Archives of Materials Science and Engineering

Volume 34
Issue 1
November 2008
Pages 48-51

International Scientific Journal
published monthly by the
World Academy of Materials
and Manufacturing Engineering

The curvilinear coordinates' approach to the smart-designs generation problem

J. Pietraszek*
Department of Mechanical Engineering, Cracow University of Technology,
Al. Jana Pawła II 37, 31-864 Kraków, Poland
* Corresponding author: E-mail address: pmpietra@mech.pk.edu.pl

Received 22.08.2008; published in revised form 01.11.2008

ABSTRACT

Purpose: The aim of this paper is to present an alternative approach to the problem of accuracy estimation for smart-designs' class (also named space-filling or intelligent design) by making use of the design's geometrical properties.

Design/methodology/approach: The assumed topographical condition: 'all design cases have to be on the defined surface' is reversed into a curvilinear coordinate system mapped on this surface. Then a sequence of irregular experimental designs is generated on the surface with various descriptive parameters and sampling from testing function is taken. Next, the identification of a general linear-quadratic model is conducted and various accuracy measures in comparison to a test function are calculated. At last the monotonic correlation analysis between accuracy measures and the design's geometrical properties is conducted.

Findings: Significant and strong correlation between the accuracy measures and some of the geometrical properties has been found.

Research limitations/implications: The correlations found are a strong suggestion for further research. The future investigations should be provided with various and more complicated testing functions and different topographical conditions. The relations between geometrical properties and accuracy measures need to be identified and their distributions and confidence intervals need to be determined.

Practical implications: The results obtained outline the method of the approximation accuracy estimation from geometrical properties of the design.

Originality/value: Worked out formulas may be of a significant value for those conducting data mining in technological data warehouses.

Keywords: Statistic methods; Design of experiment; Smart-design: Curvilinear coordinate system

METHODOLOGY OF RESEARCH, ANALYSIS AND MODELLING

1. Introduction

A typical linear-quadratic model with interactions of a static system being investigated may be formulated [1–4] as:

$$z_u = b_0 + \sum_{k=1}^{i} b_k x_{k,u} + \sum_{k=1}^{i} b_{kk} x_{kk,u}^2 + \sum_{k=1}^{i-1} \sum_{j=k+1}^{i} b_{kj} x_{k,u} x_{j,u} + e_u \qquad (1)$$

where b_0, b_k, b_{kk}, b_{kj} – parameters of the model, i – number of input factors, u – number of data points (design cases) under consideration, e_u – random error. In the model, the uth observation consists of two components: the deterministic four terms of the formula (1) and e_u, the random noise in the uth observation. Following assumption are made about the errors e_u: they have zero expectations, they are uncorrelated, they have homogeneous variance. The unknown parameters b are estimated by the method of least squares. The matrix notation is customarily used for the model (1):

$$\mathbf{Z} = \mathbf{X}\mathbf{b} + \mathbf{e} \quad (2)$$

where **Z** – vector of observations, **b** – vector of parameters, **X** – the design matrix, **e** – vector of random errors. The model (2) is said to be linear because it expresses outcome **Z** as a linear combination of the parameter vector **b**. The main purpose is to obtain estimates of the parameter vector **b**. They may be obtained as a solution of so called normal equations:

$$\mathbf{X}^T \mathbf{Z} = \mathbf{X}^T \mathbf{X}\mathbf{b} \quad (3)$$

Around so formulated model (2) exemplary data point tables called experimental design are constructed. The construction methods are focused on special properties related to the model. The design geometry is derived from statistical properties demanded for the designed experiments. It is not possible to apply this approach in passive experiments. Recently passive experiments occur more frequently due to easier collecting possibilities. In that case important reasons occur to ask the question: *Are design's geometrical properties a proper base to estimate the statistical properties of forecasted outcome*? It is an inverse problem in comparison to the approach applied in the designed experiment. The problem has been investigated recently [5-9]. In this paper the problem is being investigated for irregular topographical designs [5, 10].

2. Description of the approach

2.1. The object of investigation

A hip joint endoprosthesis is the investigated object [11]. Geometrically it is a sphere without a sector around its lower pole. Input variables are the measure point's coordinates, while the output variable is the micro hardness HV100 in this point. The initial aim of the investigation was micro hardness mapping on the endoprosthesis surface.

2.2. Coordinate system

Every experimental design appropriated for the investigation has to be topographical by a definition: all measures have to be located on the endoprosthesis surface. On the endoprosthesis surface a curvilinear coordinate system is introduced. In that way a topographical condition is built into coordinate system geometry and may be omitted in subsequent considerations. A spherical coordinate system with a zenith angle and a longitude was selected. The zenith angle changes from 0 to 120 degree. The longitude changes from 0 to 360 degree.

2.3. Test function

This investigation is the first trial to generate an irregular smart design in the curvilinear coordination system. Due to start-up difficulties a simple model is selected as a test function [11]:

$$z = 584{,}10 - 3{,}01 x_1 + 0{,}0147 x_1^2 \quad (4)$$

where z – micro hardness HV100, x_1 – zenith angle [deg]. It allows to focus on relations between geometrical properties and accuracy measures without taking a too complicated model into considerations. In the further investigations more complicated functions should be considered e.g. mentioned in [12]. There is no dependency of model (4) on the longitude x_2 because its effect is not significant. The random error related to the model (4) may be described by a normal distribution [11]:

$$\varepsilon = N(\mu = 0; \sigma = 35{,}7) \quad (5)$$

homogeneous in the whole design space. At last the test function being sampled is the sum of the deterministic part (4) and random noise (5):

$$\hat{z}(x_1, x_2) = z(x_1, x_2) + \varepsilon(x_1, x_2)\ . \quad (6)$$

2.4. Design generator

A constructive method [10] was selected for an irregular design generating process. The probability of a case selection in an unit of a design space volume is assumed as a constant. Due to the design space curvature the traditional cumulative distribution formula:

$$\mathrm{P}(\mathrm{X}_1 < \alpha, \mathrm{X}_2 < \varphi) = \frac{1}{\mathrm{V}} \int_{\substack{\mathrm{X}_1 < \alpha \\ \mathrm{X}_2 < \varphi}} \mathrm{dV} \quad (7)$$

has to be modified by specific formulas:

$$\mathrm{V} = 3\pi$$
$$\mathrm{dV} = \sin x_1 \cdot \mathrm{d}x_1 \cdot \mathrm{d}x_2 \quad (8)$$

At last, the cumulative distribution, since the coordinates are randomly independent, is formulated as:

$$\mathrm{P}(\mathrm{X}_1 < \alpha, \mathrm{X}_2 < \varphi) = \mathrm{P}_\alpha(\mathrm{X}_1 < \alpha) \cdot \mathrm{P}_\varphi(\mathrm{X}_2 < \varphi)$$
$$\mathrm{P}_\alpha(\mathrm{X}_1 < \alpha) = \frac{2}{3}(1 - \cos\alpha)$$
$$\mathrm{P}_\varphi(\mathrm{X}_2 < \varphi) = \frac{\varphi}{2\pi} \quad (9)$$

Such formulas are exceptionally convenient because they allow to generate each coordinate separately. An uniform random generator on interval [0, 1] is involved and so is the method of inverted cumulative distribution [13–15]. A rectangular 120 deg x 360 deg is assumed as the design space. The minimum and maximum distance to the nearest neighbour are assigned as the start-up parameters. Next case searching process is broken after 1000 failed trials. This causes an obvious fluctuation of the number of cases.

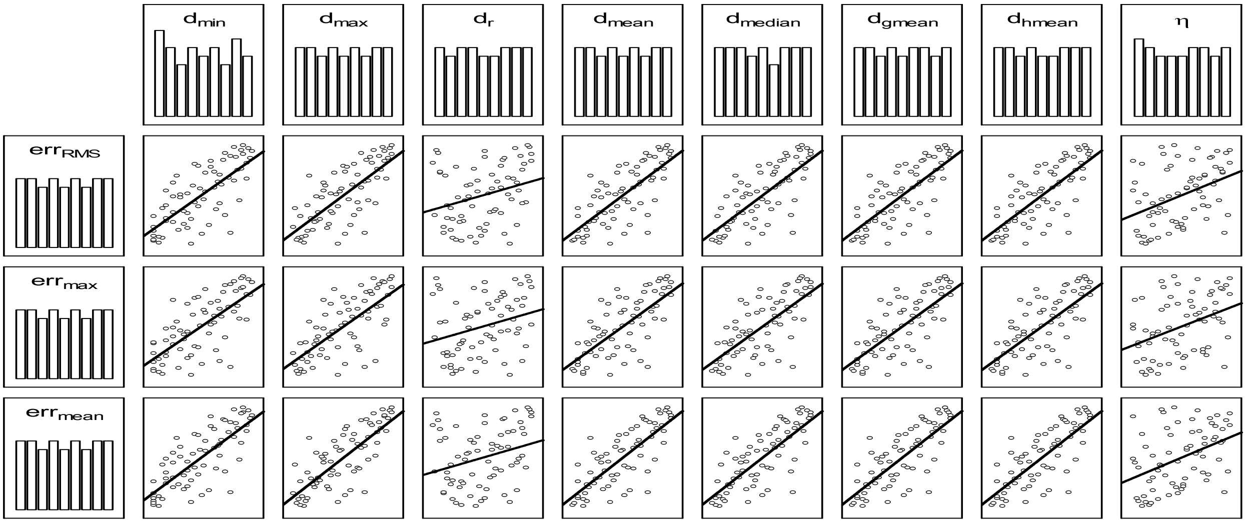

Fig.1. Matrix scatter plot for ranked data (Spearman correlation coefficient)

Table 1.
Spearman correlation between geometrical properties and accuracy measures

Geometrical property	Accuracy measure		
	err_{RMS}	err_{max}	err_{mean}
d_{min}	0.73	0.71	0.77
d_{max}	0.74	0.70	0.78
d_r	0.29	0.29	0.29
d_{mean}	0.75	0.72	0.78
d_{median}	0.75	0.71	0.78
d_{gmean}	0.75	0.72	0.78
d_{hmean}	0.75	0.72	0.78
η	0.41	0.39	0.42

2.5. Design geometrical properties

Eight geometrical properties are identified for each generated design. All properties are based on the distance to the nearest neighbour. They are: the minimum distance d_{min}, the maximum distance d_{max}, the range d_r between d_{min} and d_{max}, the mean distance d_{mean}, the median of distances d_{median}, the geometrical mean d_{gmean}, the harmonical mean d_{hmean}, the homogeneous measure η [16] calculated as quotient d_{min}/d_{max}. The distance is defined as the length of the great circle arc (orthodrome).

2.6. Model identified

A linear-quadratic model with interactions:

$$z = b_0 + b_1 x_1 + b_2 x_2 + b_{11} x_1^2 + b_{12} x_1 x_2 + b_{22} x_2^2 + \varepsilon \qquad (10)$$

is identified from the sampled data. The random noise (5) contained in the sampled data perturbs the subsequent realisations of the identified model (10). This causes differences between the model (10) forecasts and the test function's (4) outcome. The differences (errors) are aggregated in the accuracy measures and analysed.

2.7. Approximation accuracy measures

The approximation accuracy is calculated from the whole design space, not only from the design cases being the design's support. This is possible due to the fact, that the test function is known. Three accuracy measures are introduced: the maximum error err_{max}, the mean error err_{mean} and the root mean square error err_{RMS}.

3. Description of achieved results

The simulation for 7 groups of 10 designs each were conducted. Each group had its own minimum distance to the nearest neighbour parameter assigned: 0.30 ; 0.40 ; 0.50 ; 0.60 ; 0.70 ; 0.80 ; 0.90. The smaller values leaded to a too large design, the greater – too sparse. The results obtained were tabularised and analysed. The monotonic correlation analysis based on Spearman's correlation coefficient [15, 17] was conducted (Fig.1, Table 1). Technically, the geometrical properties and accuracy measures obtained from simulations were ranked. In the case of the same data, the rank is calculated as an average from the whole subgroup. In this approach the Spearman's correlation coefficient is calculated as the Pearson linear correlation coefficient for ranked data.

All correlation coefficient values obtained (Table 1) were statistically significant at the assumed significance level of 0.05. As shown in Table 1, geometrical properties d_r and η are rather useless for the accuracy forecasting. In contrast, remaining properties have rather large correlation coefficient values, which are greater than 0.70. Data shown in Table 1 reveals that potentially the most fruitful pairs are: err_{RMS} – d_{mean}, d_{median}, d_{gmean}, d_{hmean} ; err_{max} – d_{mean}, d_{gmean}, d_{hmean} ; err_{mean} – d_{max}, d_{mean},

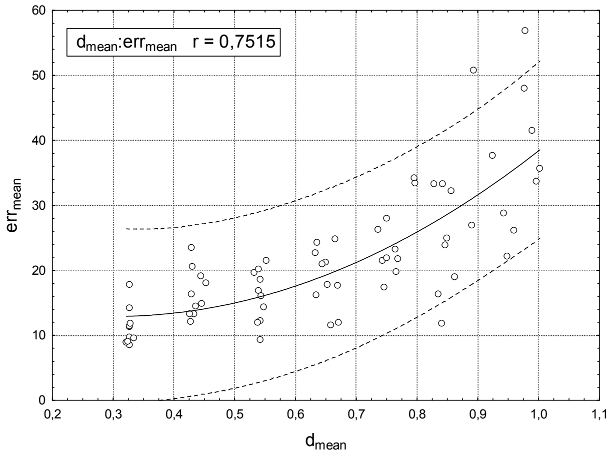

Fig. 2. Scatter plot for err_{mean} vs. d_{mean}

d_{median}, d_{gmean}, d_{hmean}. A weak correlation (Table 1) and a very large scatter (Fig.1) for the homogeneous property η is a surprise. A stronger correlation was expected for this property. In contrast, a weak correlation (Table 1) and a very large scatter (Fig.1) for d_r is not a surprise because this property as a measure is very sensitive for data changes. On the Fig.2 an exemplary dependency between err_{mean} and d_{mean} is presented. The continuous line is a second-order polynomial regression, dashed-lines are 95% confidence spreads. Spreads are calculated with a normal distribution assumption. The data normality should be precisely verified in further investigations or other distribution should be identified.

4. Conclusions

An inverse problem of the accuracy estimation from measurable geometrical properties is presented in this paper. Micro hardness data collected from a surface of a hip joint endoprosthesis was analysed. An investigated object's geometry imposed topographical condition: all experimental design cases had to be located on the object surface. A curvilinear coordination system consistent with the object geometry was introduced. In this way the topographical condition was built into coordination system geometry and it did not have to be explicitly considered. An irregular smart design random generator based on differential geometry elements was applied. The test function sampling was conducted and a linear-quadratic model was identified, basing on sampled data obtained. A data scatter appeared due to a random noise perturbing the sampled data. A monotonic correlation analysis between the accuracy measures obtained and the designs' geometrical properties was conducted. Significant and rather strong correlations between the accuracy measures and some of the designs' geometrical properties were detected. The results obtained are presented in tabular and graphical form. They strongly suggest that it is possible to forecast the accuracy of the irregular design approximation from the design geometry.

References

[1] Z. Polański, Design of experiment in technology, PWN, Warsaw, 1984 (in Polish).

[2] D.C. Montgomery, R.H. Myers, Response surface methodology. Process and product optimization using designed experiments, Wiley, New York, 1995.

[3] P.W.M. John, Statistical design and analysis of experiments, SIAM, Philadelphia, 1998.

[4] D.C. Montgomery, Design and analysis of experiments, Wiley, New York, 2004.

[5] R. Górecka, Z. Polański, Flexible experimental designs integrated with intelligent computational system, In: Computer aided metrology, Cz. Przybysz et al. (eds.), 2 (2001) 17-22 (in Polish).

[6] J. Pietraszek, Response surface methodology at irregular grids based on voronoi scheme with neural network approximator. In: Neural Networks and Soft Computing, Rutkowski L., Kacprzyk J. (eds.), Springer-Physica Verlag, Heidelberg, 2003, 250-255.

[7] J. Pietraszek, Analysis of two-factor experimental design over irregular grid generated from Voronoi scheme with ANN approximator, Journal of Surface Mining 2-3 (2003) 21-24 (in Polish).

[8] A. Skowronek, Analysis of flexible experimental design generation process, Technical Journal **7** (2007) 63-74 (in Polish).

[9] A. Skowronek, Optimization the flexible design of experiment, Proceedings of the International Conference "Engineering Optimization EngOpt'2008", Rio de Janeiro, 2008, 70-71.

[10] Z. Polański, Empirical research – methodology and computer aiding, In: Modern metrology, J. Barzykowski (ed.), WNT, Warsaw, 2004, 124-216 (in Polish).

[11] R. Dwornicka, J. Pietraszek, Identification of function modelling hardness on the surface of hip joint endoprosthesis, In: The improvement of the quality, reliability and long usage of technical systems and technological processes, A.T. Boroszow, A. Bubulis et al. (eds.), National Council of Ukraine for Mechanism and Machine Design, IFToMM Ukrainian Committee, Khmelnitsky, 2007, 96-100.

[12] Z. Michalewicz, Genetic Algorithms + Data Structures = Evolution Programs, Springer-Verlag, Heidelberg, 1996.

[13] R. Wieczorkowski, R. Zieliński, Computer generators of random numbers, WNT, Warsaw, 1997 (in Polish).

[14] A. Plucińska, E. Pluciński, Probability, WNT, Warszawa, 2000 (in Polish).

[15] J. Koronacki, J. Mielniczuk, Statistics for technical students, WNT, Warsaw, 2001 (in Polish).

[16] J. Pietraszek, The criterion of homogeneity for space filling experimental design, In: The improvement of the quality, reliability and long usage of technical systems and technological processes, Boroszow A.T., Bubulis A. et al. (eds.), National Council of Ukraine for Mechanism and Machine Design, IFToMM Ukrainian Committee, Khmelnitsky, 2006.

[17] P. Grzegorzewski, Decision support at uncertainty. Statistical methods for imprecise data, EXIT Press, Warsaw, 2006 (in Polish).