Conditions choice for conducting experiments to determine the oil cooling time

A U Yakupov, D A Cherentsov and K S Voronin

Tyumen Industrial University, 38, Volodarskogo str., Tyumen, 625000, Russia
E-mail: ya.yakupov-azamat@yandex.ru

Abstract. The paper considers the option of processing the results of a computer experiment to determine the oil cooling time in a suspended oil pipeline. The authors of the article in previous works proposed a calculation model that allows simulating the process of oil cooling. As part of the design of the experiment, it is necessary to determine the factors affecting the oil cooling time in the pipeline, which will vary in the experiment. Empirical formulas are established, a regression analysis is carried out. The estimates of reproducibility dispersions are calculated. The adequacy hypothesis was tested using the Fisher criterion. Significant regression coefficients were established.

1. Introduction
When operating pipelines through which highly viscous oils are transported, it becomes necessary to determine the temperature of the product [1, 2]. The selected temperature will affect the operating parameters. The state of the piping system and soil are assessed: at high temperatures, a thawing process of permafrost may occur, which can lead to emergency situations [3].

The process of pipeline transportation of oil already demands increased responsibility, which includes non-stationary regimes. Among considered by the authors is an emergency stop of an oil pipeline with a highly viscous product inside the fluidity of which is highly dependent on the temperature of its state. Therefore, when transportation continues after a product reaches a critical temperature, it needs an initial force. A value that may exceed the limit allowable by the strength of the pipeline or the capacity of pumps. Therefore, it is important to determine the oil cooling time to the critical temperature [4-6].

2. Materials and methods
This article describes the experience of using the method of group accounting of arguments to build a mathematical model of parameters that affect the cooling time of oil. The output value may turn out to be independent of one or several input variables presented in Table 1.
In these conditions, the authors consider it appropriate to sort out the set of successively complicated candidate models. To do this, we apply the iterative multi-series algorithm of the argument group accounting method (AGAM) [7]. According to AGAM, the candidate models are tested sequentially by a given criterion; the most frequently used are polynomial support functions in the form of the Kolmogorov-Grabov polynomial.
\[ y = b_0 + \sum_{i=1}^{M} b_i x_i + \sum_{i=1}^{M} \sum_{j=1}^{M} b_{ij} x_i x_j + \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{k=1}^{M} b_{ijk} x_i x_j x_k , \]  

where \( \bar{x} = (x_1, x_2, ..., x_M) \) are the vectors of input variables; \( \bar{b} = (b_1, b_2, ..., b_M) \) are the coefficients of the summands.

**Table 1. Input Variables**

| No. | Parameter                      | Unit       |
|-----|--------------------------------|------------|
| 1.  | Oil temperature                | °C         |
| 2.  | Specific thermal capacity of oil | J/(kg•K)  |
| 3.  | Oil viscosity                  | Pa•s       |
| 4.  | Thermal conductivity of oil    | W/(m•K)    |
| 5.  | Oil composition                | %          |
| 6.  | Oil pipeline diameter          | mm         |
| 7.  | Insulation thickness           | mm         |
| 8.  | Pipeline Depth                 | m          |
| 9.  | Soil heat capacity             | J/(kg•K)   |
| 10. | Soil thermal conductivity      | W/(m•K)    |
| 11. | Soil density                   | kg/m³      |
| 12. | Soil moisture                  | unit fraction |
| 13. | Soil temperature               | °C         |
| 14. | Ambient temperature            | °C         |
| 15. | Wind speed                     | m/s        |

The vector of coefficients \( \bar{a} \) are found from the training set of values \( \bar{x}_n, \bar{y}_n \) (\( n = \overline{1,N}, N \geq M \)) using the least squares method.

It is possible to use the combined AGAM algorithm, which does not remain constant, but expands with each new one, complicating the structures of the next form:

\[ y = b_0 + b_1 x_i, \quad i = 1, 2 ..., M. \]

The number of models is selected according to the best criterion.

In the second series, a model of a more complex structure is selected:

\[ y = b_0 + b_1 x_i + b_2 x_j, \quad i = 1, 2 ..., F; \quad j = 1, 2 ..., M; F \leq M. \]

In the third series, models of a more complex form are selected:

\[ y = b_0 + b_1 x_i + b_2 x_j + b_3 x_l, \quad i = 1, 2 ..., F; \quad j = 1, 2 ..., F; \quad l = 1, 2 ..., M. \]

Continue increasing until the value of the minimum criterion decreases.

According to the author, the most convenient to use is the multi-series algorithm [8], in which the iteration rule remains the same for all series. We use in the first series a particular description of the type.

\[ y' = b'_0 + b'_1 x_i + b'_2 x_j + b'_3 x_l + b'_4 x_i^2 + b'_5 x_j^2. \]  

(2)

In the second series:

\[ y'' = b''_0 + b''_1 y_i + b''_2 y_j + b''_3 y_i y_j + b''_4 (y_i)^2 + b''_5 (y_j)^2. \]

In the third series:
\[ y''' = b_0''' + b_1'''y_1''' + b_2'''y_2''' + b_3'''y_1'y_1'' + b_4'''(y_1'')^2 + b_5'''(y_1''')^2. \]

The process continues by analogy.

Because it is impossible to find one single model that would have a distinctive external manifestation, one can take a polynomial of any complexity and for each conduct a regression analysis and find the values of the coefficients. And in each equation, the coefficients will be chosen so that the error on the observation interval will be minimal. The iterative multi-series algorithm of the method of group accounting of arguments used has an attribute of external complement, with the help of which it becomes possible to identify a single model. External addition will be the requirement of one result to another. If the result of the first and the requirements of the second do not match, the first will have to be transformed, which is the process of self-organization. To obtain an external supplement in our case, we divide the initial values into the training and test samples. The principle of self-organization of models will be that with the gradual complication of models, the values of internal indicators will gradually decrease, passing through their lows. External indicators will also decrease, passing through their lows, which will make it possible to determine the model of optimal complexity.

A necessary step is to determine the number of values in the sample on which they will be trained, and obtain the coefficients of the arguments. It is proposed that all the studied values of the cooling time be divided into 3 parts, 2/3 of the parts to be used to obtain the coefficients, and 1/3, to check the results. The standard deviation will be the test criterion.

To build each model, you must perform the following steps on [9]:

a. Create a matrix with input parameter values \( U' \).

b. Create a vector \( O' \) in which there will be cooling time values.

c. Enumerate all possible combinations of input variables from the matrix \( U' \) and find for each case the values of the coefficients of the polynomials of equation (1).

d. Calculate cooling time values using found coefficients of approximating polynomials.

e. Calculate the standard deviation of the residuals - the differences between the vector \( O' \) and the corresponding value of each interpolated polynomial according to the formula:

\[
\sigma_{ij} = \left( \frac{1}{n} \sum_{i=1}^{n} (Y_n - (y_n')_{ij})^2 \right)^{0.5}.
\]

Polynomials with the smallest variance values will be used to construct approximating polynomials of the second series.

f. Repeat steps c, d, e for polynomials of the second series.

g. The increase in the level of the series should be performed until the variance of the residuals of the model increases.

h. To confirm the adequacy of the constructed model, it is necessary to calculate the multiple correlation coefficient:

\[
R = \frac{\sum_{k=1}^{n}(\hat{y}_k - \bar{y})^2}{\sqrt{\sum_{k=1}^{n}(y_k - \bar{y})^2}} = \sqrt{1 - \frac{\sum_{k=1}^{n}(y_k - \hat{y}_k)^2}{\sum_{k=1}^{n}(y_k - \bar{y})^2}}.
\]

where \( \hat{y}_k \) is the calculated indicator according to the model;

\( \bar{y} = \frac{1}{n} \sum_{k=1}^{n} y_k \) is the average value of the indicator;

\( \frac{1}{n} \sum_{k=1}^{n}(y_k - \hat{y}_k)^2 \) is residual variance;

\( \frac{1}{n} \sum_{k=1}^{n}(y_k - \bar{y})^2 \) is variance of the effective attribute;

\( \frac{1}{n} \sum_{k=1}^{n}(\hat{y}_k - \bar{y})^2 \) is factorial variance.

It is carried out after finding the coefficients of the model to assess its suitability. The Fisher criterion is an indicator of suitability, if the calculated value of the criterion does not exceed the
tabulated one, then with the corresponding confidence probability the model can be considered adequate.

i. To check the consistency of the obtained dependence with the indicators, it is necessary to calculate the confidence intervals of the regression equations according to the formula:

\[
y_n \pm t_{0.99}(200 - 6)\hat{\sigma}\sqrt{1 + x_n^T(D^T D)x_n},
\]

where \(D\) is the factor value matrix; \(t_{0.99}\) is the value of the student distribution quantile for a confidence probability of 0.95; \(\hat{\sigma}\) is standard deviation of random errors.

The significance of the obtained coefficients is estimated. The coefficient is significant in cases where its absolute value is greater than the confidence interval. Information on the significance of the coefficients allows us to consider the possibility of simplifying the subsequent work by eliminating some of the factors; moreover, this information will help in the interpretation of mathematical models.

The above calculations can be represented as an algorithm in Figure 1.

\[
\begin{align*}
\text{Input parameters} & \quad \bar{x} = (x_1, x_2, \ldots, x_M) \\
\text{Training sample } 2/3 & \quad \bar{x}, \bar{y} (n = 1, N, N \geq M) \\
\text{Testing sample } 1/3 & \quad \bar{x}, \bar{y} (n = 1, N, N \geq M) \\
\text{Generation of combination of series equation} & \quad y' = b'_0 + b'_1x_i + b'_2x_j + b'_3x_i x_j + b'_4x_i^2 + b'_5x_j^2 \\
\text{Finding equation coefficients by least squares method} & \\
\text{Selection of the best polynomial variants with the smallest variance values} & \\
\text{The variance values did not increase compared to the previous series.} & \quad \text{The variance values increased compared to the previous series.} \\
\text{Verification of the adequacy and consistency of the constructed model} & \\
\end{align*}
\]

\textbf{Figure 1. AGAM algorithm}

3. Conclusion

The cooling rate of oil can also depend on the conditions of convective heat transfer [10] inside the pipeline, which can have a significant effect, but are not taken into account in this technique. For example, when convective heat transfer is taken into account, it is likely that the angle of inclination of the pipeline will affect the cooling time.
Thus, this technique allows determining the regression equation, evaluating the significance of the parameters, identifying significant parameters for the experiment to determine the cooling time of oil in a suspended underground pipeline.

References
[1] Yakupov A U et al. 2019 Evaluation of the effect of seasonal cooling devices on the time of safe shutdown of the pipeline News of higher educational institutions. Oil and gas 3 120-126
[2] Yakupov A U et al. 2019 Influence of structural features of thermostabilizers on the cooling time of oil in a stopped oil pipeline News of higher educational institutions. Oil and gas 6 140-148
[3] Kvasov I N, Shendaleva E V, Shitengauer O V and Zemenkova M Yu 2019 Prediction of environmental risks in industrial accidents at main and technological oil pipelines News of higher educational institutions. Oil and gas 6 103-117
[4] Voronin K S et al. 2018 Modernization of the individual device for temperature stabilization of the soil IOP Conf. Ser. Mater. Sci. Eng. 445 (1) 012011
[5] Yakupov A U, Voronin K S and Cherentsov D A 2019 Temperature condition of a stopped underground oil pipeline IOP Conf. Ser. Mater. Sci. Eng. 663 012013
[6] Markov E V, Pulnikov S A and Sysoev Yu S 2018 Methodology for calculating the safe stop time of underground pipeline with high pour point oil IJCET 1699-1705
[7] Ivakhenko A G 1981 The inductive method of self-organization of models of complex systems (Kiev: Science. Dumka)
[8] Adler Yu P, Markova E V and Granovsky Yu V 1976 Planning an experiment in the search for optimal conditions (Moscow: Nauka)
[9] Zemenkova M Yu and Chekardovsky S M 2016 Methodology of scientific research in the oil and gas industry. Monograph (Tyumen: TIU)
[10] Kutateladze S S 1979 Fundamentals of the theory of mass transfer "Fundamentals of the theory of mass transfer" (Moscow: Atomizdat)
[11] Zemenkova M Yu, Zemenkov Y D, Pimnev A L and Kurushina E V 2016 System of Controlling the Reliability of Hydraulic Machinery in Oil and Gas Facilities IOP Conference Series: Materials Science and Engineering 127 (1) DOI: 10.1088 / 1757-899X / 127/1/012055
[12] Zemenkova M Y, Shipovalov A N and Zemenkov Y D 2016 Mathematical Modeling of Complex Hydraulic Machinery Systems When Evaluating Reliability Using Graph Theory IOP Conference Series: Materials Science and Engineering 127 (1) DOI: 10.1088 / 1757-899X / 127/1/012056
[13] Chizhevskaya E L, Zemenkova M Y and Mayss N A 2020 Directions to Improve Effectiveness of Management of Management of the Industry Company Personnel IOP Conf. Series: Earth and Environmental Science 459 052093 DOI:10.1088/1755-1315/459/5/052093