s standard deviation, and σ_y is the *FR*'s standard deviation. S_i denotes the coefficient of skewness and K_i denotes the coefficient of kurtosis.

To verify the influence of probability distribution types on the final assembly variations, we suppose the distribution types of *FE*s follow normal distribution, skewed distribution and uniform distribution respectively, and four probability combinations will be used as shown in Table.4, where skewed distribution is simulated by virtue of Beta approximation.

Table 4. Distribution type and parameters of the *FE*s.

N o.	Distribution pattern	Distribution parameter			
		β_1	β_2	β_3	β_4
1	Normal	$\mu = 0$ $\sigma = \dfrac{0.005}{405}$	$\mu = 0$ $\sigma = \dfrac{0.005}{410}$	$\mu = 0$ $\sigma = \dfrac{0.005}{400}$	$\mu = 0$ $\sigma = \dfrac{0.005}{440}$
2	Beta(Right-skewed)	$\alpha = 2$ $\beta = 5$	$\alpha = 3$ $\beta = 4$	$\alpha = 2$ $\beta = 8$	$\alpha = 1.5$ $\beta = 3.2$
3	Beta(Left-skewed)	$\alpha = 5$ $\beta = 2$	$\alpha = 4$ $\beta = 3$	$\alpha = 8$ $\beta = 2$	$\alpha = 3.2$ $\beta = 1.5$
4	Mixing(Normal, Beta, Uniform)	$\mu = 0$ $\sigma = \dfrac{0.005}{405}$	$\alpha = 5$ $\beta = 2$	$\alpha = 2$ $\beta = 5$	$a = \dfrac{-0.015}{440}$ $b = \dfrac{0.015}{440}$

The results of variation distribution parameters of the *FR* can be worked out according to Eq. (13) to (16), as listed in Table.5. Based on Eq. (5) to (10), we can further obtain the values of a_0, b_0, b_1, b_2, and deduce the variation distribution pattern by reference to Table.1, which has been shown in the last column of Table.5.

Table 5. Distribution type and parameters of the *FR*.

Parameter	Combination 1	Combination 2	Combination 3	Combination 4
Mean	0	0.0604	0.1396	0.0457
Variance	0.0125	0.0189	0.0189	0.0115

Skewness	0	0.3734	-0.3734	0.2417
Kurtosis	3	2.8210	2.8210	2.9311
Pattern	IV	VI	VI	VI

Step 4: Solve PDF using Pearson distribution family formulas (pattern IV and VI). With the aid of Matlab function, we can easily get the probability denisty of the Pearsons distribution with mean 'mu', standard deviation 'sigma', skewness 'skew' and kurtosis 'kurt', evaluated at the values in X.

Calculate the qualification rate of four combinations, comparing with the MC results. In order to balance accuracy with efficiency, the simulation number of MC is set to 10000 in this case. With regard to the tolerance limits, *FR* should fall in the qualified range of 0.038mm in accordance with engineering requirements. Results are shown in Table.6, and Fig.3 illustrates the distribution curve as well as MC statistical histograms.

Table 6. Qualification rate of four combinations.

Method	Combination 1	Combination 2	Combination 3	Combination 4
Proposed method	99.88%	11.59%	0.00%	26.28%
MC method	99.84%	11.55%	0.00%	26.26%
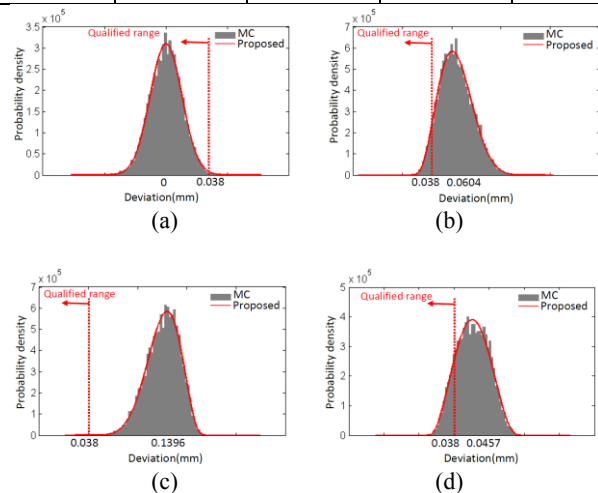
Figure. 3. Variation distribution curve and histogram: (a) combination 1, (b) combination 2, (c) combination 3, and (d) combination 4.

5 Comparison and discussion

By comparing results of the suggested method to the MC ones, it is observed that the probability density curves drawn by Pearson distribution family method agree with the generated histogram statistics using MC simulation fairly well, which have been illustrated in Fig.3 (a) to (d). Meanwhile in Table.6, the error between qualification rates produced by the two ways is less than 0.5 percent off. This demonstrates that the proposed method has the same accuracy of MC method. Additionally, the former will calculate only once in the entire simulation process while the latter will calculate ten thousand times repeatedly. Pearson distribution family method has vastly enhanced the work efficiency.

In addition, as can be seen from the diagram, when the distribution pattern of *FE*'s tolerance zone is changed, the shape of the variation distribution of the *FR* will bring about great changes, which appears to have happened at the assembly qualified rate as well. In the four groups of experiments, combination 1 (normal distribution) can guarantee the high pass percentage of 99.8% while combination 3 (left-skewed distribution) almost causes all products to be unqualified as a result of overlarge mean shift (=0.1396mm). All above results demonstrate that probability distribution types of the parts variations have a significant impact on the final assembling variation, and the traditional hypothesis of normal distribution cannot describe the true error conditions. Therefore, during the phase of product design, the practically processing technology and the level of processing must be taken into account to determine each component's real variation distribution. Only this can make results be more reasonable and reliable.

6 Conclusion

This work was attempting to propose a novel variation analysis method for complex assembly with multi joints. A calculating scheme was illustrated on a realistic case originating in aero-engine subassembly. Following conclusions could be drawn: first, for aero-engine rotors which contain lots of geometric tolerances, and planar and cylindrical features, Jacobian-Torsor theory can effectively solve the 3D variation modelling; second, compared to MC results with 10000 simulation times, Pearson distribution family method has the same accuracy, but much higher efficiency (runs only once); third, the result of *FR* depends on each *FE*'s probability distribution pattern significantly. In order to obtain more accurate predictions, the real distribution states of parts variations should be detected and used, which is beneficial to tolerance design and performance forecast.

Acknowledgments

References

1. Mantripragada R, Whitney DE. Modeling and controlling variation in mechanical assemblies using state transition models. *IEEE International Conference on Robotics and Automation, 1998 Proceedings*2002, p. 219-26 vol.1.

2. Thimm G, Britton GA, Cheong FS. Controlling Tolerence Stacks for Efficient Manufacturing. *Int J Adv Manuf Technol*. 2001; 18: 44-8.

3. Forouraghi B. Worst-Case Tolerance Design and Quality Assurance via Genetic Algorithms. *Journal of Optimization Theory & Applications*. 2002; 113: 251-68.

4. Fei C, Bai G. Extremum selection method of random variable for nonlinear dynamic reliability analysis of turbine blade deformation. *Propulsion & Power Research*. 2012; 1: 58-63.

5. Kim C-B, Pulat PS, Foote BL, Lee DH. Least cost tolerance allocation and bicriteria extension. *International Journal of Computer Integrated Manufacturing*. 1999; 12: 418-26.

6. Lin EE, Zhang HC. Theoretical Tolerance Stackup Analysis Based on Tolerance Zone Analysis. *International Journal of Advanced Manufacturing Technology*. 2001; 17: 257-62.

7. Ata MY. A convergence criterion for the Monte Carlo estimates. *Simulation Modelling Practice & Theory*. 2007; 15: 237-46.

8. Yang Z, Mcwilliam S, Popov AA, Hussain T. A probabilistic approach to variation propagation control for straight build in mechanical assembly. *International Journal of Advanced Manufacturing Technology*. 2013; 64: 1029-47.

9. Ghie, Walid. Statistical analysis tolerance using jacobian torsor model based on uncertainty propagation method. *International Journal of Multiphysics*. 2016; 3: 11-30.

10. Ahsanullah, Hamedani, G. G. Characterizations of certain continuous univariate distributions based on the conditional distribution of generalized order statistics. *Pakistan Journal of Statistics*. 2012; 28: 253-8.

11. Kharoufeh JP, Chandra MJ. Statistical tolerance analysis for non-normal or correlated normal component characteristics. *International Journal of Production Research*. 2002; 40: 337-52.

12. Provost SB. A Brief Derivation of the Asymptotic Distribution of Pearson's Statistic and an Accurate Approximation to Its Exact Distribution. 2016.

13. Desrochers A, Ghie W, Laperrière L. Application of a Unified Jacobian—Torsor Model for Tolerance Analysis. *Journal of Computing & Information Science in Engineering*. 2003; 3: 2-14.

14. Karpov IG, Zyryanov YT. On modified pearson distributions and their identification. *Automatic Control & Computer Sciences*. 2015; 49: 366-72.