Using predictive analytics to unlock unconventional plays

Julie Vonnet¹ and Gudmund Hermansen²* introduce a multi-disciplinary, integrated reservoir modelling workflow that gathers data of different natures, sources and scale in order to predict sweet spots locations.

The dramatic expansion in computing power over the past two decades and the huge amount of data generated within organisations have led to a proliferation of new methods to identify patterns and trends among these large datasets.

Data mining and predictive analytics are cross-disciplinary approaches that consist of advanced mathematical and statistical methods that retrace patterns from petabytes of data. Such methods and algorithms are able to extract information, correlations and interplays and turn them into structured sets of interactions for predicting the behaviour of a system – even under unknown conditions.

Sales forecasts, credit checks for lending, insurance underwriting and disease prognoses are just a few application examples of data mining and predictive analytics within healthcare, finance and retailing, all being domains where decisions made are critical to the business results.

These advances in data mining and predictive analytics can also be applied to the upstream oil and gas industry and particularly to unconventional reservoirs. In many fields, success or failure is held in the balance by just a few key sweet spots (defined as reservoir rock with characteristics suggesting improved production).

In unconventional reservoirs, sweet spots recognition is essential to reducing uncertainty, high-grading acreage, and improving field economics. Furthermore, the weak price regime supported by the current production glut is leading operators to look at their data in new ways to determine how to best optimize production costs.

In this article, we introduce a multi-disciplinary, integrated reservoir modelling workflow that gathers data of different natures, sources and scale in order to predict sweet spots locations.

We will describe how a classification algorithm, such as the k-nearest neighbour (kNN), can be used to learn and predict patterns from datasets, either in time or in depth, and how Emerson’s reservoir modelling workflow, Roxar RMS can produce both 3D grid properties and average maps showing the location of potential sweet spots from a user-defined set of characteristics.

Through new predictive analytics tools and machine learning algorithms, we will demonstrate how data sets generated from unconventional reservoirs can be mined to their maximum potential to identify sweet spots and deliver improved decision-making on where to drill, what production strategies to adopt, and how to unlock the value of operator assets.

In order to do this, however, it is first necessary to provide an overview of predictive analytics, its benefits and the interpretation challenges faced by unconventional reservoirs. We will then demonstrate how Emerson’s new workflow can unlock sweet spots for higher returns.

Predictive analytics – definition, workflows and algorithms

As already mentioned, predictive analytics has become a key tool for organisations in predicting current behaviour in order to predict future outcomes – for example, whether a person will be able to keep up with his mortgage payments;
Many algorithms are available today to identify and learn underlying patterns and structures in data. Below are some of the most-used methods that can be used alone or in different combinations.

- **Regression** is probably the most common algorithm. From linear regression to regression splines, such algorithms evaluate the relationships between one dependent variable and one or several independent ones;
- **Decision trees** are scenario-based and show the outcome of a tree made of different, sequential decisions;
- **Cluster analysis** is a clustering method that groups items into categories that have similar characteristics;
- **Time series algorithms** measure data points at points in time and are able to predict future values based on previously observed values;
- **Ensemble models** use the results from a large set of models to analyse the uncertainty in the predicted values;
- **Support vector machines** are used to analyse data and recognise patterns that can be applied to classification or regression analysis;
- **Association rules** are designed to identify strong rules in databases;
- **Bayesian algorithms** and statistics focus on probability.

We focus in this article on one of the most common classification algorithms and its application on unconventional reservoirs – the kNN (k-nearest neighbour) algorithm.

### The k-nearest neighbour algorithm

Classification algorithms attempt to classify data to known, or predefined, classes. They are opposed in that sense to the unsupervised learning algorithms, where classes are unknown, like cluster analysis for instance. The use of a classification algorithm requires some characteristics of one class to be well defined, which is particularly relevant with sweet spots properties. Wells showing high production very often also have specific characteristics that can be used as training data points – a data set used to derive the classification of new points.

A typical predictive analytics workflow would tend to consist of two main stages and a number of processes as illustrated in Figure 2:

- The first phase is to gather patterns from data and derive rules from them. This step of data mining identifies underlying interplays among the pool of available data. Quality controlling the data to filter out outliers or missing values is crucial at this stage, since skewed predictions can have a negative impact later on.
- The second stage involves using known results to train a model against the available data with future results predicted through using these models. This is the prediction step, which can potentially include scenarios, uncertainty and assumptions, and is the basis for strategic decision-making. One must be aware, though, of the inherent inaccuracy of predictive models. No predictive model is 100% accurate and requires the handling of multiple scenarios to better capture uncertainties around the predictions.
Furthermore, the BP Energy Outlook 2030 Report predicts that growing production from unconventional sources of oil is expected to provide all of the net growth in global oil supply to 2020. While the US remains the clear leader in production from unconventional reservoirs, many parts of Europe, Africa, Australasia and South America (among other regions) are now embracing unconventional reservoirs.

Yet, despite their potential, operators still face significant challenges in making informed decisions across the prospect lifecycle of unconventional reservoirs with a need to navigate and understand the complex relationships between the vast terabytes of recorded field data as well as the fragmented workflow and different engineering technologies.

The data challenges in unconventionals
– A North American example
The Bakken Shale/Three Forks play covers 200,000 square miles across North Dakota and Montana with huge amounts of data being generated from over 4000 horizontal wells. In such a vast region, factors such as depth, thickness and total organic carbon (TOC) can vary wildly with the application of analytical tools vital for identifying sweet spots.

Figure 3
Illustration of kNN classification – from left to right: k = 1, k = 2, k = 3. In Figure 3, for example, the test sample (grey circle) should be classified either to the first class of blue circles or to the second class of red circles.

In Figure 3, for example, the test sample (grey circle) should be classified either to the first class of blue circles or to the second class of red circles.

Given the set of training data (characterizing either blue or red), the algorithm first calculates the distance between each data needing to be classified (grey point in this example) and the data set. The class assignment is then done depending on the set of k closest training points. Note that the algorithm is not bounded to two-dimensional spaces, it is well-defined in high-dimensional spaces and also has multiclass problems. Moreover, the distance used to measure the space between samples does not have to be the Euclidean distance, it could be a problem specific distance measure that emphasis or retains different types of prior knowledge.

When k=1, the grey sample is set to blue, if k=2, it can be either red or blue, if k=3, it is red. In case there is a tie (k=2 in our example), it is very common to make a random assignment but it is also possible to weigh the result based on a risk scenario. This is what is done when mapping the sweet spots, as will be explained in the last section of this article. Here, we will show how the kNN algorithm plays a key role in constructing 3D maps of sweet spots predictions with a corresponding level of confidence.

Before we examine the application of the kNN algorithm in greater detail, however, it is necessary to examine some of the key data and interpretation challenges facing unconventional reservoirs and how predictive analytics can play such an important role in addressing them.

Data and interpretation challenges facing unconventional reservoirs
Unconventional oil and gas production is revolutionising the upstream oil and gas sector and global energy supply today. According to Bobby Ryan, the VP for global exploration at Chevron, in the US alone shale gas and tight oil production has created more than two million jobs and added $2.4 trillion to GDP (World Oil).

Identifying sweet spots, for example, requires a detailed understanding of complex reservoir properties and how these influence the productivity of the wells. Inevitably, this involves large amounts of data – from the static (horizons and faults) through to lithological properties and organic content.

For the past few years, traditional approaches to mapping out and predicting reservoir behaviour have tended to combine geophysics, geology and reservoir engineering as well as predictive statistical methods to estimate reservoir properties – both at local and regional scales. These statistical methods are also used as a means of finding the most productive areas. Such processes, however, can also be time-consuming.
Predicting the locations of unconventional hydrocarbons and the reservoir’s behaviour requires multi-disciplinary teams working together with different skills but with an integrated vision and an integrated workflow. It also requires the effective and accurate identification of the huge amounts of data available.

Alongside more traditional approaches such as petrophysics, well correlation, structural and petrophysical modelling, predictive analytics comes to the fore. It provides a mathematical framework to take these huge, disparate datasets and classifies the reservoir into specific classes used for optimal wells placement. The next section will look at a specific predictive analytics workflow developed by Emerson.

**A new workflow for unconventional fields – mapping the sweet spots**

**The workflow principles**
- Identify highest quality wells and attributes;
- Correlate these attributes with 3D data to get the training pattern;
- Identify similar patterns in areas with low well control;
- Generate sweet-spots likelihood maps.

The new workflow, that takes place within Emerson’s reservoir modelling software RMS, consists of the analysis of measurements that classify samples into specific groups characterized by predictive patterns. Using this computational power on large datasets – composed of well data and seismic cubes within a single integrated workflow – can reap significant benefits for operators in unconventional and expensive. There are also specific challenges to identifying sweet spots:

- As already referenced, unconventional plays come with their own reservoirs, complexities and combination of properties that generate many terabytes of data. The development of unconventional plays can also be expensive due to the complexity of the reservoirs and the difficult-to-understand relationships between different variables as well as the techniques required to enhance permeability and to drain hydrocarbons. In addition, there are faults and complex fracture network distributions that are so crucial to accurate interpretation – not to mention other variables, such as thermal regimes, brittleness, and reservoir thickness.

- For a well to perform to its full potential in a shale gas field, it is also important for operators to remain in contact with the shale for as long as possible and, due to shale’s low permeability, expose the shale to the pressure drop that allows the gas to flow in fracture treatment programmes. This requires a complete understanding of the reservoir and, in particular, the identification and mapping of sweet spots.

- From an economics standpoint, there is also the cost of failure. Whereas conventional reservoirs can be plugged and abandoned if the relevant zone of interest is considered uneconomic (often before too many financial resources have been committed) operators in unconventional fields can normally only make a full field evaluation once sweet spots have been identified and the well brought into production. It is therefore crucial to accurately identify sweet spots and predict reservoir behaviour beforehand.

---

Figure 4: An overview of the 3D geomodel built to integrate data measured across the area of interest – a facies property is shown here.
required here is to integrate all these data into one geomodel, which can be divided into sub-areas or not, depending on the global size of the studied area.

Production data is also really valuable but their variations in time sometimes make their integration complex. To avoid this, it can be useful to deal with the data per stages, for instance at the frac stages. Production data can be translated into well properties, reflecting the productivity of each well across time and depth.

The importance of a 3D geomodel
It is also important to stress the need to build a 3D geomodel either in time or in depth, bringing together seismic interpretation, well correlation and properties modelling. This is a vital step as it enables the kNN algorithm to be run on a full 3D model, rendering both 3D grid properties and average maps that show the location of potential sweet spots. Filtering dependent properties beforehand,

reservoirs as well as negating the weaknesses and limitations of traditional workflows.

The goal of this workflow is to predict the potential interest of areas that are located far from the wells by identifying sets of indicators that show clear trends that correlate with the best-producing wells. These most relevant properties can be used as a training pattern and the classification of samples can be expanded to the entire target region by exploring the nearest neighbourhood of each sample to classify all points according to their similarities. In the example presented below, data has been used to identify the most promising areas based on existing well information and the various rock properties available.

Selection of the data
Relevant data sets can be log measurements, 3D grid properties as well as seismic data. The only preliminary step
Two properties are generated after the algorithm is run. One reflects the classification of the area of interest into two different classes: a potential sweet spot versus a non-sweet spot. The probability of the class to be correctly computed is then calculated providing indications as to the quality of the prediction. This is particularly important during risk assessment phases in which several scenarios can be envisaged as a prelude to decision-making. Figure 6 illustrates this process.

**Seeking sweet spots and optimizing crucial decisions**

Once geologists and reservoir engineers have generated these maps, it is possible to use them to plan new wells, predict production and even calibrate these models using existing production history. Figure 7 illustrates future target locations.

**Conclusions**

As unconventional resources become more and more prevalent, reservoir modelling is playing a crucial role in ensuring their economic viability. As part of this, emerging workflows are developing that combine the latest innovative technologies in geosciences alongside the most advanced predictive modelling and analytics methods. This article demonstrates that through a step-by-step approach and the use of innovative predictive analytic tools, many of the challenges in shale can be addressed and sweet spots can be accurately identified for future reservoir development.

This new workflow can unravel the most complex interactions between static and dynamic data to create a complete picture of unconventional fields and predict future behaviour. By combining traditional multi-disciplinary workflows with the latest data-mining developments, the relevance of geomodels can be enforced; classification algorithms can be used to capture complex interplays; and predictive analytics can optimise investments.

Just as other sectors have benefited from predictive analytics, it is encouraging to see that the upstream oil and gas sector is now doing the same, enabling operators to enhance recovery and unlock the value of their assets.

**References**


International Conference on Data Mining (ICDM) [2006] Top 10 algorithms in data mining.