**Abstract**

Forecasting fluid flow is essential to support optimum development decisions of subsurface resources such as groundwater, geothermal, and oil and gas. Forecasting offers invaluable information for understanding complicated rock and fluid systems to improve well productivity (water, hydrocarbon, and heat), maximize resource recovery, and project economics. However, production forecasting depends on multivariate reservoir parameters and complex non-linear fluid flow behavior that is challenging to capture through conventional forecasting techniques.

In subsurface data analytics and machine learning, advances enable new methods and workflows to integrate widely available historical production data into multi-step time series forecasting. Current methods to forecast well performance with surrogate flow models achieve high correlation coefficients during testing. However, they are typically black-box models with low interpretability and without diagnostics to evaluate the quality of the resulting model. In addition, these methods are often limited to deterministic forecasts that ignore uncertainty and are often limited to one-at-time single-well analysis.

To address these limitations, we propose a novel workflow for a data-driven surrogate flow model for subsurface forecasting using an attention-based deep neural network architecture. The proposed workflow through the problem formulation allows the inclusion of static and dynamic predictor features and forecasting across multiple wells. To include location in the TFT-based surrogate flow model, we propose two static predictor features, and to improve model interpretability, we calculate SHAP values to determine the most important features at every time step. The results indicate that our proposed workflow can be applied over a diverse range of practical subsurface resource scenarios to improve accuracy, provide uncertainty models, and enhance learning from our models to improve subsurface resource decision-making. To test the interpretability and performance of our proposed data-driven workflow, we construct a realistic, fit-for-purpose multi-well production dataset using numerical simulation with scheduled events for training and testing for primary and secondary recovery scenarios.