Topological Data Analysis of Marcellus Play Lithofacies

Andrea Cortis, AYASDI, Menlo Park, CA, USA

Summary

We present a study on the lithofacies characterization of the Marcellus shale gas formation. The data set consists of nine vertical wells, each with petrophysical logs of composition (quartz, calcite, clay, and total organic carbon) and elastic parameters (density, and compressional and shear velocity) along four geological sections, i.e., the Mahatango Formation, the Upper Marcellus, the Cherry Valley carbonate formation, and the Lower Marcellus members of the Marcellus Formation.

We successfully used a new mathematical technique known as Topological Data Analysis (TDA) to identify lithofacies groups in the vertical profiles that possess well-defined marginal distributions in the velocity-density plane.

Motivation

The definition of lithofacies in a vertical log profile is a complex multidisciplinary task that often rests on the professional experience of groups of petrophysicists, geophysicists, geologists, and sequence stratigraphers. Statistical and machine learning data-mining methods such as Principal Component Analysis, k-means and hierarchical clustering, Gaussian Mixture models, and Neural Network Classifiers are often used to make sense of vertical log profiles data. These methods, however, often produce less than satisfactory predictions and we find ourselves in want of more efficient and rigorous mathematical methods.

In this work, we consider nine vertical wells taken from the Marcellus gas shale play along four geological sections, i.e., the Mahatango gray shale, the Upper Marcellus, the Cherry Valley carbonate formation, and the Lower Marcellus. The petrophysical logs for these wells include composition (quartz, calcite, clay, and total organic carbon content) and elastic parameters (density, and compressional and shear velocity). Figure 1 shows the total organic carbon (TOC) content vertical profiles for these nine wells. Note that the absolute scale for the TOC content and the absolute top and thickness for each of the wells has been removed to ensure data anonymity. From the TOC vertical profiles, however, it is straightforward to identify the location of each of the producing layers. The total number of variables is seven (four for composition and three for elastic properties) and the total number of measurements for all the wells is 18,650.