Evaluation of a Machine Learning Approach to Detecting Channels in Seismic Volumes

Marc Servais, Estanislao Kozlowski, Graham Baines, Andrew Davies

Halliburton

Abstract

Channelized systems form important hydrocarbon reservoirs, and significant effort has been made to streamline their identification in seismic data. Here we apply machine learning to improve recognition of channelized elements in seismic surveys. A stratigraphic forward model specifically designed to simulate fluvial/deep marine meandering systems was used to simulate 3D grids with age, facies and grain size information. From these simulations, synthetic seismic volumes are generated using empirical petrophysical relationships to define velocity and density. Using the synthetic seismic volume as input and channel masks derived from the facies grids as labels, we adopt a supervised learning approach to train a machine learning model that can predict the presence or absence of channels in a seismic volume. We also discuss the use of data augmentation methods when training a model. When a large training dataset is available, results show that a machine learning model taking 3D seismic cubes as input produces a more accurate prediction than one based on 2D seismic slices. However, when the amount of labelled data is limited - for example, if training on real data that has been manually labelled - training 3D models in a reliable way becomes more challenging, e.g. due to overfitting. In this case, 2D models (which require less training data) can be a useful alternative. We examine the relationships between the size of the training dataset, model complexity, and prediction accuracy. For the case where the quantity of labelled data is limited, we explore the benefit of using multiple models trained on 2D slices of different orientations (cross sectional and map view). We also present a "2½ D" method based on using several 2D slices, this approach requires less training data than a 3D approach but yields improved results compared to 2D models. Our contribution is twofold: First, producing models trained on synthetic data that can be used to predict the presence of channels in seismic data, without the potential limitations of interpreter bias. Secondly, we show how to use synthetic data to estimate how much labelled data is required when training is based on real (non-synthetic) data - for 2D, 2½D and 3D models.