

A Machine Learning-Based Workflow for Source Rock Thermal Maturity Estimation from Palynological Samples

Mustafa Al Ibrahim, Christian Cesari, Mokhles Mezghani

Saudi Aramco

Abstract

Estimating thermal maturity of source rocks is an important step in constructing a comprehensive understanding of petroleum systems, and therefore successful hydrocarbon exploration and production operations. Traditionally, the estimation is done by an expert by studying vitrinite reflectance or by estimating it from pyrolysis measurements. When pyrolysis is not viable, alternative methods are based on palynological samples/slides. Thermal alteration index is one metric used to quantify thermal maturity from palynological samples. This method relies heavily on human expertise to identify specific organic matter in the slide and quantify their colors. Because it is a manual procedure, there is room for improving the efficiency and accuracy of the analysis. This study presents a low-cost automated methodology for estimating thermal maturity, specifically estimating the thermal alteration index, using palynological sample images. The workflow utilizes image processing techniques and machine learning to construct a prediction model. First, a dataset of palynological sample images with known thermal maturity is collected. The organic matter in each image is extracted by color or gray thresholding. The histograms are truncated at low pixel values to remove overlapping organic matter. Red, green, and blue (RGB) histograms of the remaining organic matter are then compiled and fitted with mathematical functions, e.g., polynomial. The weights of the functions fitted are used as an input for a prediction model with thermal maturity being the output. Note that whereas thermal alteration index relies on the identification of specific organic particles, the automated workflow outlined in this study relies on the whole image, so manual identification is not required. The workflow is applied on a dataset of nineteen samples. Data augmentation is applied by subsampling and rotating images. The procedure outlined to extract the organic matter histograms are followed. Different machine learning algorithms, specifically tree-based regressors, show comparable results. In the unseen test data with coefficient of determination between prediction and estimation values of 0.95, 0.95, and 0.94 for CatBoost, Random Forest, and XGBoost algorithms respectively. Overall, the automated procedure can provide accurate and consistent thermal maturity estimation and help tackle the large number of available samples.