Statistical control of DOF maintenance condition

T Aleksandrova and E Gromakov
National Research Tomsk Polytechnic University (TPU), 30 Lenin Avenue, Tomsk, Russia, 634050

E-mail: tata09@mail.ru

Abstract. This work aims to develop and rationale statistical and control tools set to ensure correct interpretation of measurements and data received from cluster well in the automated mode. Robust procedure is offered for control chart interpretation that allows maximizing the determination of real faults and minimizing false alarms counts.

1. Introduction

DOF, Digital Oil Field is an automatic control system for oil and gas extraction that ensures the optimization of field maintenance using the information about its state collected from the continuously refined supercomputer geological and hydrodynamic models and equipment condition data [1]. The DOF term is based on the concept of intelligent control that is why digital oil field also referred to as intelligent.

Known realizations of that technology are directed onto automated control of unmanned energy resources extraction with decision-making center of maintenance based on objective data about field.

The aim of this work is to develop and study the composition of the statistical processing and control tools to ensure correct interpretation of measurements and data received in the automated mode from cluster well.

2. Shewhart control charts

Three methods can be used for statistical processing. The first one is based on Neyman-Pearson criterion and represented as Shewhart control chart [2]. It is historically the very first common tool to control the variable parameters of maintenance processes. The second one is based on the repeated application of Wald sequential analysis results [3]. It puts into practice in the form of control charts of cumulative sums. Finally, third approach of violations detection is based on exponential smoothing of statistical results (EWMA charts). Hereinafter the features of the use of the first two methods are considered.

Control chart in DOF as in Production Quality Control is a chart of regular measurements samples of some process indicator ***X***. As ***X*** can be used the output parameter of the energy resources extraction process, such as oil extraction and/or parameters characterizing the state of the equipment (for example, the temperature of the bearing assembly, vibration and etc.). The statistical data collected by telemetry can be combined in the form of samples diagrams on the map ***g(X)***.

Sample on this diagram is the sequence of n independent observations $x_1, x_2, \dots, x_n$ of indicator ***X***, related to some time period (day, month, year).

Suppose that $f(x, \theta)$ describes the distribution of random variable $\boldsymbol{x}$ at a certain value of statistical parameter (average value of sample, spread and others) θ.

Using the control chart the problem of detecting unnatural variability of observed parameter that may be a change of field operating mode. This variability is observed on the basis of changes of selected for chart parameter of distribution function θ of sequential observations.

The features of the energy resources extraction process and state of its technical equipment are the slow changes of variability in operating parameters trends.

We assume that the statistical properties of the time series that characterizes these states and the properties of the reasons that generates its changes remain unchanged or change slowly in the selected intervals of samples formation. Current control is reduced to monitoring of the regular sample observations and detecting of the observed time series θ properties changing from θ_0 to θ_1. The need to change the operational mode of the field may occur at unknown time t_0. We call this event as disorder of the operational process. Useful problems that are solved by the automated system of the DOF are the detection of the disorder at an early stage of its development, the timely control of the operating parameters and the prediction of the operational state of the field.

The generalized statement of the considered range of the practical problems related to the operational state of the field is characterized by random sequence $\{x_1, x_2, \dots, x_n\} = \{x_1^n\}$, that at the time moment t_0 changes its own properties that are uniquely determined by vector of parameters θ, $dim\{\theta\}$= r. Starting from the time t_0 vector of parameters becomes $\theta = \theta_1$. Moment of disorder t_0 is revealed by set criterion for the evaluation of sequence $\{x_1^n\}$, $N \to \infty$, at appearance of the next point x_n.

Each time for each new sample the hypothesis $\theta = \theta_0$, pointing to the absence of the unnatural variability is tested. We call this hypothesis as zero hypothesis and sign it H_0. Competing hypothesis H_1 suggests an unnatural change of the parameter $\theta = \theta_1$. The errors may occur when hypothesis is adopting.

Neyman and Pearson in their work indicated that the checking of the alternative hypotheses may reveal the errors of two kinds which are highly undesirable for making the decision about changing of the operational mode of the field. The error of the first kind is the declination of the hypothesis H_0 (no disorder) while it is true. The error of the second kind is the acceptance of the hypothesisH_0while the competing hypothesis is true H_1 (disorder).

Error of the first kind α risk while the operational parameters are processed may occur when operational state of the field is in statistically controlled state but the monitored parameters are out of control limits at random.

As a result, the actions that are generated by non-existent cause are called as "excessive regulation".

The error of α risk is equal to the probability of making a wrong decision based on the false alarm. Manager actions based on such kind of alarms lead to unnecessary costs.

Therefore, to statistical processing of operational information about the field from all options of the algorithms that detect the unnatural variability of the indicators that determine the current state and prospects of the field exploitation you need to choose the ones that provide the maximum probability of the correct detection of unnatural variability in case that the probability of the false alarm (α risk) not exceeds the predetermined value.

The error of second kind ($\beta\ risk$) occurs when the operational state of the field disorder and the control points do not indicate a lack of control. In this case the corrective action is not performed (lack of control). $\beta\ risk$ is equal to probability of the hypothesis H_1 acceptance when it is false. Sometimes $\beta\ risk$ referred to as the risk of real failure or the risk of missing signal.

It is necessary to minimize the probability of these errors through the statistical processing of such important information as the operational state of the field. To solve this problem we can us the criterions of the variability evaluation that are based on the Shewhart control charts.

Let us suppose that $\boldsymbol{X}$ is a normally distributed random variable with expectation μ and dispersion σ^2. Then the probability that deviation of random variable $\boldsymbol{X}$ from its expectation μ in absolute value less than the predetermined value $\varepsilon = \varkappa\sigma > 0$ can be calculated by the following formula:

$$P(|X-\mu|<\varkappa\sigma)=P\left(\left|\frac{X-\mu}{\sigma}\right|<\varkappa\right)=P\left(-\varkappa<\frac{X-\mu}{\sigma}<+\varkappa\right)=$$
$$=P\left(\frac{X-\mu}{\sigma}<\varkappa\right)-P\left(\frac{X-\mu}{\sigma}<-\varkappa\right).$$

The desired probability is

$$P(|X-\mu|<\varkappa\sigma)=\ \Phi(\varkappa)-\Phi(-\varkappa)==2\Phi(\varkappa)-1 \qquad (1)$$

This probability depends only on $\varkappa$. Values of $\Phi(\varkappa)$ for different values of $\varkappa$ can be found with help of special tables [4]. Particularly, using formula (1) we can calculate

for $\varkappa=1$: $P(|X-\mu|<\sigma)=\ 2\Phi(1)-1=2*0{,}84135-1=0{,}6827$;

for $\varkappa=2$: $P(|X-\mu|<2\sigma)=\ 2\Phi(2)-1=2*0{,}97725-1=0{,}9545$;

for $\varkappa=3$: $P(|X-\mu|<\sigma)=\ 2\Phi(3)-1=2*0{,}99865-1=0{,}9973$;

that is 99.73% quality characteristic data will fall in interval $(\mu-3\sigma;\mu+3\sigma)$.

The probability of the results falling out of the interval $(\mu-3\sigma;\mu+3\sigma)$ for statistically stable process is equal to 0.0027 and this value points to the promising application of the Shewhart control charts for evaluation of the operational parameters of the field.

3. ARL for Shewhart control charts

The probability of the operation disturbance recognition is called as sensibility of the control chart. In analysis of the control chart sensibility the average run length (ARL) of the controlled parameters measurement samples.

The mean length of the ARL series of the control chart is a calculated (expected) value of the samples that is necessary to process for the detection of the shift of process mean value. Usually ARL are calculated separately for the zero shift of mean controlled parameter (in-control ARL_0) and non-zero shift (out-of-control ARL_1).

Appropriate metrics of ARL for α error are defined as:

$$ARL_0=\frac{1}{\alpha},$$

and for β error ($\mu\neq\mu_0$) as:

$$ARL_1=\frac{1}{1-\beta}.$$

If there is no real change of process then ARL_0 value must be high.

Since every sample can be associated with time period then the length of ARL_0 may be considered as the prediction depth of the statistically controlled operational state of the field.

In case of statistically uncontrolled exploitation parameter (if the shift is significant) then it is necessary for ARL_1 to have a low value for the fast detection of the process state change.

In other words, when the real state of the process is unsatisfactory the decision rule should give the earliest possible alarm about process change. There can be used different combinations of criteria for determining the absence of control [5].

A significant disadvantage of Shewhart control charts is that they do not allow you to quickly find disorder of the process with a slight shift of the characteristic parameter.

Feature of the methodology of construction of Shewhart control charts is that they give the evaluation of the operational state on the basis of some accumulated array of samples.

4. CUSUM charts

Since Shewhart control charts only take into account the error of the first kind [2] then to improve the reliability of the disorder variability evaluation in addition to control chart we propose to use the Wald criterion [3] that was developed by Page in [6] in the form of sequential probability ratio test for two simple hypotheses:

IOP Conf. Series: Materials Science and Engineering **93** (2015) 012053 doi:10.1088/1757-899X/93/1/012053

H_0 (no disorder): $\theta = \theta_1$ and
H_1 (disorder): $\theta = \theta_2$,
where θ is an parameter of the probability distribution function $\omega(x_t/\theta)$ of the observation x_t.

To evaluate the hypotheses on each step of control the probability ratio is calculated as follows

$$\omega(x_t/\theta_2)/\omega(x_t/\theta_1).$$

If

$$B < \frac{\omega(x_t/\theta_2)}{\omega(x_t/\theta_1)} < A\,,$$

then there is no change of exploitation and we continue the observation of indicator ***X***.

If

$$\frac{\omega(x_t/\theta_2)}{\omega(x_t/\theta_1)} \geq A\,,$$

then the hypothesis H_0 is declined and H_1 is accepted.

If

$$\frac{\omega(x_t/\theta_2)}{\omega(x_t/\theta_1)} \leq B\,,$$

then H_0 is accepted.

Constants ***A*** and ***B*** are determined in [8] as

$$A = \frac{1-\beta}{\alpha};$$

$$B = \frac{\beta}{1-\alpha},$$

where α is the probability of the error of first kind; β is the probability of the error of second kind.

In practice, it is easier to calculate the logarithm of the ratio $\omega(x_t/\theta_2)/\omega(x_t/\theta_1)$.

Let the write it as $z_t = \ln\frac{f(x_t/\theta_2)}{f(x_t/\theta_1)}$.

The variable z_t is called as the cumulative sum (*CUSUM*). Through the statistical regulation the values of the cumulative sum z_t accumulates sequentially while adding of samples.

If $\ln B < z_t < \ln A$, then monitoring continues and indicator ***X*** is sequentially evaluated.

If $z_t \geq \ln A$, then the hypothesis H_0 is declined and H_1 is accepted.

If $z_t \leq \ln B$, then H_0 is accepted.

The evaluation of the operational state of the field on the basis of this algorithm is performed each time on the appearance of the sample and unlike the Shewhart control charts there is also performed the current control of the observed operational indicator.

The discussion of the used in construction of *CUSUM* mathematical principles can be found in paper of Woodall [7].

If there is a series of quality variable values $x_1, x_2, \ldots, x_n$, then the construction of *CUSUM* would be as follows:

$$S_1 = (x_1 - k)\,,$$
$$S_2 = (x_1 - k) + (x_2 - k) = S_1 + (x_2 - k),$$
$$\vdots$$
$$S_n = \sum_{i=1}^{n}(x_i - k),$$

where x_i is the value of the observed variable;
k is the constant that represents the standard value;
i is the number of the sample.

The calculated and plotted in order of appearance cumulative sums S_n form a *CUSUM* – chart.

The constant k can have any value but often k is equals to the expectation then the k is close to the nominal value of the process variable.

The decision rule then can be formulated as follows.

If the distance between the current value of the cumulative sum and the lowermost one of the preceding points on the graph is greater or equal than h then it is necessary to perform corrective actions because there is a shift of observed parameter mean value.

Thus, the parameters of *CUSUM*-chart are

n is the sample size;

k is the standard value;

h is the interval of the solution.

The shift of the observed parameter mean value may be expressed as $(\mu - k)$. In case of normal distribution of the parameter the shift can be standardized as follows

$$k_{\text{ст}} = \frac{(\mu - k)}{\sigma/\sqrt{n}} = \frac{(\mu - k)\sqrt{n}}{\sigma}.$$

Similarly, the interval of the solution

$$h_{\text{ст}} = \frac{h}{\sigma/\sqrt{n}} = \frac{h\sqrt{n}}{\sigma}.$$

Thus, if the mean value of the process parameter deviation from the nominal value increases then the *CUSUM* increases too. Similarly, if the mean of the process variable decreases then the graph tends downwards.

If we plot the graph of the cumulative sum of deviations from the nominal values for consecutive sample means then even small regular shifts of the mean value of the parameter (minor disorder of operating parameter) would gradually lead to the accumulation of a significant sum of deviations.

In order to establish the control limits in *CUSUM*-charts it was proposed to use the procedure known as the *V*-mask that sequentially moves on the graph after the plotting of the point for the latest sample. We may assume that *V*-mask is an upper and lower control limits for the cumulative sums. However, instead of being parallel to the center line these straight lines are converged at a predetermined angle to the right, thereby forming a shape like a lying letter *V*. Even if one point lies outside of the V then the process is suspected of being out of control.

This type of control charts is particularly well suited for detection of the small regular shifts of the operational parameter that may be overlooked in the application of the classical $\bar{x}$ Shewhart control chart. For example, when due to depreciation of equipment the process slowly "slips" out of control that results the violation of the regulated specifications. The usage of *CUSUM*-charts gives us the monotonically increasing (decreasing) chart of the cumulative sums of deviations from the planned specifications.

CUSUM-chart in comparison with Shewhart control charts reacts on small variability of the operational parameter faster and allows to quickly determine its origin.

As is shown in [7], ARL *CUSUM*-charts are more effective in comparison with ARL Shewhart control charts because they have high values for the processes with zero shift of the controlled parameter mean value and small values for the processes with significant shift of the process parameter from the standard [6].

Sequential state control of the field operational parameter using the both instruments allows us to decrease the number of errors in parameter variability interpretation that leads to increasing of statistical control robustness.

5. Conclusion

The above statistical control tools allow the DOF monitoring of operational state of the field with low level of decisions at least in three dimensions of real time:

1) in operative real-time of daily management of the field;
2) in retrospective archival real-time with depth of established by regulatory documents;
3) in the forecast time that is defined by the strategic objectives of the enterprise.

References

[1] Lo C 2013 *Making the most of the digital oil field Offshore Technology*
[2] State Standard of Russian Federation P 50779.42 - 99 (ISO 8258 - 91). *Statistical methods for Shewhart control charts. Intro.* 1999-04-15, 1994. Moscow: Standard Publishing House, pp: 99
[3] Wald A 2004 *Sequential analysis* (Mineola, Courier Corporation)
[4] Storm R 2001 *Wahrscheinlichkeitsrechnung, mathematische Statistik und statistische Qualitätskontrolle* (Leipzig, Fachbuchverlag Leipzig)
[5] Wheeler D and Chambers D 1992 *Understanding Statistical Process Control* (Knoxville, SPC Press)
[6] Page E 1954 *Continious Inspection Schemes* (London, Biometrika)
[7] W. H. Woodall, The distribution of the run length of one-sided CUSUM procedures for continuous random variables, *Technometrics* 25 (1983), no. **3**, 295–301
[8] Murdoch J 1979 *Control Charts* (London, Palgrave Macmillan)