

Detecting changes simulation of the technological objects' information states

*Aleksandr Skatkov*¹, *Aleksey Bryukhovetskiy*¹, *Dmitriy Moiseev*^{1,*}, and *Raisa Litvinova*¹

¹Federal State Autonomous Educational Institution of Higher Education «Sevastopol State University», Russian Federation, 299053, Sevastopol

Abstract. The purpose of the production processes' states simulation has two constituents. The first is to ensure the effective operation of technological facilities. The second is to improve the quality and competitiveness of the produced products. The proposed new scientific approach aims to develop the known statistical methods in the following direction. They are to be adopted for monitoring accuracy and stability of production processes: real-time processing of large amount of current data, large volumes of a priori information used for decision-making, adjusting the parameters of the monitoring system, critical technological conditions detection. It can prevent great economic damage or even catastrophic consequences.

1 Introduction

The completeness of use of latest methods and tools of computer modelling greatly affect the effectiveness of technological processes management. Simulation of multiple options for the production cycle development, including, a variety of precritical emergencies and technical incidents that may occur, makes it possible to choose the most productive and safe technological mode, appropriate management system and structure.

In modern conditions, special attention is to be paid to the control of the products' quality and to the production processes' states monitoring. This is because the usage of production processes intensification, intellectual technologies implementation and high energy capacities. At present, so-called critical systems and infrastructures are becoming widespread. Their criticality is manifested in the fact that sometimes small changes in parameters or structures lead to catastrophic failures, which are of the nature of accidents and catastrophes. Therefore, the operation of such facilities requires the constant operation of the statistical control of the parameters that accompanies them. The main task of such solutions is the systematic analysis and interpretation of processes occurring in a heterogeneous production environment.

Organization of the technological objects state monitoring is connected with the following conflict situation. On the one hand, the real-time provision dictates the need of small-volume samples usage. On the other hand, it is necessary to provide the required level of hypotheses acceptance reliability on the results of control. So it is needed to find a

* Corresponding author: dmitriymoiseev@mail.ru

compromise decision in conditions of a priori information deficit and initial statistical uncertainty. It should ensure minimization of technological objects operation risks, meeting the requirements of speed and reliability of situation recognition. Such solutions should be sought by compressing large volumes of processed data without losing their validity. According to the authors' opinion, the most pressing problems that require urgent solution are the following: information saturation determination using statistical methods and models for assessing the technological processes' states; increase in the level of states classification reliability; event classification for small samples; decision-making in conditions of a priori information deficit.

2 State of the art

The solutions for operational monitoring and technological processes monitoring in the oil and gas industry are known [1]. In the paper models and classification of monitoring systems are considered from the viewpoint of queuing theory. The basic core product model is presented in [2]. It is necessary for successful functioning of PLM-systems of reliable, complete and efficient data models. Extensive models of product data are highlighted: OAM, PSRL, Product family evolution model, Heterogeneous material model, Mechatronic device model, Embedded system model, Product lifecycle management (PLM) system. PLM technologies combine techniques and tools for information support of products throughout all stages of their life cycle. In [3] the author presents probabilistic-statistical methods of decision theory, explaining "how, from the observations of the realization of a random process, to detect spontaneously arising effects, when the properties of the process change spasmodically". In conditions of production processes, most of the data analysis problems are related to the study of stochastic systems. Known methods for solving the problems of state monitoring can not be a perfect operational tool, because they do not provide the required quality control of the state, in particular, with their help it is difficult to trace the moments of change in individual properties of the technological objects' states, such as accuracy, stability, quality indicators and others. It is advisable to search for such solutions within the framework of the statistical theory of hypothesis acceptance [4]. Since the possibilities for an analytical solution of such problems are very limited, it is suggested to use the simulation methods as an alternative. One of the tasks of technological objects monitoring is measuring metrics of information states of process technology objects made on the basis of small samples based on specified decision-making criteria to assess the information situation. It is suggested to use the Spearman criterion as an information measure [5].

The article considers a method for detecting changes in the state of a technological object. The method is based on the notion of heteroscedasticity as the inhomogeneity of observations, which is expressed in the unequal variance of the random error of the regression model. This concept is widely used in econometrics [5]. The fact of detecting of the object's state heterogeneity (G-effect) can be used to detect changes in equipment conditions under the influence of external disturbances. The process of detecting the G-effect includes such steps as: setting the law of distribution and intensity of the input stream, calculating the processing time for applications, constructing regression dependencies, and deciding whether a G-effect exists. The assumption of the residues variance constancy is known as the assumption of homoscedasticity. If this assumption is violated and the variance of the residues is not constant, then it is said that the estimates are heteroscedastic (G-effect). It is the state's heterogeneity assessment that can be used to detect the production process instability and to identify the technological process disruption under external influences that lead to the operation equipment modes violation, the accuracy reproducing of the quantitative and qualitative products' characteristics and their compliance with normative and technical

documentation. The main advantages of the proposed method for assessing the state of a technological object on the basis of an information measure are the following: sensitivity to changes in the technological objects' states, low computational complexity, and adaptability to external influences.

3 Results of experimental studies

For sufficiently considered general statement of the problem the evaluation of the results of object's states observations identify the significance of its qualitative change. A collection of observations is a set of measurements of the object's states $X = \{X_1, X_2, \dots, X_k\}$, $X_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}$, x_i – value of the measured parameter of technological object, k – number of samples, n – the volume of each sample.

Statistics describing the process of modelling [5] is calculated using (1):

$$t_{expr} = \rho\sqrt{n-2}/\sqrt{1-\rho^2}, \quad (1)$$

where ρ – Spearman's rank correlation coefficient; t_{expr} – the observed value of Student's statistics during the experiment.

Inequality (2) is analysed for assessing the object's state by the hypothesis accepting criterion about the G-effect presence.

$$|t_{expr}| > t_{st(0.01;n-2)}, \quad (2)$$

where n – number of experiments ($n = 40$), t_{st} – critical value. When inequality (2) is satisfied the G-effect takes place.

In this paper, the results of simulating QMS using the example of processing applications distributed are studied according to the gamma law. To simulate the process of changing the state of objects, the simulation environment AnyLogic has been used.

The influence of the following parameters has been considered:

- $\lambda, \sigma(\lambda)$ – intensity of requests generation;
- $\mu, \sigma(\mu)$ – intensity of requests' service;
- ρ – system load;
- n – sample sizes;
- τ – moments of time characteristics measurement;
- p – significance levels.

The influence of these parameters was estimated using gamma distribution models based on Spearman's criterion for nonparametric statistics.

Experimental researches of qualitative change in technological object condition were made on the basis of heteroscedasticity effect.

The characteristics of the considered single-channel QMS are the following:

- intensity of generation of requests $M(\lambda)=0,7$ and $\sigma(\lambda)=0,09$;
- intensity of requests' service $M(\mu)=0,8$ and $\sigma(\mu) = [0,17; 0,2; \dots; 0,4]$;
- storage capacity – 5, functioning algorithm – FIFO.

Gamma distribution of requests' intensity. A series of experiments has been performed in which the intensity of receipt and processing of applications in the system was distributed according to gamma law. A random value generator using the gamma distribution – gamma (α, β) has two parameters: α – form parameter, β – scale parameter.

Scale parameter β and form parameter α of this distribution is associated with the mathematical expectation and variance based on the system of equations (3).

$$\begin{cases} \alpha * \beta = M, \\ \alpha * \beta^2 = \sigma^2, \\ \alpha > 0, \\ \beta > 0. \end{cases} \quad (3)$$

The solution of system of equations (3) makes it possible to determine the most plausible values of the parameters of the gamma distribution. Parameters' values transformation from normal distribution into gamma distribution is done. The following intensity of input generation requests' flow is got: for $M(\lambda) = 0.7, \sigma(\lambda) = 0.09 \Rightarrow \alpha = 5.5, \beta = 0.128$. The values of requests' service intensity are given in table 1.

Table 1. Converting the parameters of the normal distribution to the gamma distribution.

μ	$\sigma(\mu)$	α	β
0,8	0,17	3,7	0,215
0,8	0,2	3,2	0,25
0,8	0,24	2,7	0,297
0,8	0,26	2,45	0,326
0,8	0,32	2	0,4
0,8	0,33	1,92	0,416
0,8	0,35	1,8	0,445
0,8	0,37	1,73	0,463
0,8	0,38	1,7	0,473
0,8	0,4	1,6	0,5

The simulation results using the transformed parameters are shown on the Fig. 1.

The results for different samples are presented in table 2, where different colours mean different significance levels for the observed values $p=0,001$ – blue, $p=0,005$ – green, $p=0,01$ – yellow, which correspond to the G-effect. Results, that are represented in table 2, are depicted in the graph form for the sample $n = 40, \tau = 31$ on the fig. 2.

With a sample size $n = 60$ the model is constantly in the heterogeneous area. When comparing the results obtained for a gamma distribution with a normal distribution the most stable areas of G-effect manifestation have been observed, as before, with a sample size $n = 40$.

Table 3 is constructed for different τ times for the same source data, $n = 40$.

Table 2. Results of experiments for different volumes of n .

α	β	$t(20-2)$	$t(40-2)$	$t(60-2)$
3,7	0,215	2,71026823	3,4551096	5,84330347
3,2	0,25	3,82052841	3,31260798	4,77561313
2,7	0,297	1,79406021	3,53239149	4,67969141
2,45	0,326	4,79421496	6,1844856	7,35062798
2	0,4	3,83611859	2,84421792	4,97467211
1,92	0,416	6,74909246	6,70973064	6,68590119
1,8	0,445	3,85179474	6,50216333	6,65408968
1,73	0,463	3,78960209	5,31104172	4,34127791
1,7	0,473	4,04699038	6,27757451	9,63782085
1,6	0,5	1,92696007	4,29294489	7,06779261

Table 3. The results of experiments for different instants of time τ for $n = 40$.

α	β	$\tau=21$	$\tau=31$	$\tau=41$	$\tau=51$
3,7	0,215	4,607527	3,45510	3,321075	2,804199
3,2	0,25	3,498866	3,312607	3,432512	3,084391
2,7	0,297	4,364688	3,532391	3,79828	4,128619
2,45	0,326	3,469068	6,184485	4,132653	6,177916
2	0,4	2,893882	2,844217	3,707346	3,476063
1,92	0,416	6,584664	6,709730	5,850022	4,733136
1,8	0,445	7,729158	6,502163	5,760100	4,993624
1,73	0,463	5,853054	5,311041	2,792867	2,899508
1,7	0,473	6,035988	6,277574	8,73027	8,840128
1,6	0,5	6,064622	4,292944	4,124588	7,588412

The obtained data, presented in table 3, leads to the conclusion that the most stable G-effect is manifested when $\tau = 31$ and $n = 40$ (fig. 2).

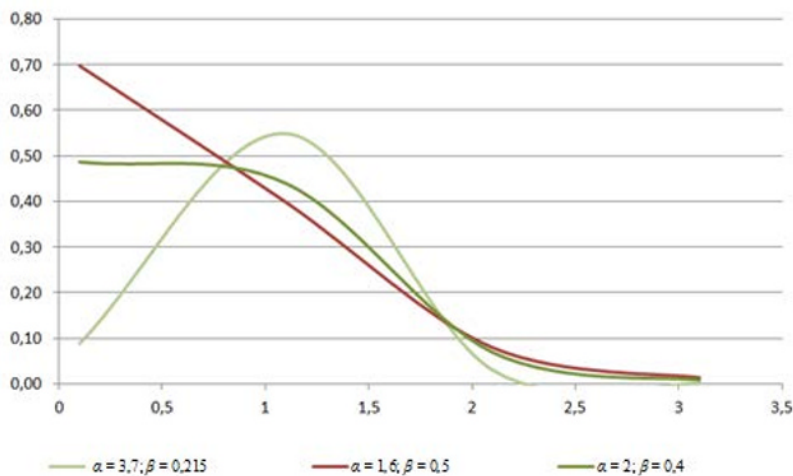


Fig. 1. Examples of distribution function graphs with different parameters .

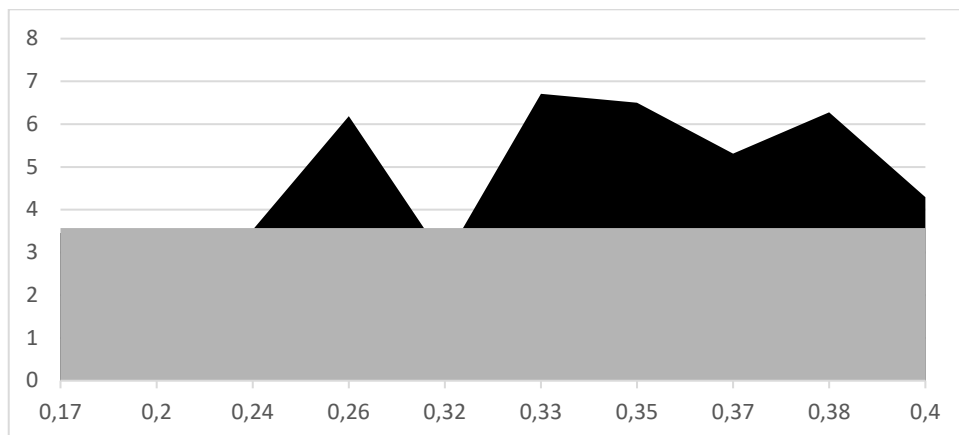


Fig. 2. Hetero areas for $n = 40$, $\tau = 31$.

4 Conclusion

1. The experiments have been carried out to determine the influence of a different parameter values on changes in the state of technological objects using the gamma distribution. The obtained results allow to state that the established gamma distribution parameters allow to ensure the adequacy of the simulation model. The simulation results for the gamma distribution confirm the stable detection regions of the G-effect obtained earlier [6] for the normal law of the distribution of requests' intensities. Moreover, in the model, based on the gamma distribution of the hetero regions, they have a bigger time length than the model using the normal distribution law.

2. Experiments have shown that the most optimal choice for gamma distribution, which provides high reliability of G-effect detection, is the use of sample size $n=40$. When using a sample of $n=20$ heteroscedasticity manifests itself unstably, and when using a sample $n=60$ the model was constantly in the heterogeneous area.

3. Various times τ were investigated. Thus, it was established that for $\tau = 21$, $\tau = 41$, $\tau = 51$ the G-effect was unstable. On the basis of the above, it is recommended to use $\tau=31$ with the aim of increasing the reliability of the results of the assessment of the state.

4. A general recommendation for assessing the detection of G-effect is the following. Depending on the purpose of the maintenance, whether it is critical or not, the decision-maker has the right to set the significance level – p , for which, on the one hand, high reliability, on the other hand, a minimum number of errors of the first and second kind is achieved, which means that risks will be reduced when making erroneous decisions, which, in turn, will ensure minimal losses.

In the plans for further studies, it is proposed to evaluate the complex effect of the values of the input parameters on the statistical sensitivity, statistical stability, areas of confident / uncertain recognition, for a different number of intervals, and also build a decision support system for G-effect detection and risk assessment.

This work was supported by the Russian Foundation for Basic Research (project № 18-47-920007).

References

1. E. Ohotnikov, Information technologies of modeling and management, **8**, (2004)
2. P. Xuwei, F. Lijun, W. Yiming, Applied Mechanics and Materials, **58–60** (2011,)
3. A. Shiryaev Probabilistic-statistical methods in the theory of decision-making, (2014)
4. S. Aivazyan, V. Mkhitarian, Unity Publishing House, (1998)
5. N. Kremer, UNITY-DAN, (2006)
6. A. Skatkov, A. Bryukhovetskiy, V. Shevchenko, D. Voronin AICT, (2016)