e-ISSN 1607-3274 Радіоелектроніка, інформатика, управління. 2019. № 1
p-ISSN 2313-688X Radio Electronics, Computer Science, Control. 2019. № 1

UDC 004.94:658

THE METHOD OF MULTIVARIATE STATISTICAL ANALYSIS OF THE TIME MULTIVARIATE CRITICAL QUALITY ATTRIBUTES OF MANUFACTURE PROCESS WITH THE DATA FACTORIZATION

Havrylko Ye. V. – Dr. Sc., Associate Professor, Professor of Computer Sciences Department, State University of Telecommunications, Kyiv, Ukraine.

Kurchenko O. A. – PhD, Associate Professor, Head of the Department of Management of Information and Cyber Security, State University of Telecommunications, Kyiv, Ukraine.

Tereshchenko I. V. – PhD, Associate professor of the Department of info-communication engineering, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Tereshchenko A. I. – Post-graduate student of the Department of Management of Information and Cyber Security, State University of Telecommunications, Kyiv, Ukraine.

ABSTRACT

Context. This paper presents a method for solving the problem of product's quality assurance at the stage of the initial manufacture process design in accordance with the process-analytical technology for the design of modern certified manufacturing – QbD. The method uses the information technologies of multivariate statistical analysis (MSA) to evaluate the influence of time multivariate critical process parameters (CPPs) on the time product critical quality attributes (CQAs). Preparatory transformation of clusters of critical process (manufacture process) parameters into factors of product critical quality attributes was carried out.

Objective. To disclose the method of multivariate statistical analysis for assessing the character and features of the influence of time multivariate critical process parameters on time multivariate critical quality attributes at the design stage of the manufacture process.

Method. The method consistently uses: statistical procedures of exploratory multivariate data analysis; transformation the homogeneous observed values matrices of CPPs and product CQAs into data frame (table) with factorized data; construction the regression trees of multivariate CPPs with a multivariate responses (CQAs). The method is implemented the R language packages software.

Results. Factorized time multivariate CPPs make it possible to use methods of multivariate statistical analysis for evaluating the influence of CPPs factors on the time multivariate CQAs.

Conclusions. This method of statistical analysis, together with statistical multivariate canonical analysis, represents an up-to-date information technology for detailed estimation the influence of time multivariate CPPs objects and some CPPs components on CQAs.

KEYWORDS: quality-by-design, critical quality attributes, critical process parameters, design of experiment, multivariate statistical analysis.

ABBREVIATIONS

CCA is a canonical correspondence analysis;
CMAs is a critical material attributes;
CPPs is a critical process parameters;
CQAs is a critical quality attributes;
DoE is a design of experiment;
MBSE is a model based systems engineering;
MRT is a multivariate regression trees;
MSA is a multivariate statistical analysis;
NDI is a non-development item;
NMDS is a nonmetric multidimensional scaling;
PAT is a process analytical technology;
PCA is a principal component analysis;
PDCA is a plan-do-check-act cycle;
QbD is a quality-by-design;
QFD is a quality function deployment;
RDA is a redundancy analysis;
SEMMA is a sample, explore, modify, model, assess
TPQP is a target product quality profile;
VAT is a visual assessment of cluster tendency;
VS is a variables selection.

NOMENCLATURE

X is a observations array of random multivariate manufacture process variables (CPPs observations);

Y is a observations array of random multivariate manufacture responses (CQAs observations);

X^j is a m-dimensional vector of the j-th ($j = 1, 2, \ldots, n$) observation of m random CPPs;

Y^j is a k-dimensional vector of the j-th ($j = 1, 2, \ldots, n$) observation of k random CQAs;

x_m is a m-th component of X^j, $X^j = \{x_1, ..., x_m\}$;

y_k is a k-th component of Y^j, $Y^j = \{y_1, ..., y_k\}$;

X_v^j is a point in the s-dimensional space of attributes: $X_1^j, X_2^j, ..., X_s^j$;

X_v^q is a $q \times m$ ($q = 1, 2, ..., n$) sample from the multivariate population at the v-th attribute.

© Havrylko Ye. V., Kurchenko O. A., Tereshchenko A. I., 2019
DOI 10.15588/1607-3274-2019-1-16

INTRODUCTION

Market conjuncture and increased rivalry objectivity set the task of developing and improving methods for ensuring high quality products in stage of the initial manufacture process design with further management of the entire production cycle (PDCA, [1]) in accordance with the TPQP [2].

One of the actual ways to solve this problem is the QbD methodology [2, 3]. This methodology shifts the emphasis of product quality assurance towards information technology of statistical research and attentive experiment planning at the stage of product development and manufacture process design. The QbD concept is a significant transformation in product quality regulation from an empirical process to a more scientific and risk-based approach. The QbD concept combines categories under the general term "critical quality attributes": CMAs, CPPs and CQAs with TPQP. In turn, TPQP determines the quantitative parameters of the product safety and efficiency [2] and allows them to be compared with the requirements of standards, as shown in Fig. 1.

The dependence of product CQAs on CPPs and CQAs is considered as a quality function: $CQAs = F(CPPs)$. This quality function is structured according to the main stages of the QFD [4] manufacture process, for which critical quality attributes are determine and QbD principles are applied, as shown in Fig. 1.

As an important tool for QbD, PAT [5, 6] is considered, which actively uses MSA information technologies to provide and maintain critical quality attributes within the standards requirements in the design space. The results of the data multivariate statistical analysis of the multifactor experiment at the design stage make it possible to establish the permissible variability ranges of CPPs and CMAs based on the extent of influence on CQAs.

It should be noted that with the strict requirements of standards (ISO 9001 [7], ISO 22000 [8]) for CMAs, the influence of CPPs on CQAs is more often considered [5]. In this case, the time multivariate statistical data CPPs generally have a group/object influence on CQAs and therefore need a software factorization of the time multivariate computer format data.

The object of the study is the process of product's quality assurance at the stage of the initial manufacture process design.

In accordance with the abovementioned, the task of products quality assurance at the stage of the initial manufacture process design is concretized as the definition of CPPs and their objects, which significantly affect on CQAs.

The solution of this problem is proposed to implement by consistently application of the following MSA methods:

– Defining the structure and hierarchy of time multivariate data: CPPs and CQAs;

– Defining of qualitative and quantitative measures of relations between the formed objects (factors) CPPs and CQAs;

– Comparison of the results of factor and component analysis of the influence of time multivariate CPPs on time multivariate CQAs.

Together with computer-intensive information technologies of statistical canonical analysis [9], the factorized multivariate data of critical quality attributes provide additional principal opportunities for multivariate statistical analysis of the affect of CPPs on CQAs.

The method of Genichi Taguchi is known for solving this problem by fractional and factored planning of the experiment with subjective test sets of design parameters and noise factors. These data form the matrix experiments using orthogonal arrays [10]. According to the effects of noise factors, orthogonal design parameters are evaluated for robustness and analyze possible losses of the quadratic quality function.

In contrast to this decision, it was suggested to apply MSA and computer-intensive information technologies for carrying out a statistical full-factor experiment with direct factorization of each of the observation tables after the exploratory data analysis. At the same time, statistical estimates of multivariate factor analysis are adequately supplemented by estimates of the component affect of CPPs on CQAs.

The subjects of studies are informational technologies for estimating the factor influence of manufacture CPPs on product CQAs.

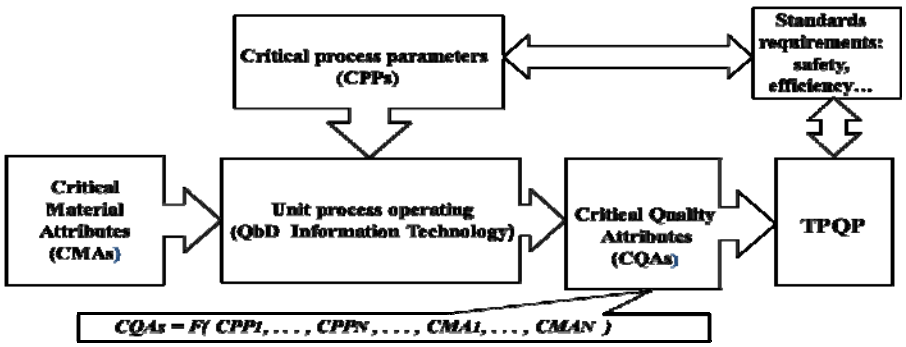


Figure 1 – The relationship between categories of critical quality attributes and TPQP according to the QbD concept

© Havrylko Ye. V., Kurchenko O. A., Tereshchenko A. I., 2019
DOI 10.15588/1607-3274-2019-1-16

The proposed method of MSA of the time multivariate critical quality attributes of manufacture process with the data factorization is should be considered as the evolution of the QbD approach of product quality assurance at the design development stage. The QbD concept and method are aimed at an experimental study of the influence degree of CPP*s* on CQAs.

1 PROBLEM STATEMENT

The statistics of the experiment operates with data from object-oriented *n* observations formed into two arrays of random multivariate variables *X* and *Y,* $X \to Y$.

A random variable *X* potentially determines the properties of the random variable *Y*: $X \supset \left\{X^j\right\}$, $Y \supset \left\{Y^j\right\}$. The category $X \to Y$ is described by the quality function of the manufacture process:

$$CQAs = F(CPP_i), CQAs \supset \left\{CQA_k\right\}.$$

Formally, the results of monitoring (sampling) of random multivariate values of *X* and *Y* are data of the QbD development space, but do not represent a categorized information space of relations $X \to Y$.

To determine and characterize the causal relations in the information space of the variables *X* and *Y*, it is necessary to inquire into the structure and hierarchy of observational data *X* and *Y*, as well as the qualitative and quantitative measures of the relations between them.

The problem of solving this task is that empirical arrays of *X* and *Y* data have a computer format in the form of time interrelated matrices that are not ready for issue. To conduct multivariate statistical analysis, the *X* and *Y* data arrays must undergo a preparatory processing step in the form of statistical procedures of exploratory data analysis, clustering, factorization and interpretation of the results [11]. In practice, observations of exogenous variables are not offline, but are groups of related CPPs. These groups form clusters with an interpreted dependence of variables for which it is of interest to estimate the influence on the multivariate responses of the manufacture process (Y^j). It is suggested to cluster the time random multivariate data CPPs (X^j) and identify them with the influence factors on the multivariate responses of the manufacture process CQAs (Y^j).

These operations make it possible to determine the potential trends and relationships between factorized observations of variables *X* and an array of multivariate responses *Y*, and also allow MSA to be used to research dependence $X \to Y$ as a quality function: $CQAs = F\left(CPP_i\right)$.

2 REVIEW OF THE LITERATURE

The QFD process methodology [4, 12], the methods of assurance the TPQP with the detailed QbD [2, 5, 13] and the risk-oriented process approach [4, 14, 3] – are the dominants of modern quality management systems [7].

High requirements to product quality and market competition – lead to the need of products quality assurance at the QbD development stage [13, 5] at an decomposition of the quality function (QFD) [4]. An adequate tool for QbD are the PAT methods [6, 15], which are based on information MSA technologies.

At the design stage, the PAT methods are focused on the research of multifactorial relationships between CPPs and CQAs, exogenous factors and their influence on the on the final product quality [12, 16].

In [17], an actual systematic review of a large number of references on the topics of applications of MSA methods is given. The known methods for determining the nominal values of CQAs, CPPs and permissible variation for them offer statistical solutions to the multivariate VS. The VS statistical methods (essentially MSA methods) research the contribution of exogenous factors to the variability of the quality function values (a multivariate endogenous variable).

Thus, MSA methods create an adequate reasoned basis for transformation functional requirements into design parameters and for justification of design parameters permissible variation [18]. The adequate CPPs and CMAs are determined on the basis of an reasonable (standards-oriented) change in the characteristics of the quality function and the sensitivity of this function to critical quality attributes.

This problem can be considered from the position of system engineering [19, 20, 21], when models play a key role at one or several stages of the development process and are part of the process design by the MBSE [22]. The main stages of research and design of the product/process from the general requirements to the system [21] within the background of robust design [18] inevitably lead to the architecture (structure + logic) of the experiment [16, 21, 23]. One of the general trends of this logic is the decomposition of the quality function (QFD) [3, 4, 18] using the DoE technology [19], similar to the methodology of careful experiment planning (QbD) [2, 3, 5] with PAT tools [6].

It is also important to note one of the features of modern process design, which is the use of NDI technologies (for example, software packages of MSA methods on R language) [24].

At the design and experimentation level, this direction is getting widespread use in connection with the intensive introduction of information technologies of object-oriented MSA and the high confidence of engineers in the developed and tested software packages.

Thus, the review of literature enables to emphasize the following features of the task:

– the definition of the requirements for methods of technological approximation to the specified values of CQAs attributes, as a some quality function of the CPP*s* values, is made at the stage of the initial manufacture process design;

© Havrylko Ye. V., Kurchenko O. A., Tereshchenko A. I., 2019
DOI 10.15588/1607-3274-2019-1-16

– the guaranteed and certified product quality at the hierarchical level of the initial manufacture process design is achieved by computer-intensive information MSA technologies of manufacture process and products in the multivariate space of design parameters.

3 MATERIALS AND METHODS

To ensure the standardized quality of products on the initial manufacture process design, the task is to determine from the data of *n* observations the *m* CPPs – $\{x_1,...,x_m\}$ and their objects (factors) that significantly affect on *k* CQAs of product – $\{y_1,...,y_k\}$ multivariate process responses.

The cluster analysis methods combine samples X^j of the observable data set into groups of objects that are in some sense homogeneous and called clusters or classes, as shown in Fig. 2.

Suppose that each of the observations X^j is given as a point – X_v^j in the *s*-dimensional space of attributes: $X_1^j, X_2^j, ..., X_s^j$. The set of these points can be treated as a sample $q \times m$ – X_v^q ($q = 1, 2, ..., n$ – the number of observations corresponding to the *v*-th attribute) from the multivariate population at the *v*-th attribute.

A number of observations of the *q*-th attribute form an object (cluster), which is a factor of the homogeneous affect of *q* observations on CQAs. Some observations Y^j are grouped into objects by this attribute by conducting a constrained ordination. Based on the results of such an analysis, a diagram is constructed – an "ordination three plot", which shows the variability of the CQAs with relation to the two CCA1-CCA2 (canonical correspondence axis) projection axes and the statistical relationships with each critical process parameter – $\{x_1,...,x_m\}$.

The research used partitioning algorithms of the R language that decompose the data set of *n* observations into *s* clusters with previously unknown parameters. This algorithms searches for centroids, i.e. the centers of points concentration with the minimum scatter within each cluster as far as possible from each other and with the verification of clustering quality [11].

The interrelation between the components of multivariate data of *X* and *Y* arrays was established under the multivariate statistical canonical analysis with the means of software packages of the R language [11]. In particular, the mathematical apparatus of multivariate canonical ordination was used: RDA, CCA and NMDS.

These methods are positioned as an extension of multivariate statistical regression analysis in modeling the multivariate response of the manufacture process (Y^j) to the multivariate value of the critical process parameter (X^j).

Thus, the sequence of performing statistical procedures for clustering and factorization of multivariate data CPPs and CQAs creates a model of a statistical full-factor experiment that is realized by methods of multivariate statistical causal analysis.

4 EXPERIMENTS

In the case under consideration, the experiment assumes the study of cause-effect or causal dependencies. To determine the causal relationships of multivariate X^j and Y^j variables, three conditions must be fulfilled [25]:

1. The events and facts related with a X^j variable must precede events and facts related with a Y^j variable.

2. The variation of the X^j variable must be accompanied by a variation of the Y^j variable.

3. The influence of the third variable, which implies an alternative explanation to empirical observations and a statistically confirmed relationship between X^j and Y^j should be excluded.

Thus, is proposed a scheme of the experiment in the form of a sequence of actions above multivariate observational data as critical attributes of the quality of the manufacture process:

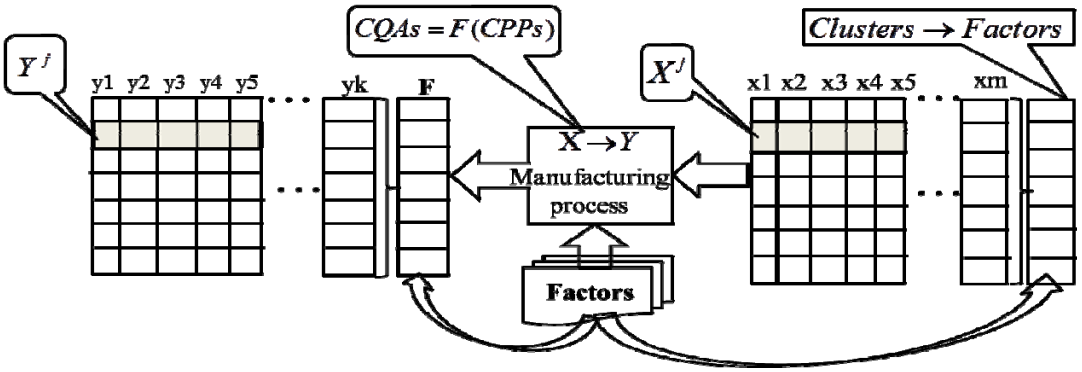


Figure 2 – Formation of factors of time multivariate data X^j and Y^j

© Havrylko Ye. V., Kurchenko O. A., Tereshchenko A. I., 2019
DOI 10.15588/1607-3274-2019-1-16

1. The carrying out an exploratory analysis of exogenous variables data: determination of correlated variables, estimation of clustering tendency.

2. The carrying out of clustering of the constrained values (X^j) and quality estimation of this clustering.

3. The transformation of random multivariate homogeneous arrays (data frames) of variables X^j and Y^j into data frames with factor variables.

4. The carrying out and estimating of unconstrained ordination of responses Y^j and constrained ordination of responses Y^j by factors of exogenous variables X^j.

4a. The comparison of the results of unconstrained and constrained ordination of multivariate responses Y^j (without and with affect the factors of exogenous variables X^j).

4b. The comparison of estimates the constrained ordination of the responses Y^j with the canonical analysis by CCA and NMDS methods.

4c. The formation of the MRT of an exogenous *X* variables array with a multivariate responses *Y*.

5. The estimation the MSA results of the relations between an array of exogenous variables *X* and an array of multivariate responses *Y*.

In general, this algorithm corresponds to the SEMMA methodology [26] for determining physical regularities at statistical analysis. SEMMA focuses on structuring processes (such as QFD), as well as applying statistical analysis and visualization of results at each stage of the design process and gives the researcher the freedom to choose the concept and implementation of explore.

5 RESULTS

The results correspond to the multivariate statistical analysis of time-data arrays of predictors *X* and responses *Y* specified in the form of uniformly distributed random values of the design space within the ranges specified in standards.

The exploratory data analysis was used for descriptive statistics of sample properties. This analysis included: processing and systematization of empirical data, visual presentation of multivariate data in the form of tables and graphs, as well as quantitative characteristics of the basic statistical parameters. Estimation of the weight or variable importance of predictors X^j and their tendency to clustering, taking into account the low multicollinearity, was evaluated using the principle components analysis and VAT, as shown in Fig. 3a.

The search for an optimal scheme for combining predictors into clusters was performed by permutation of multisets at various combinations of the groups number, distances metrics and clustering methods using 30 quality parameters (indices).

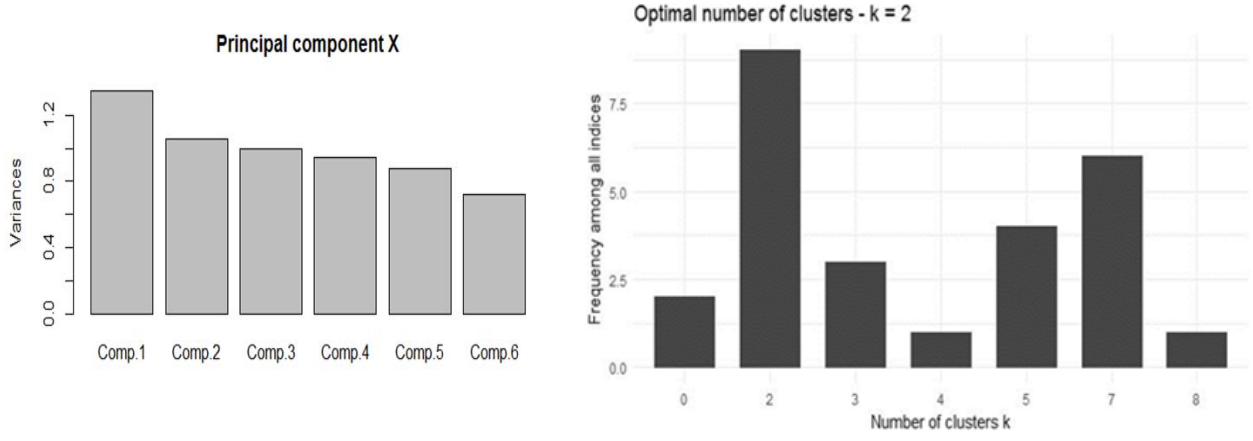
Based on the results of the calculations, the optimal number of clusters was chosen, as shown in Fig. 3b.

The operation of transforming homogeneous arrays of numerical variables *X* and *Y* into tables with factor (cluster) attributes was performed by identifying the factor with the cluster of observations X^j and corresponding group of multivariate responses Y^j.

As a result, tables (data.frame) with factorized data arrays *X* and *Y* were formed, as shown in Fig. 2.

Unlike unconstrained ordination, constrained ordination allows to connect variability of multivariate values of responses Y^j with affect of predictors X^j.

In Fig. 4 shows the ordination of the multivariate response components *Y* on the principal components plane (unconstrained ordination) and the distribution of these values over the objects of the *X* array factors on the principal components plane (constrained ordination).



a b

Figure 3 – The result of estimating the weight of the predictor components X^j – a. The result of determining the optimal number of clusters according to the estimates of different indices – b

© Havrylko Ye. V., Kurchenko O. A., Tereshchenko A. I., 2019
DOI 10.15588/1607-3274-2019-1-16

An unconstrained ordination analysis of responses matrix Y makes it possible to obtain a mapping in the orthogonal coordinate system of only some particular structural features of the research data array in the form of graphic projections of the components of multivariate response Y^j on the principal components plane, as shown in Fig. 4a.

In the diagram of Fig. 4b shows the grouping of observations by factor characteristics on the background of the explanatory components x_m (dashed lines) associated with the eigenvectors of the X array (ordination three plot).

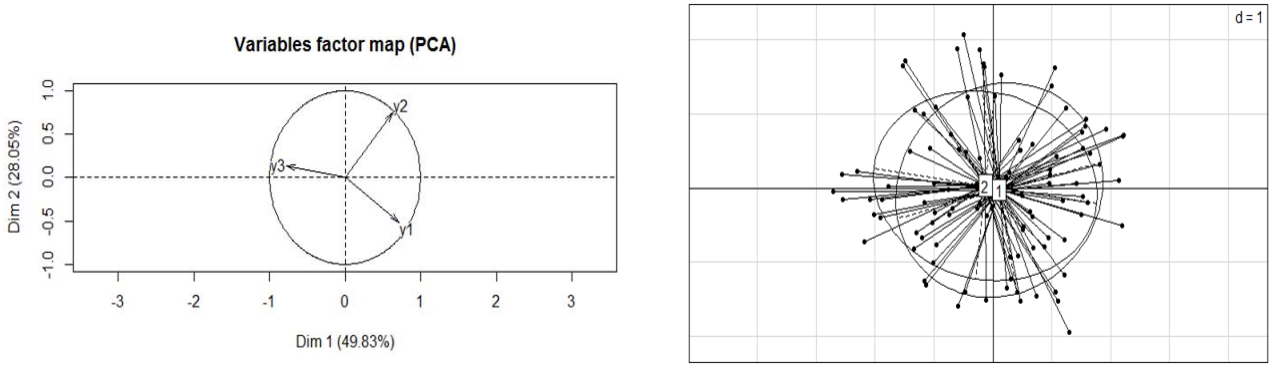
Note, that the use of PCA-ordinance is a method of smoothing random data fluctuations and contributes to the construction of more stable (robust) cluster structures. The practice of direct ordination allows combining the operations of hierarchical formation of objects on the separating predictors of the X array with the projection of multivariate data of responses array Y into the space of the principal components.

To assess the component relationship of two arrays of multivariate variables X and Y, used statistical canonical ordination procedures: CCA and NMDS.

To represent the results of NMDS, when the least distorted distance metric between the two-dimensional projections of the Y^j values on the principal component axis is determined. This distance metric is determined by the Spearman correlation coefficient for each of the variants of the distance metric in the Y^j response spaces and predictors X^j.

The NMDS results: the projections of the response Y^j values on the ordinate axis NMDS1-NMDS2 and the relations with the components of predictor variables x_m are shown in Fig. 5.

At constructing of regression trees of an exogenous variables array X with multivariate responses Y (MRT), the task of estimating the affect of the components of exogenous variables x_m on the common variability of the components of the multivariate response y_k is solved, as shown in Fig. 6.



a b

Figure 4 – Diagrams of unconstrained ordination – a and constrained ordination – b of the array of responses Y on principal components plane (PCA)

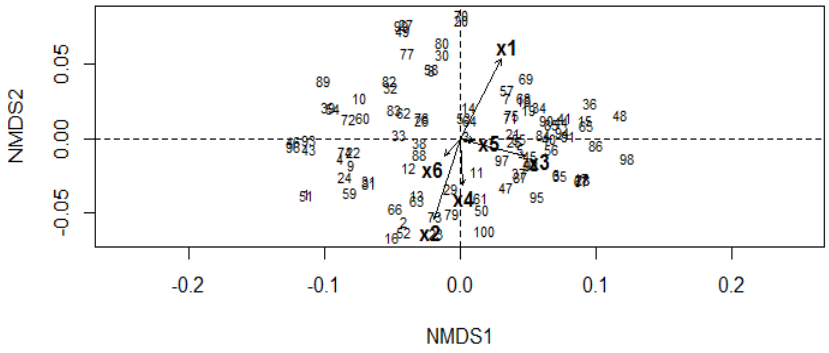


Figure 5 – The ordinance results of the response values Y^j on the NMDS1-NMDS2 axis and the relations with the predictors components x_m

© Havrylko Ye. V., Kurchenko O. A., Tereshchenko A. I., 2019
DOI 10.15588/1607-3274-2019-1-16

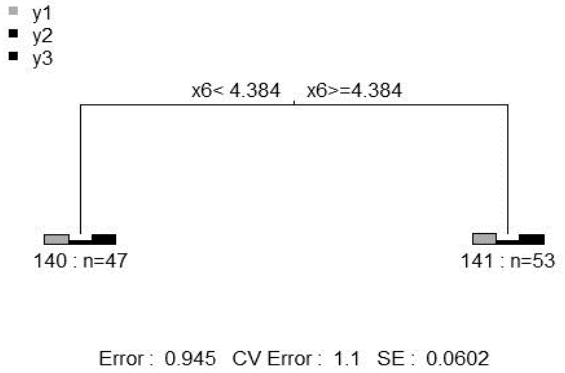


Figure 6 – MRT with multivariate response y_k for clusters of an array of exogenous variables X

MRTs are formed as a result of a recursive operation of dividing the rows of the responses data array Y into homogeneous subsets using an array of exogenous variables X as the separating parameters.

MRTs provide visualization of the effect of the components of exogenous variables x_m on the formation of an array of multivariate responses Y and show which components of the array initiate the decomposition of the researched population Y in the tree nodes and determine the composition of the formed objects Y.

The results demonstrate the great potential of MSA at the analysis of time multivariate data for a practical application of products quality assurance at the stage of the initial manufacture process design.

6 DISCUSSION

The article presents the results of the research at the design stage of the manufacture process, during which a multivariate statistical analysis of the cause-effect relationships between the variability of the critical quality attributes and the action of certain factors identified with the critical process parameters. The study contains elements of exploratory, descriptive and analytical statistical analysis.

The evaluation of the x_m predictors variable importance and the tendency to clustering the X array, taking into account the possible multicollinearity of the components x_m, were estimated using the PCA-method and visualization of the trends formation of clusters (VAT).

The PCA-method yields the relative expected contribution of each component x_m (Com.m) to the common data variation, as shown in Table 1.

Note that all components of a multivariate X array are characterized by an almost equal contribution (proportions of variance 16.6%) to the expected distribution and together they give a common contribution of 100%. The clustering tendency available in this test is characterized by Hopkins statistics (0.5087), which shows a moderate tendency to form clusters at a random distribution of data.

Table 1 – The expected contribution of components x_m to the common variation of the X data array

Comp. / Param.	Com.1	Com.2	Com.3	Com.4	Com.5	Com.6
Standard deviation	1.15	1.02	0.99	0.97	0.937	0.84
Prop-n of variance	0.22	0.17	0.16	0.15	0.14	0.12
Cum-ve prop-n	0.22	0.40	0.57	0.73	0.878	1.00

The criterion for the effectiveness of clusterization is based on the metric hypothesis of compactness. If it is possible to find such a division of observations into groups that the distances between observations from one group (intra-cluster distances) will be less than a definite value $\varepsilon > 0$, and between observations from different groups (cross-cluster distance) – bigger than ε and therefore clustering is successful.

According to this criterion, the results of assessing the quality of clustering according to the Zagoruiko test for clusters of an array of multivariate data X were obtained (Fig. 3b), as shown in Table. 2. In Table 2 on the main diagonal are located average intra-cluster distances, which are smaller than the cross-cluster distances (off-diagonal elements of the table), that indicates the possibility of division the value space of an X array into two clusters.

Table 2 – The quality data of clustering estimation by the Zagoruiko test

Cl. dist / №	Cl. dist. 1	Cl. dist. 2
1	3.044665	3.623451
2	3.623451	3.009617

The operation of transforming a homogeneous table of endogenous variables observations Y into a table with factor variables was carried out after research the characteristics of this array.

The results of the projection in the principal components orthogonal coordinate system of some features of the internal relation of the multivariate response array components Y (unconstrained ordination) are shown in Table 3.

Table 3 – The results of unconstrained ordination of the data Y array on the principal components axis

Dim. / Y	Dim. 1	Dim. 2	Dim. 3
$y1$	43.56287	10.7237730	45.71336
$y2$	8.30943	88.6417708	3.04880
$y3$	48.12770	0.6344562	51.2378

From Table 3 that the components of the responses array y_k in value of their contribution to the components of the decomposition Dim1, Dim2, Dim3, the factors do not form and have a weak correlation. The unconstrained ordination is limited only to an estimate of the distribution

© Havrylko Ye. V., Kurchenko O. A., Tereshchenko A. I., 2019
DOI 10.15588/1607-3274-2019-1-16

of the elements y_k on the principal components plane and the interpretation of the contribution of each component. The constrained ordination allows to establish correspondence the variability of multivariate values of responses Y^j with affect of exogenous variables x_m.

The constrained ordination of the responses array (Fig. 4b) is shown against the background of the components (dotted lines) and corresponds to relation the internal features of responses Y^j with the factorized effects of the critical process parameters X^j. The location of responses Y^j in both factor objects indicates a weak correlation of the components y_k and the predominance of random affects of the components x_m on the response components variability Y^j. This fact makes it necessary the research of the component influence of predictors x_m on the variability of the components of a multivariate dependent random variable Y^j. The predominance result of the influence the random constituent of the predictors variability x_m on the components of multivariate responses Y^j is confirmed by the NMDS of projections of the array values Y on the ordinate axis NMDS1-NMDS2.

The NMDS method is most applicable for data arrays with a significant influence of random or unaccounted factors, as in the case under consideration.

The results of application of the NMDS method are summarized in Table 4.

Table 4 – The NMDS results of the response array *Y* on the NMDS1-NMDS2 axis and the relations with the predictors x_m

NMDS / X	NMDS1	NMDS2	r2	Pr(>r)
*x*1	0.48665	0.87360	0.0329	0.201
*x*2	–0.33369	–0.94268	0.0292	0.239
*x*3	0.97049	–0.24114	0.0213	0.327
*x*4	0.06534	–0.99786	0.0087	0.665
*x*5	0.99269	–0.12071	0.0010	0.960
*x*6	–0.71850	–0.69552	0.0024	0.900

From Table 4 it follows that with a low statistical significance of the influence of exogenous variables x_m on the responses Y^j, there is a high correlation of some components x_m with the ordinate axes (for example, *x5* with NMDS1 and *x4* with NMDS2).

At interpreting (Fig. 5), it should be borne in mind that the result of the nonmetric scaling of multivariate responses Y^j is imposed on the components of exogenous variables x_m, which situated depending on their calculated correlation coefficients with the ordinate axes NMDS1-NMDS2.

From Fig. 5 that the NMDS model, even with a small number of multivariate data components *X* and *Y*, gives only a conditional idea of the affect of exogenous variables x_m on the responses Y^j.

The detailing impact of the components x_m is usually performed by projection on the constrained ordinance of the responses Y^j of isolines contours of the smoothed components x_m of the data array *X*. Note, that with increasing number of components, as well as with a weak tendency to clustering of *X* and *Y* array values, the analysis becomes more complicated.

At the final stage of the analysis, an MRT of an array of exogenous variables *X* with a multivariate response *Y* was formed, as shown in Fig. 6. For this MRT, "leaves" are clusters of objects *Y* with minimal differences between points in a multivariate space within each cluster. Note, that in order to explain the effect of variables x_m on responses Y^j the MRT software package has the option of selecting a various number of clusters. For this MRT (Fig. 6), the dividing variable is the component *x6*, which forms equivalent clusters with an almost equal number of observations (47 and 53) at the absence of explicit dominance of the components y_k.

The discussion shows that the experimental data of the computer format require processing, analysis and interpretation procedures that determine the potential trends and dependencies between unit components and objects of the time multivariate data for the purpose of practical interpretation at the stage of the initial manufacture process design.

CONCLUSIONS

The idea of design a quality function of the manufacture process at each of QFD stages in order to obtain predictable values of the attributes of the product critical quality at variability of the critical process parameters values within the tolerances is realized using the multivariate statistical analysis methods.

The scientific novelty of obtained results is that the MSA method uses the preparatory transformation of clusters of critical process parameters of manufacture process into factors of product critical quality attributes is carried out.

This method is used under statistical experiments with multivariate arrays of critical quality attributes of computer format to products quality assurance at the stage of the initial manufacture process design.

The researches results are consequence of combining the causal methods of multivariate statistical analysis that allow to obtaining the numerical characteristics with an adequate **practical interpretation** of the influence the critical process parameters on the product critical quality attributes.

© Havrylko Ye. V., Kurchenko O. A., Tereshchenko A. I., 2019
DOI 10.15588/1607-3274-2019-1-16

The **prospect of further research** consists in the synthesis or application of robust design methods to provide standard values of product critical quality attributes that are resistant to the affects of various environmental factors and to the variability of critical process parameters.

REFERENCES

1. Deming W. E. Out of the crisis, *Cambridge, Mass: Massachusetts Institute of Technology, Center for Advanced Engineering Study; Cambridge: Cambridge University Press,* 1986, 507 p.
2. Lionberger R. A., Lee S. L., Lee L. M., Raw A., X. Yu. Lawrence Quality by Design: Concepts for ANDAs, *American Association of Pharmaceutical Scientists journal.* Jun. 2008, 10(2), pp. 268–276. DOI: 10.1208/s12248-008-9026-7
3. ICH Q8. Pharmaceutical Development. [Electronic resource]. Access mode: http://www.ich.org/products/guidelines/quality/quality-single/article/pharmaceutical-development.html
4. Eshan S. J. A Case Study on Quality Function Deployment (QFD), *International Organization of Scientific Research Journal of Mechanical and Civil Engineering (IOSR-JMCE), Nov-Dec*, 2012, Volume 3, Issue 6, pp. 27–35. https://doi.org/10.9790/1684-0362735
5. Zhang L., Mao S. Application of quality by design in the current drug development, *Shenyang Pharmaceutical University, Wenhua Road, Shenyang, China: Asian journal of pharmaceutical sciences,* 2017 (12), No.103, pp. 1–8. https://doi.org/10.1016/j.ajps.2016.07.006
6. Guidance for Industry PAT A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. [Electronic resource]. Access mode: http://www.fda.gov/cder/OPS/PAT.htm
7. ISO 9001: 2015. Quality management systems. [Electronic resource]. Access mode: https://www.iso.org/standard/62085.html
8. ISO 22000: 2005 Food safety management systems. [Electronic resource]. Access mode: https://www.iso.org/standard/35466.html
9. Legendre P. Short course on Advanced spatial ecology. Research Center of Plant Ecology and Biodiversity Conservation, *Institute of Botany, Chinese Academy of Sciences,* 1–5 July 2006, pp. 575–600. [Electronic resource]. Access mode: http://biol09.biol.umontreal.ca/PLcourses/Legendre_Canonical.pdf
10. Rao S., Samant P., Kadampatta A., Shenoy R. An Overview of Taguchi Method: Evolution, Concept and Interdisciplinary Applications, *International Journal of Scientific & Engineering Research,* October 2013, Vol. 4, Issue 10, pp. 621–626. [Electronic resource]. Access mode: https://www.ijser.org/onlineResearchPaperViewer.aspx?An-Overview-of-Taguchi-Method-Evolution-Concept-and-Interdisciplinary-Applications.pdf
11. Coghlan A. A Little Book of R For Multivariate Analysis, Release 0.1, February 19 2017, 47 p. [Electronic resource], Access mode: http://a-little-book-of-r-for-time-series.readthedocs.org/
12. Food and Drug Administration. Final Report on Pharmaceutical cGMPs for the 21st Century. A Risk Based Approach. [Electronic resource]. Access mode: https://www.fda.gov/downloads/drugs/developmentapprovalprocess/manufacturing/questionsandanswersoncurrentgoodmanufacturingpracticescgmpfordrugs/ucm176374.pdf
13. [Kannissery P., Tahir M. A., Charoo N. A., Ansari S. H., Ali J.]Pharmaceutical product development: A quality by design approach, *International Journal of Pharmaceutical Investigation*, Jul-Sep 2016, 6(3), pp. 129–138. DOI: 10.4103/2230-973X.187350.
14. ISO 31000 Risk management. [Electronic resource]. Access mode: https://www.iso.org/iso-31000-risk-management.html
15. Guidance for Industry. PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. [Electronic resource]. Access mode: https://www.fda.gov/downloads/drugs/guidances/ucm070305.pdf
16. Mohamed I. Progressive Modeling: The Process, the Principles, and the Applications, *Procedia Computer Science,* 2013, Vol. 16, pp. 39–48. https://doi.org/10.1016/j.procs.2013.01.005
17. Araujo F., Pimentel P., Fogliatto F. S. Variable selection methods in multivariate statistical process control: A systematic literature review, *Department of Industrial Engineering, Federal University of Rio Grande do Sul, 90035-190 Porto Alegre, RS, Brazil, Computers & Industrial Engineering.* January 2018, Vol. 115, pp. 603–619. https://doi.org/10.1016/j.cie.2017.12.006
18. Göhler S. M., Husung S., Howard T. J. The Translation between Functional Requirements and Design Parameters for Robust Design, *Procedia College International pour la Recherche en Productique (CIRP),* 2016, Vol. 43, pp. 106–111. https://doi.org/10.1016/j.procir.2016.02.028
19. Gausemeier J., Gaukstern T., Tschirner C. Systems Engineering Management Based on a Discipline-Spanning System Model, *Procedia Computer Science,* 2013, Vol. 16, pp. 303–312. https://doi.org/10.1016/j.procs.2013.01.032
20. MacCalman A., Kwak H., McDonald M., Upton S. Capturing Experimental Design Insights in Support of the Model-based System Engineering Approach, *Procedia Computer Science*, 2015, Vol. 44, pp. 315–324. https://doi.org/10.1016/j.procs.2015.03.030
21. Kaiser L., Bremer C., Dumitrescu R. Exhaustiveness of Systems Structures in Model-Based Systems Engineering for Mechatronic Systems, *Procedia Technology,* 2016, Vol. 26, pp. 428–435. https://doi.org/10.1016/j.protcy.2016.08.055
22. Mac Calman A., Lesinski G., Goerger S. Integrating External Simulations Within the Model-Based Systems Engineering Approach Using Statistical Metamodels, *Procedia Computer Science*, 2016, Vol. 95, pp. 436–441. https://doi.org/10.1016/j.procs.2016.09.309
23. Lemazurier L., Chapurlat V., Grossetête A. // An MBSE Approach to Pass from Requirements to Functional Architecture, *International Federation of Accountants (IFAC)-Papers OnLine*, July 2017, Vol. 50, Issue 1, pp. 7260–7265. https://doi.org/10.1016/j.ifacol.2017.08.1376
24. Clement S., Register A., Wise R. Key Enablers for Leveraging Non-Development Items in a System, *Procedia Computer Science,* 2015, Vol. 44, pp. 164–173. https://doi.org/10.1016/j.procs.2015.03.009
25. Cohen J., Cohen P., West S. G., Aiken L. S. Applied multiple regression/correlation analysis for the behavioral sciences. Mahwah, New Jersey, US: Lawrence Erlbaum Associates Publishers, Third Edition, 2003, 703 p. https://doi.org/10.1177/019394598000200320
26. Azevedo A., Santos M. F. KDD, SEMMA and CRISP-DM: a parallel overview, *IADIS European Conference on Data Mining, Amsterdam, the Netherlands 22–27 July 2008, Proceedings of the IADIS European Conference on Data Mining,* 2008, pp. 182–185. [Electronic resource]. Access mode: https://pdfs.semanticscholar.org/7dfe/3bc6035da527deaa72007a27cef94047a7f9.pdf

Received 14.06.2018.
Accepted 01.10.2018.

© Havrylko Ye. V., Kurchenko O. A., Tereshchenko A. I., 2019
DOI 10.15588/1607-3274-2019-1-16

УДК 004.94:658

МЕТОД БАГАТОВИМІРНОГО СТАТИСТИЧНОГО АНАЛІЗУ ЧАСОВИХ БАГАТОВИМІРНИХ КРИТИЧНИХ АТРИБУТІВ ЯКОСТІ ПРОЦЕСУ ВИРОБНИЦТВА З ФАКТОРИЗАЦІЄЮ ДАНИХ

Гаврилко Є. В. – д-р техн. наук, професор кафедри комп'ютерної інженерії, Державний університет телекомунікацій, Київ, Україна.

Курченко О. А. – канд. техн. наук, доцент, завідувач кафедри управління інформаційною та кібернетичною безпекою, Державний університет телекомунікацій, Київ,Україна.

Терещенко І. В. – канд. техн. наук, доцент, доцент кафедри інфокомунікаційної інженерії, Харківський національний університет радіоелектроніки, Харків, Україна.

Терещенко А. І. – аспірант кафедри управління інформаційною та кібернетичною безпекою, Державний університет телекомунікацій, Київ, Україна.

АНОТАЦІЯ

Актуальність. У статті запропоновано метод розв'язання задачі забезпечення якості продукції на етапі первинного проектування процесу виробництва відповідно до актуальної процесно-аналітичної технології конструювання сучасних сертифікованих виробництв – «якість через дизайн» (QbD). Об'єктом дослідження є процес забезпечення якості продукту на ранніх етапах проектування. Метод використовує інформаційні технології багатовимірного статистичного аналізу MSA (Multivariate Statistical Analysis) для оцінки впливу часових багатовимірних критичних параметрів процесу виробництва на часові атрибути критичного якості продукту. Проводиться трансформація кластерів критичних параметрів процесу виробництва в чинники критичних атрибутів якості продукту.

Метод. Метод послідовно використовує: статистичні процедури розвідувального аналізу багатовимірних даних; трансформацію однорідних матриць значень спостережень критичних параметрів виробничого процесу і критичних атрибутів якості продукту в таблиці з факторізованими даними; побудову дерев регресії багатовимірних критичних параметрів процесу виробництва з багатовимірним відгуком. Метод реалізований за допомогою програмних пакетів мови R.

Результати. Факторізовані часові багатовимірні критичні атрибути якості процесу виробництва надають додаткові ступені свободи використання методів багатовимірного статистичного аналізу для оцінювання впливу факторів критичних параметрів виробничого процесу на часові багатовимірні критичні атрибути якості продукту.

Висновки. Запропонований метод статистичного аналізу представляє актуальну інформаційну технологію детального оцінювання впливу об'єктів часових багатовимірних даних критичних параметрів процесу виробництва і окремих складових на критичні атрибути якості продукту.

КЛЮЧОВІ СЛОВА: якість через дизайн, атрибути критичної якості, критичні параметри процесу, дизайн експерименту, багатовимірний статистичний аналіз.

УДК 004.94:658

МЕТОД МНОГОМЕРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА ВРЕМЕННЫХ МНОГОМЕРНЫХ КРИТИЧЕСКИХ АТРИБУТОВ КАЧЕСТВА ПРОЦЕССА ПРОИЗВОДСТВА С ФАКТОРИЗАЦИЕЙ ДАННЫХ

Гаврилко Е. В. – д-р техн. наук, профессор кафедры компьютерной инженерии, Государственный университет телекоммуникаций, Киев, Украина.

Курченко О. А. – канд. техн. наук, доцент, заведующий кафедрой управления информационной и кибернетической безопасностью, Государственный университет телекоммуникаций, Киев, Украина.

Терещенко И. В. – канд. техн. наук, доцент, доцент кафедры инфокоммуникационной инженерии, Харьковский национальный университет радиоэлектроники, Харьков, Украина.

Терещенко А. И. – аспирант кафедры управления информационной и кибернетической безопасностью, Государственный университет телекоммуникаций, Киев, Украина.

АННОТАЦИЯ

Актуальность. В статье предложен метод решения задачи обеспечения качества продукции на этапе первоначального проектирования процесса производства в соответствии с актуальной процессно-аналитической технологией конструирования современных сертифицированных производств – «качество через дизайн» (QbD). Объектом исследования является процесс обеспечения качества продукта на ранних этапах проектирования. Метод использует информационные технологии многомерного статистического анализа MSA (Multivariate Statistical Analysis) для оценки влияния временных многомерных критических параметров процесса производства на временные атрибуты критического качества продукта. Проводится трансформация кластеров критических параметров процесса производства в факторы критических атрибутов качества продукта.

Метод. Метод последовательно использует: статистические процедуры разведывательного анализа многомерных данных; трансформацию однородных матриц значений наблюдений критических параметров производственного процесса и критических атрибутов качества продукта в таблицы с факторизованными данными; построение деревьев регрессии многомерных критических параметров процесса производства с многомерным откликом. Метод реализован с помощью программных пакетов языка R.

Результаты. Факторизованные временные многомерные критические атрибуты качества процесса производства предоставляют дополнительные степени свободы использования методов многомерного статистического анализа для оценки влияния факторов критических параметров производственного процесса на временные многомерные критические атрибуты качества продукта.

Выводы. Предложенный метод статистического анализа представляет актуальную информационную технологию детальной оценки влияния объектов временных многомерных данных критических параметров процесса производства и отдельных составляющих на критические атрибуты качества продукта.

© Havrylko Ye. V., Kurchenko O. A., Tereshchenko A. I., 2019
DOI 10.15588/1607-3274-2019-1-16

КЛЮЧЕВЫЕ СЛОВА: качество через дизайн, атрибуты критического качества, критические параметры процесса, дизайн эксперимента, многомерный статистический анализ.

ЛІТЕРАТУРА / ЛИТЕРАТУРА

1. Deming W. E. Out of the crisis / W. E. Deming. – Cambridge, Mass: Massachusetts Institute of Technology, Center for Advanced Engineering Study; Cambridge: Cambridge University Press, 1986. – 507 p.
2. Quality by Design: Concepts for ANDAs / [R. A. Lionberger, S. L. Lee, L. M. Lee et al.] // American Association of Pharmaceutical Scientists journal. – Jun. 2008 – 10 (2) – P. 268–276. DOI: 10.1208/s12248-008-9026-7
3. ICH Q8. Pharmaceutical Development. [Electronic resource] – Access mode: http://www.ich.org/products/guidelines/quality/quality-single/article/pharmaceutical-development.html
4. Eshan S. J. A Case Study on Quality Function Deployment (QFD) / S. J. Eshan // International Organization of Scientific Research Journal of Mechanical and Civil Engineering (IOSR-JMCE), Nov-Dec. – 2012. – Volume 3, Issue 6 – P. 27–35. https://doi.org/10.9790/1684-0362735
5. Zhang L., Mao S. Application of quality by design in the current drug development / L. Zhang, S. Mao // Shenyang Pharmaceutical University, Wenhua Road, Shenyang, China: Asian journal of pharmaceutical sciences. – 2017 (12). – No. 103. – P. 1–8. https://doi.org/10.1016/j.ajps.2016.07.006
6. Guidance for Industry PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. [Electronic resource] – Access mode: http://www.fda.gov/cder/OPS/PAT.htm
7. ISO 9001: 2015. Quality management systems. [Electronic resource] – Access mode: https://www.iso.org/standard/62085.html
8. ISO 22000: 2005 Food safety management systems. [Electronic resource] – Access mode: https://www.iso.org/standard/35466.html
9. Legendre P. Short course on Advanced spatial ecology. Research Center of Plant Ecology and Biodiversity Conservation / P. Legendre // Institute of Botany, Chinese Academy of Sciences. – 1–5 July 2006. – P. 575–600. [Electronic resource] – Access mode: http://biol09.biol.umontreal.ca/PLcourses/Legendre_Canonical.pdf
10. An Overview of Taguchi Method: Evolution, Concept and Interdisciplinary Applications / [S. Rao, P. Samant, A. Kadampatta, R. Shenoy] // International Journal of Scientific & Engineering Research. – October 2013. – Vol. 4, Issue 10. – P. 621–626. [Electronic resource] – Access mode: https://www.ijser.org/onlineResearchPaperViewer.aspx?An-Overview-of-Taguchi-Method-Evolution-Concept-and-Interdisciplinary-Applications.pdf
11. Coghlan A. A Little Book of R For Multivariate Analysis, Release 0.1 / A. Coghlan, February 19 2017. – 47p. [Electronic resource] – Access mode: http://a-little-book-of-r-for-time-series.readthedocs.org/
12. Food and Drug Administration. Final Report on Pharmaceutical cGMPs for the 21st Century – A Risk Based Approach. [Electronic resource] – Access mode: https://www.fda.gov/downloads/drugs/developmentapprovalprocess/manufacturing/questionsandanswersoncurrentgoodmanufacturingpracticescgmpfordrugs/ucm176374.pdf
13. Pharmaceutical product development: A quality by design approach / [P. Kannissery, M. A. Tahir, N. A. Charoo, S. H. Ansari, J. Ali] // International Journal of Pharmaceutical Investigation – Jul-Sep 2016. – 6(3). – P.129–138. DOI: 10.4103/2230-973X.187350.
14. ISO 31000 Risk management. [Electronic resource] – Access mode: https://www.iso.org/iso-31000-risk-management.html
15. Guidance for Industry. PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. [Electronic resource] – Access mode: https://www.fda.gov/downloads/drugs/guidances/ucm070305.pdf
16. Mohamed I. Progressive Modeling: The Process, the Principles, and the Applications / I. Mohamed // Procedia Computer Science. – 2013. – Vol. 16. – P. 39–48. https://doi.org/10.1016/j.procs.2013.01.005
17. Araujo F. Variable selection methods in multivariate statistical process control: A systematic literature review / F. Araujo, P. Pimentel, F. S. Fogliatto // Department of Industrial Engineering, Federal University of Rio Grande do Sul, 90035-190 Porto Alegre, RS, Brazil, Computers & Industrial Engineering. – January 2018. – Vol. 115. – P. 603–619. https://doi.org/10.1016/j.cie.2017.12.006
18. Göhler S. M. The Translation between Functional Requirements and Design Parameters for Robust Design / S. M. Göhler, S. Husung, T. J. Howard // Procedia College International pour la Recherche en Productique (CIRP). – 2016. – Vol. 43. – P. 106–111. https://doi.org/10.1016/j.procir.2016.02.028
19. Gausemeier J. Systems Engineering Management Based on a Discipline-Spanning System Model / J. Gausemeier, T. Gaukstern, C. Tschirner // Procedia Computer Science. – 2013. – Vol. 16. – P. 303–312. https://doi.org/10.1016/j.procs.2013.01.032
20. Capturing Experimental Design Insights in Support of the Model-based System Engineering Approach / [A. MacCalman, H. Kwak, M. McDonald, S. Upton] // Procedia Computer Science. – 2015. – Vol. 44. – P. 315–324. https://doi.org/10.1016/j.procs.2015.03.030
21. Kaiser L. Exhaustiveness of Systems Structures in Model-Based Systems Engineering for Mechatronic Systems / L. Kaiser, C. Bremer, R. Dumitrescu // Procedia Technology. – 2016. – Vol. 26. – P. 428–435. https://doi.org/10.1016/j.protcy.2016.08.055
22. Mac Calman A. Integrating External Simulations Within the Model-Based Systems Engineering Approach Using Statistical Metamodels / A. Mac Calman, G. Lesinski, S. Goerger // Procedia Computer Science. – 2016. – Vol. 95. – P. 436–441. https://doi.org/10.1016/j.procs.2016.09.309
23. Lemazurier L. An MBSE Approach to Pass from Requirements to Functional Architecture / L. Lemazurier, V. Chapurlat, A. Grossetête // International Federation of Accountants (IFAC)-Papers OnLine. – July 2017. – Vol. 50, Issue 1. – P. 7260–7265. https://doi.org/10.1016/j.ifacol.2017.08.1376
24. Clement S. Key Enablers for Leveraging Non-Development Items in a System / S. Clement, A. Register, R. Wise // Procedia Computer Science. – 2015. – Vol. 44. – P. 164–173. https://doi.org/10.1016/j.procs.2015.03.009
25. Applied multiple regression/correlation analysis for the behavioral sciences / [J. Cohen, P. Cohen, S. G. West, L. S. Aiken]. – Mahwah, New Jersey, US: Lawrence Erlbaum Associates Publishers. Third Edition, 2003. – 703 p. https://doi.org/10.1177/019394598000200320
26. Azevedo A. KDD, SEMMA and CRISP-DM: a parallel overview / A. Azevedo, M. F. Santos // IADIS European Conference on Data Mining, Amsterdam, the Netherlands 22–27 July 2008: Proceedings of the IADIS European Conference on Data Mining, 2008. – P. 182–185. [Electronic resource] – Access mode: https://pdfs.semanticscholar.org/7dfe/3bc6035da527deaa72007a27cef94047a7f9.pdf

© Havrylko Ye. V., Kurchenko O. A., Tereshchenko A. I., 2019
DOI 10.15588/1607-3274-2019-1-16