UDC 622.244.442:519.233.5

Volovyk V. A., PhD student, specialty G16 Mining and Oil and Gas Technologies
Scientific supervisor: Khomenko V. L. PhD, Associate Professor, Department of Petroleum Engineering and Drilling
(Dnipro University of Technology, Dnipro, Ukraine)

APPLICATION OF REGRESSION ANALYSIS FOR THE OPTIMIZATION OF DRILLING FLUID FORMULATION

Drilling fluids are a critically important, yet costly, component in the overall well drilling budget. Their formulation optimization is a key engineering task aimed not only at the direct reduction of chemical reagent costs but also at ensuring a smooth and safe drilling process. An improperly selected or overly "cheap" formulation can lead to significantly greater total losses: damage to the productive formation due to high filtration, reduced well flow rates, and increased non-productive time (NPT) due to complications such as differential sticking [1]. Furthermore, formulation optimization is directly related to environmental and logistical costs; effective composition management and fluid reuse can reduce the total volume of waste (cuttings) by up to 30%, which significantly lowers disposal costs. Thus, true optimization seeks not the minimum formulation cost, but the minimum *total cost* of drilling operations, while ensuring compliance with all technological parameters.

The transition from traditional, often empirical approaches to formulation development to mathematically substantiated methods is an urgent requirement. Multiple Regression Analysis (MRA) serves as a powerful statistical tool that allows for the quantitative description of the relationship between the drilling fluid composition and its final properties . The essence of the method lies in constructing a mathematical model that links dependent and independent variables.

Dependent variables (Y) include key technological parameters and risks that need to be controlled: rheological properties (Plastic Viscosity PV, Yield Point YP, Apparent Viscosity AV), density, solids content, as well as critical quality indicators like fluid loss, filter cake thickness, and the differential-sticking coefficient (DSC) [2]. *Independent variables* (X) include the engineer's 'levers' – factors that can be controlled: concentrations of formulation components (bentonite, biopolymers, starch, salts, lubricants, weighting agent), as well as operational conditions (primarily, temperature).

The standard linear regression model takes the form $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon$. The main value of MRA lies in the interpretation of the regression coefficients (β_i) . Each coefficient quantitatively shows by how many units the dependent variable (e.g., plastic viscosity in cP) will change when the corresponding independent factor (e.g., viscosifier concentration by 1 kg/m³) changes, assuming other factors remain constant. This allows for a direct economic assessment: to calculate the 'efficiency per dollar' for different analogous reagents and select the one that provides the necessary property enhancement at the lowest cost.

Since chemical and physical processes in drilling mud are highly non-linear, simple linear models are often insufficient. MRA allows for this non-linearity by using polynomial models, for example, quadratic ones: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X_1^2 + \beta_5 X_2^2$. Such models are extremely valuable for optimization. Interaction terms (e.g., $\beta_3 X_1 X_2$) allow modeling *synergy* or *antagonism* between reagents (e.g., how adding salt affects the efficiency of a starch viscosifier).

Quadratic terms (e.g., $\beta_4 X_1^2$) model the *saturation point* (law of diminishing returns), showing that after a certain concentration, further addition of the reagent becomes ineffective.

Having a system of validated regression equations for each key property (PV, YP, density, DSC), the chemical optimization task is transformed into a classic mathematical programming problem. The *objective function* is the minimization of the total formulation cost:

Minimize(Cost) $= \sum(Price_i \cdot C_i)$. The *constraints* are the regression models themselves, which must satisfy the technological requirements: $PV_{model}(C_1, \ldots, C_n) \geq PV_{required}$; $Density_{model}(C_1, \ldots, C_n) = Density_{required}$; $DSC_{model}(C_1, \ldots, C_n) \leq DSC_{max}$. Solving such a system allows for finding a mathematically substantiated, cheapest formulation that is guaranteed to meet all technical specifications [3].

Although modern machine learning methods, particularly Artificial Neural Networks (ANN), often demonstrate higher accuracy in real-time property prediction [3], they function as 'black boxes.' Their advantage is the speed and accuracy of prediction, but they do not provide a clear understanding of the impact of each component . Multiple Regression Analysis, in contrast, is a 'white box': it provides an interpretable equation where each coefficient has a physical and economic meaning. That is why for the task of *a priori* formulation optimization, where understanding the trade-offs and economic impact of each component is necessary, MRA remains an indispensable, reliable, and highly informative tool.

References

1. Горбійчук М. І., Семенцов Г. Н. Оптимізація процесу буріння глибоких свердловин. Івано-Франківськ : Факел, 2003. 493 с.
2. Farghat M. S. A New Approach for Predicting Effective Mud Annular Viscosity During Drilling. *Mansoura Engineering Journal (MEJ)*. 2014. Vol. 39, Iss. 1. P. 75–86.
3. Elkatatny S. Real-Time Prediction of Rheological Parameters of KCl Water-Based Drilling Fluid Using Artificial Neural Networks. *Arabian J. Sci. Eng*. 2017. Vol. 42. P. 1655–1665. DOI: 10.1007/S13369-016-2409-z.