Automation of information entropy assessment process for large production systems and their components

Y V Polishchuk and R V Chumakov
Orenburg State University, Orenburg, Russian
E-mail: Youra_Polishuk@bk.ru

Abstract. In the framework of this work, an assessment method is considered and an application is proposed that implements the ability to calculate information entropy for large production systems. The importance of informational entropy for large production systems and their components allows a group of decision-makers to assess the prospect of generating managerial decisions from the perspective of completeness of knowledge about the state of the system based on available factual data.

1. Introduction
The process of operating large production systems (LPS) is associated with the need for managerial decisions aimed at their effective functioning. By LPS, we mean large production systems consisting of a set of enterprises and organizations that function as a whole within the framework of one technological process.

During the operation of the LPS, factual data on its condition is transmitted to a group of decision makers (GDM) in the form of formalized information from SCADA systems and through electronic documents the content of which contains factual data. In this case, the receipt of factual data for GDM occurs with a certain delay. For this reason, there is a discrepancy between the actual state of the LPS and its display in factual data. In this case, the probability of making incorrect managerial decisions of GDM for LPS increases, which is due to the increase in informational entropy of LPS [1].

Filling support operational documentation (SOD) with new factual data is realized by means of pulsed information flows (IF), in which factual information is transmitted in the content of electronic documents. Each IF is characterized by a pulse pitch since the information on it enters the SOD after a given time interval. The duration of the IF pulse step is determined by GDM and can be adjusted if necessary, for example, in the event of non-standard situations on the LPS such as an accident, repair, etc.

The factual data stored in the SOD identifies the state of the LPS at a particular point in time. Among the SOD data, there are two main types of parameters. The first type is static parameters that do not change for the LPS depending on time. The second type is dynamic parameters. This kind of parameters is characterized by "obsolescence", i.e. the discrepancy between the real value of the LPS parameter at the moment and stored in the SOD, which is due to their dependence on time. To assess the discrepancy between the values of the dynamic parameters, GDM can use the obsolescence function.

During the implementation of the LPS management process, GDM forms an IF and determines for them the values of pulse pitch in such a way as to minimize the "obsolescence" of factual data and thereby prevent the adoption of incorrect management decisions. However, non-standard situations and various technological problems are associated with the operation of the LPS, which lead to a discrepancy
between the values of the real LPS data and their display in the SOD. In practice, this makes it difficult or even impossible for GDM to identify the state of the LPS at the time of the decision.

As an example, we consider a diagram of the informational entropy of the steady-state operation of a conditional LPS (Figure 1).

Data on the condition of the conditional LPS comes in the form of IF, the number of which is equal \(N_{IF}\), two of them are shown in the upper diagrams of Figure 1. According to the first and last IFs, the GDM receives data on the state of dynamic parameters having various functions of aging and pulse pitches. The initial IFs make various contributions to the total entropy of the LPS, which is shown in the lower diagram. For the remaining IFs, \(2, \ldots, N_{IF} - 1\) GDM receives information on the state of static parameters, which do not change over time, and therefore make a constant fixed contribution to the total entropy of the LPS. The threshold level of permissible information entropy, which defines GDM, is highlighted on the diagram of the total IF. In cases where the graph of the total informational entropy is below the acceptable level of entropy, the GDM stores a sufficient amount of relevant information to identify the state of the LPS and make correct management decisions. Otherwise, additional information about the state of the managed system is required.

Figure 1. Diagram of informational entropy of the steady-state operation process of LPS.

2. Problem formulation
The task of monitoring the informational entropy of LPS is to automatically keep its values within acceptable limits (threshold values).

To calculate information entropy, we apply the following formula:

\[
H_{\text{inf}} = -\log_2 \left( \frac{F_{st}(P a_{st}, K_{st}) + F_{dyn}(P a_{dyn}, F_{obs}, K_{dyn})}{2 + F_{st}(G P a_{st}(T_{op}), K_{st}) + F_{dyn}(G P a_{dyn}(T_{op}), F_{obs}, K_{dyn})} + 0.5 \right).
\]
where $F_{st}, F_{dyn}$ – functions that calculate the sum of the products of cardinalities of the sets of static and dynamic (taking into account the "obsolescence" function) parameters by weight coefficients, respectively;  
$P_{a_{st}}, P_{a_{dyn}}$ – sets of known, static and dynamic (with the time of their existence) system parameters, respectively;  
$F_{obs}$ – obsolescence function;  
$K_{st}, K_{dyn}$ – sets of weights corresponding to static and dynamic parameters;  
$GP_{a_{st}}, GP_{a_{dyn}}$ – functions generating sets of theoretically possible, respectively, static and dynamic parameters of the system for a specified period;  
$T_{op}$ – system operating time.

3. Implementation

As an example, we consider the process of calculating the informational entropy of a collector-beam collection system (CCS) for gas condensate field products [3].

The CCS includes wells, which are separate systems. The contours of the influence of CCS wells may intersect, and when transporting products along a loop, in some cases, “crushing” of wells may occur, leading to a decrease in the efficiency of operation of the CCS as a whole. The boundaries of the system under consideration are the gas-bearing layer and the block of input strands (BIS). The internal characteristic of the CCS is the gas pressure, since the boundaries of the CCS are characterized by known pressure values [4]. The reservoir gas pressure is characterized by the geology of the reservoir. The pressure on the BIS is set by the technological mode of operation.

When moving products from the formation to the BIS, static pressure losses occur in the near-well zones of the wells, when the gas rises to the wellhead and during friction of its wall, as well as in the loop [5]. Correct control of GDM can be implemented only when the structural parameters and characteristics of the system are known for the CCS. In the conditions of the Orenburg gas condensate field (OGCF), the design parameters of the CCS are not fully known for all production collection systems. The design parameters of the CCS relate to a variety of static parameters that characterize the completeness of knowledge about the state of the system.

The dynamic parameters of CCS include the values of all types of well pressures and the flow rates of gas, condensate and water for all wells connected to the production system. The resulting document on conducting hydrodynamic research of wells (CHR) contains factual information, including the values of reservoir and bottomhole pressures of the wells. Information about the flow rates of wells and the value of wellhead pressures comes to the SOD from the geological and technological report (GTR).

Under the conditions of the OGCF, for the dynamic parameters of the CCS we will use the following obsolescence function:

$$F_{obs} = \exp(-\sqrt{T}),$$

where $T$ – number of months from receipt of the parameter.

The choice of the exponential dependence as the function of “obsolescence” is due to a single exponential form of the base curves of the fall of the main operational indicators of the OGCF. For the dynamic parameter obtained in the current month ($T = 0$), the value of the obsolescence function is one ($F_{obs} = 1$). The significance of dynamic parameters is assessed by multiplying the value of the parameter weight by the value of the obsolescence function.

During the operation of the LPS, as a rule, the actual amount of factual data is less than the amount of the given GDM. For example, well logging in the conditions of OGCF should be performed for each of the production wells once a quarter, but the complexity of ensuring planned production volumes at the field does not allow stopping wells for well logging four times a year, since the well stops producing products for the duration of the research. It is worth noting that the design parameters are known in full not by all OGCF collection systems.
Thus, when making managerial decisions of GDM, it is necessary to take into account the magnitude of
information entropy for CCS. The latter can be implemented in an automated mode through the use of the “Information Entropy Assessment” program, which is a convenient tool for analyzing the state of
system objects for GDM. Let's consider it in more detail.

For the convenience of using the program, it is implemented through a Web interface using the JavaScript language [6] and can be used by GDM remotely using an Internet browser. The main window of the “Information Entropy Assessment” program is shown in figure 2.

Figure 2. The main window of the program “Information Entropy Assessment” with loaded data.

It contains control buttons for loading, saving and updating data, as well as a drop-down menu for selecting the desired well. The presented program window displays information on the value of information entropy for the current well, and also the general information entropy for CCS, taking into account the existing dynamic and all existing parameters.

Let us consider in more detail the initial data for calculating the informational entropy of CCS, which are presented using the JSON data format for convenience (figure 3).

The following information is defined in the JSON source file for CCS:

- obsolescenceFunction – obsolescence function;
- endDate – date of assessment of informational entropy for an object;
- relief – general static parameters (in the example, this is a relief);
- title – static parameter name;
- weight – static parameter weight.

Further in the document, in the data section, information is described for all the wells included in the CCS, including data on their dynamic and static parameters. The example describes the parameters for one of the wells CCS - k1_317.

The following information was determined by the dynamic parameters of the wells:

- title – dynamic parameter name;
- theoreticalPeriod – theoretical period between measurements, in months;
• weight – dynamic parameter weight;
• measurementDate – months in which the parameters are actually obtained.

The following information was determined from the static parameters of the wells:

• title – static parameter name;
• weight – static parameter weight;
• startDate – date obtained static parameter.

To calculate the value of the informational entropy of the CCS, when adding new parameter values, it is necessary to adjust the assessment date using the parameter in endDate.

In the example, the previously considered dependence for the obsolescence function is used (2).

Thus, the completeness of information for dynamic parameters is assessed by how accurately the sequence of practical measurements follows the theoretical requirements, taking into account the obsolescence of their relevance from time to time, and for static parameters - by their actual existence.

```json
{
    "obsolescenceFunction": "exp(-sqrt(T))",
    "endDate": "13.12.2018",
    "relief": {
        "title": "Relief",
        "weight": 20
    }
},
"data": {
    "k1_317": {
        "meta": {
            "startDate": "20.11.2010",
            "title": "K1-317 well"
        }
    },
    "dynamicParameters": {
        "GCW_Debit": {
            "title": "GCW debit",
            "theoreticalPeriod": 1,
            "weight": 1,
            "measurementDate": [4, 8, 20, 24, 40, 48, 56, 60]
        }
    },
    "staticParameters": {
        "wellSchematic": {
            "title": "Well schematic",
            "weight": 30,
            "startDate": "20.11.2010"
        }
    }
}
```

Figure 3. Example file with data, retrieved from SOD in JSON format (2-column text for compactness).

4. Conclusion

The described method for assessing informational entropy for LPS increases the efficiency of making managerial decisions due to the automated assessment of the completeness of knowledge about the system.

The program “Information Entropy Assessment” allows GDM to set parameters characterizing its state for an object, to determine their contribution to minimizing the value of information entropy, taking into account their obsolescence. As the advantages of using the program for GDM, it is worth noting the automatic calculation of the value of information entropy, as well as the ability to save and adjust the information of the source file.
References

[1] Petrov B N 1983 Selected works. T. 1. Theory of Automatic Control (Moscow: Nauka) p 432
[2] Polishchuk Y V 2015 Mechatronics, automation, control 6 396-401
[3] Volkov N M, Myheev A P and Kolev K A 1989 Gas Worker Handbook (Moscow: Nedra) p 296
[4] Polishchuk Y V 2007 Oil field 6 60-3
[5] Yermilov O M, Remizov V V, Shirkovsky A P and Chugunov L S 1996 Formation physics, gas production and underground storage (Moscow: Nauka) p 541
[6] Nicholas C. Zakas 2005 Professional JavaScript for Web Developers 2nd (New Jersey: John Wiley & Sons) p 670