MULTIVARİATE CHARTS OF STATISTICAL CONTROL OF THE DYNAMIC PROCESS OF OIL FIELD DEVELOPMENT MANAGEMENT

Oleg Dyshin[1], Fuad Maharramov[2]
•
[1] Research Institute "Geotechnical Problems of Oil, Gas and Chemistry", Azerbaijan State Oil and Industry University, engakadem@gmail.com
[2] Azerbaijan State Oil and Industry University, oilqazgeology@gmail.com

Abstract

The properties of the procedure for constructing multidimensional Hotelling maps are investigated for the case of individual observations of a dynamic process. The features of the behavior of such maps are noted in comparison with multivariate statistical control of controllability and stability, using sample observations at a certain fixed point in time or in a short period of time. At the same time, detrending of non-stationary time series, describing the dynamics of each feature separately, as well as the transformation of the multidimensional sample distribution of observations to a joint multivariate normal distribution is used.

Keywords: optimal oil recovery, multivariate dynamic Hotelling charts, process detrending, multivariate normal distribution, statistically controlled process

I. Introduction

Development of oil fields is a complex process that depends on many geological and technological factors and must meet the requirements of the most optimal oil recovery [1,2]. In order to maintain the optimal mode of oil production, the task is to stabilize the development process based on the application of the concept of a statistically controlled process using Hotelling's multivariate control charts [3-5].

Typically, control charts (CC) are used to control the quality of some large volume of single-item products that are divided into subgroups (lots, lots) of the same volume n. At the same time, checks (observations) of the quality of each unit of production can be carried out according to one indicator (onevariate CC) or several indicators (multivariate CC) at the same moment in time with a random selection of the unit of the tested product. For $n = 1$, these observations are called individual.

When using CC to check the statistical controllability of dynamic processes (in particular, the process of developing an oil field), the following features should be taken into account.

For a multidimensional dynamic process defined by time series of discrete values of indicators in a sequence of time values t, each of these series describes a one-dimensional non-stationary random process with mathematical expectation, which is generally a function of time (process trend) and only for a stationary process its first moment (expectation) and the second moment (covariance) do not depend on time [6].

Therefore, to construct CC, the first differences of the initial time series are usually used, as is done in the "STATISTIKA" package. In this case, it is assumed that such a procedure leads to a stationary random process. However, only after taking the second time difference, the obtained time series can be approximately considered stationary [7, 8]. In this regard, it is advisable to construct a multidimensional dynamic CC to use for each indicator not the initial time series, but the corresponding detrended time series obtained by detecting and rejecting a trend, which is obviously stationary with a high accuracy of the calculated trend.

II. Statistical process control using control charts.

When constructing CC, two different phases are distinguished [3,4]. Phase I, carried out with a hindsight analysis, involves testing the statistical control of the process based on original or subgrouped observational data. This phase is commonly referred to as the start-up-stage of the process, in which a set of data (training sample) is obtained from which control limits are set for monitoring the process. The purpose of this phase is to identify statistical controllability and find the upper control limit (UCL) and lower control limit (LCL). In the second phase, on the basis of the obtained control limits, corrective control is carried out, including the detection of points of instability (outliers) of the process and subsequent regulation of the process with maintaining its statistical stability.

In this paper, we consider phase I of a multivariate statistical control with individual observations representing the values x_{ij} of several indicators xj, given in a certain sequence of times t_i, $i = 1,\ldots, m$. With regard to the process of developing an oil field x_{ij} - oil production indicators by years t_i, $i = 1,\ldots, m$.

III. Multivariate charts of statistical control at the start-up stage for individual observations.

Suppose that we are dealing with individual observations, that is, the sample of observations consists of m subgroups of observations with the same volume of observations n = 1. Let $\boldsymbol{x} = \left(x_1,\ldots,x_j,\ldots,x_p\right)$ there be a vector of p variables (indicators of the process under study) $x_1,\ldots, x_j,\ldots, x_p$ and x_{ij} - the value of the variable x_j in the i-th ($i = 1,\ldots, m$) observations. Let's introduce the notation:

$$\boldsymbol{X} = \left(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_i,\ldots,\boldsymbol{X}_m\right)^T ;$$

$$\boldsymbol{X}_i = \left(x_{i1},\ldots,x_{ip}\right)^T \left(i = 1,\ldots,m\right)$$

$$\bar{\boldsymbol{X}}_m = \left(\bar{x}_1,\ldots,\bar{x}_j,\ldots,\bar{x}_p\right) ,$$

$$\bar{x}_j = \frac{1}{m}\sum_{i=1}^{m} x_{ij}$$

$$\boldsymbol{S}_m = \frac{1}{m-1}\sum_{i=1}^{m}\left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_m\right)\left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_m\right)^T$$

where $\bar{\boldsymbol{X}}_m$ and $\boldsymbol{S}_m$ - sample mean vector and covariance matrix; T is the transposition sign.

To construct a multivariate CC based on Hotelling's T^2-statistic, it is assumed [9] that observations are independent and identically distributed (i.i.d.) random variables satisfying a p-dimensional normal distribution $\boldsymbol{N_p}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ with a mean vector $\boldsymbol{\mu} = \left(\mu_1,\ldots,\mu_p\right)^T$ and a covariance matrix

$$\Sigma = \left(\boldsymbol{X} - \boldsymbol{\mu}\right)\left(\boldsymbol{X} - \boldsymbol{\mu}\right)^T .$$

If the vector $\boldsymbol{X} = \left(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_i,\ldots,\boldsymbol{X}_m\right)$ obeys d-variate normal distribution, then the asymmetry indices $b_{1,d}$ and curvatures $b_{2,d}$ satisfy the relations $b_{1,d} = 0$ and $b_{2,d} = d\ (d + 2)$. Sample estimates of these indicators $b_{1,d} = \frac{1}{m}\sum_{h=1}^{m}\sum_{i=1}^{m} g_{hi}^3$, $b_{2,d} = \frac{1}{m}\sum_{i=1}^{m} g_{ii}^2$ where

$$g_{hi} = \left(\boldsymbol{X}_h - \bar{\boldsymbol{X}}\right)^T \cdot \frac{1}{m} \boldsymbol{Q}_m^{-1} \left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}\right),$$

$$\boldsymbol{Q}_m = (m-1)\boldsymbol{S}_m = \Sigma\left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}\right)^T,$$

$$\boldsymbol{X}_i = (x_{i1}, ..., x_{id}) \text{ (i=1,…,m)},$$

$$\bar{\boldsymbol{X}} = \bar{\boldsymbol{X}}_m = \left(\bar{X}_1, ..., \bar{X}_d\right),$$

$$\bar{X}_j = \frac{1}{m} \sum_{i=1}^{m} x_{ij}.$$

Mardia [10] showed that if $\boldsymbol{X} \in N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then for d> 2 and m> 50

$$A = \frac{1}{6} m b_{1,d} \sim \chi_f^2,$$

$$f = \frac{1}{b} d(d+1)(d+2)$$

and

$$B = \frac{b_{2,d} - d(d+2)}{\left[8d(d+2)/m\right]^{1/2}} \sim N_1(0,1)$$

where χ_f^2 is the χ²-distribution with f degrees of freedom, $N_1(0,1)$ is the standard one-dimensional normal distribution.

T_i^2 - Hotelling statistics corresponding to the observation $\boldsymbol{X}_i$ is written in the form $T_i^2 = \left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_m\right)^T \boldsymbol{S}_m^{-1} \left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_m\right)$. Assuming that the estimates $\bar{\boldsymbol{X}}_m$ and $\bar{S}_m$ characterize a sample from a p-dimensional normal population (general population) with a mean $\boldsymbol{\mu}$ and a covariance matrix Σ , the statistics T_i^2 obey the Pearson χ²-distribution with p degrees of freedom. In this case, at the initial stage, the lower control limit is written as $LCL = \chi^2(1-\alpha/2, p)$, and the upper limit is written as $UCL = \chi^2(\alpha/2, p)$ where $\chi^2(\alpha, p)$ - $(1-\alpha)$- percentile χ^2 - distribution with p degrees of freedom; α - a given level of significance.

Assuming that the i-th observation $\boldsymbol{X}_i$ does not depend on $\bar{\boldsymbol{X}}_m$ and $\boldsymbol{S}_m$ the statistics T_i^2 (for a fixed i) obeys Fisher's F-distribution with degrees of freedom p and m-p [11]. In this case, the lower control limit has the form

$$LCL = \frac{p(m-1)(m+1)}{m(m-p)} F(1-\alpha/2; p, m-p)$$

and the top one is,

$$UCL = \frac{p(m-1)(m+1)}{m(m-p)} F(\alpha/2; p, m-p)$$

where $F(\alpha; p, m-p)$ is the $(1-\alpha)$ percentile of the F-distribution with degrees of freedom p and m-p.

Checking the mutual independence of random variables can be carried out using the appropriate test [12, p. 612].

If the above assumption about the independence of observation $\boldsymbol{X}_i$ from $\bar{\boldsymbol{X}}_m$ and is not fulfilled, the specified equalities for LCL and UCL may be violated. As shown in [12], statistics T_i^2 (for a fixed i) has a beta distribution, $T_i^2 \sim \frac{(m-1)^2}{m} B\left(p/2, (m-p-1)/2\right)$ which can be correctly used in the case of individual observations when constructing control limits at the initial stage to check the statistical controllability of the process. In this case, the lower control limit is set as,

$$LCL = \frac{(m-1)^2}{m} B\left(1-\alpha/2; p/2, (m-p-1)/2\right)$$

and the top one is,

$$UCL = \frac{(m-1)^2}{m} B\left(\alpha/2; p/2, (m-p-1)/2\right)$$

where $B\left(\alpha/2; p/2, (m-p-1)/2\right)$ is the $(1-\alpha)$ percentile of the beta distribution with the parameters p/2 and (m-p-1)/2.

If the tables for the beta distribution are difficult to access, you can use the following relationship between it and the *F* - Fisher distribution:

$$\frac{\left[p/(m-p-1)\right]\cdot F(\alpha; p, m-p-1)}{1+\left[p/(m-p-1)\right]F(\alpha; p, m-p-1)} = B(\alpha; p/2, (m-p-1/2)$$

$$LCL = \frac{(m-1)^2}{m}\cdot\frac{\left[p/(m-p-1)\right]\cdot F(1-\alpha/2; p, m-p-1)}{1+\left[p/(m-p-1)\right]\cdot F(1-\alpha/2; p, m-p-1)} \quad \text{and}$$

$$UCL = \frac{(m-1)^2}{m}\cdot\frac{\left[p/(m-p-1)\right]\cdot F(\alpha/2; p, m-p-1)}{1+\left[p/(m-p-1)\right]\cdot F(\alpha/2; p, m-p-1)} \, .$$

In many cases, LCL is assumed to be 0 because any shift in the mean results in an increase in the T^2 statistic, thus allowing the LCL values to be ignored. However, T_i^2 it is sensitive not only to shifts of the mean vector, but also to changes in the covariance matrix of observations. If it strongly depends on the volume of observations, then this can lead to a violation of normality at small values T_i^2. To detect such deviations from normality, non-zero LCL values should be used. Statistics T_i^2 going beyond the control values UCL or LCL indicates a violation of statistical control at time $t = t_i$. The task of identifying an indicator or several indicators that caused this violation and the subsequent regulation of the process (phase 2) will be considered separately.

Detrending of the process. Let be $\{x_k\},(k=1,...,k_0)$, $x_k = x(t_k)$, $t_1 < t_2 < ... < t_{k_0}$ - the initial time series of some indicator of the investigated dynamic process. We will assume that, $k_0 = s\cdot N_s$ where s and N_s are positive integers. We divide the full time interval of observations $[t_1, t_{k0}]$ into intervals of length *s*. For each $\nu = 1,\ldots, \nu_0$, $\nu_0 = N_s$, we construct a polynomial regression $x = \sum_{r=0}^{m_\nu} c_r^{(m_\nu)} t^r$ of order $m(m=1,...,m_o)$. The best value m_ν^* of the degree of the polynomial of model is obtained by enumerating $m_\upsilon = 1,...,m_0$ the minimum sum of squares of residuals SSR (see details of the detrending algorithm in [13]). Then $x_k^{trend} = \sum_{r=0}^{m_\nu^*} c_r^{(m_\nu^*)} t_k^r$

$x_k^{stas.} = x_k - x_k^{trend}$ $k=(\nu-1)s+i$ $i=1,\ldots,s$ $\nu=1,\ldots,\nu_o$.

IV. Normalization of a multivariate random variable

To normalize d-dimensional vector of variables $x = \left(x_1,...,x_j,...,x_d\right)$, it is proposed [14] to use the vector of parameters $\lambda = \left(\lambda_1,...,\lambda_j,...,\lambda_d\right)$ and the transformed vectors

$$x_j^{(\lambda_j)} = \begin{cases} \dfrac{x^{\lambda_j} - 1}{\lambda_j}, \lambda_j \neq 0, \\ \log x_j, \lambda_j = 0; x_j > 0 \end{cases}$$

corresponding to the variables x_j $(j = 1,\dots,d)$. Wherein $\lambda_j (j = 1,\dots,d)$ are chosen so that the vector $x^{(\lambda)} = \left(x^{(\lambda_1)},\dots,x_j^{(\lambda_d)}\right)$ consists of independent identically distributed random variables satisfying a joint normal distribution $N_d(\boldsymbol{\mu},\boldsymbol{\Sigma})$ with a mean vector μ and a normal matrix $\boldsymbol{\Sigma}$. . In this case, the maximum likelihood function is written as $L\frac{1}{2}log|\Sigma|\sum_{i=1}^{d}(\lambda_j - 1)\sum_{i=1}^{m} log\, x_{ij\max}$, where x_{ij} is the j-th element of the vector $\boldsymbol{X}_i = \left(x_{i1},\dots,x_{id}\right)^T$ (i = 1,...,n), and ML is the estimate (maximum likelihood estimate) $\sum$ for the matrix $\boldsymbol{\Sigma}$. The function $(-L_{\max}(\lambda))$ minimum λ is determined by the coordinate descent method [13].

V. Conclusion

The construction of multivariate dynamic control charts for statistical control of the oil field development process has distinctive features in comparison with commonly used (onevariate or multivariate) CC based on quality control of products for its various indicators, but issued at a certain fixed time of control (that is, the task statics). Even a sample of observations, represented by the values of the same indicator of the production process, is a non-stationary time series, the mathematical expectation (trend) of which is generally characterized by a non-linear dependence on time, and only the residual time series obtained by detrending the initial time series is stationary a time series in which both the first moment (mathematical expectation) and the second moment (variance) do not depend on time. So, it is more accurate to investigate the statistical stability of the stationary remainder of the time series, and the variance of this series will characterize the stability band of the square deviation of the stationary remainder from its constant mean, which fundamentally diverges from the analysis using Shewhart's CC. Another feature of multivariate analysis, in comparison with the traditionally used sequence of onevariate control charts, is to take into account the interdependent influence of a set of features on the response of the process. Thus, the synergistic nature of the multivariate control of the dynamic process is revealed, in which various interactions of different signs are manifested.

Thus, the proposed methodological approach to multivariate statistical control of oil field development is of both theoretical and practical significance and can be applied in the statistical control of any dynamic production processes.

References

[1] Mirchink, M. F., Mirzajanzade, A. Kh., Zheltov, Yu. V. Physical and geological problems of increasing oil recovery. M. Nedra, 1975.

[2] Bagirov, B. A., Salmanov, A. M. Methodology for monitoring the development of oil deposits. Abstracts of reports. "Application of probable-statistical methods in drilling and oil and gas production." Baku, AzIU, 1991, p. 25-26.

[3] Wheeler, D., Chambers, D. Statistical Process Management: Business Optimization Using Shewhart Control Charts; - M .: Alpina Business Books, 2009.-409 p.

[4] Montgomery, D. C. Introduction to statistical Quality Control. 6th Edition. John Wiley and Sons, Inc. 2009.-734 p.

[5] Adler, O. V., Maximova, V., Shper, L. Shewhart's control charts in Russia and abroad: a brief overview of the current state (statistical aspects).

[6] Bendat, J., Pirsol, A. Applied analysis of random data. / Per. from English M., Mir, 1989.-540 p.

[7] Box, G., Luceno A. Statistical Control by monitoring and Feedback Adjustment. - John Wiley and Sons, Inc.

[8] Chua, M-K. , and Montgomery, D. C. Investigation and characterization of a control scheme for multivariate quality control // *Quality and Reliability Engineering International*, 1992, vol. 8, pp. 37-44.

[9] Johnson, R. A. and Wichern, D. W. Applied Multivariate Statistical Analysis. - Prentice-Hall, Englewood Gliffs, New Jersey, 1988.

[10] Mardia, K. V. Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness stadies // *Sankhya* B, 36,1974,115-128.

[11] Tracy, N. D., Young, J. C., Mason, R. L. Multivariate Control Charts for Individual Observations // *Journal of Quality Technology*, 1992, v. 24, No.2, p.88-95.

[12] Seber, G. A. F. Multivariate Observations. 2-th ed., 2004, John Wiley Sons, Inc. 686 p.

[13] Dzhanakhmedov, A. Kh., Dyshin, O. A., Javadov, M. Ya. Synergetics and fractals in tribology. Baku: "Apostroff", 2014. -504 p.

[14] Andrews, D. F., Gnanadesikan R. and Warner, J. L. Transformation of multivariate data // *Biometrics*, 1971, 27, 825-840.

[15] Ortega, J., Reinboldt V. Iterative methods for solving nonlinear systems of equations with many unknowns. M.: Mir, 1975.-558 p.