

Introduction to Modern Intelligent Data Analysis

1.1 Introduction

During 2020-21, the oil and gas industry has been faced with one of the most difficult challenges in its entire history. The combined impacts of a global pandemic and decreased demand have resulted in a plunge in oil prices. The collapse has pressured companies to take a serious look at their operations to reduce capital and operational expenses. Therefore, digital transformation initiatives may be launched to optimize production, reduce costs, and minimize environmental impacts of hydrocarbon recovery. By moving to the digital world, companies can also create additional revenue streams to help the business bottom-line.

Intelligent data analysis is the backbone of any digital transformation initiative aimed at optimizing engineering workflows and reducing costs. With access to rich datasets and decades of expert engineering knowledge, the oil and gas industry is in an excellent position to benefit from these algorithms and workflows. The increase in popularity of these methods is evident from the number of papers published in conferences and journals as can be seen in Figure 1.1.

However, applying advanced machine learning algorithms is not without pitfalls and risks. Hajizadeh (2019) performed a SWOT analysis of ML application in the oil and gas industry. For instance, analytics-based solutions must be repeatable and robust to initial starting conditions. Additionally, several industry and organizational-related issues may deaccelerate innovation efforts. The risk-averse nature of our industry and slow and top-down management style can stagnate data science projects. Figure 1.2 summarizes some of these challenges based on performing literature review and expert interviews (Baier *et al.*, 2019).

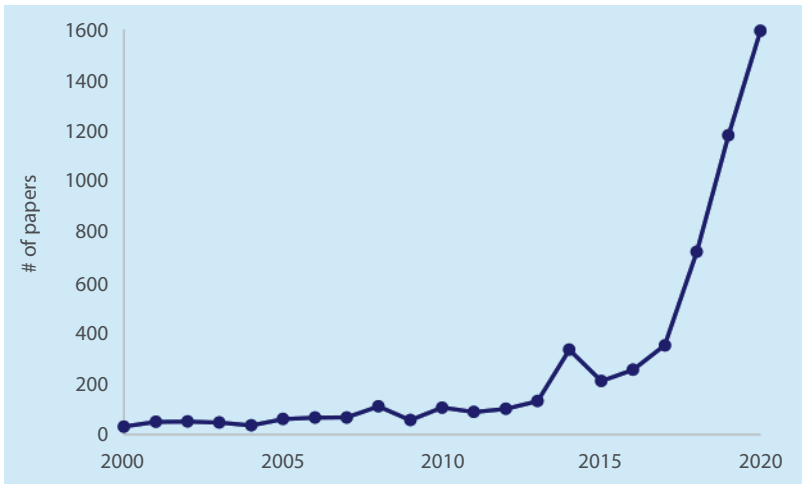


Figure 1.1 Explosion of number of papes in OnePetro between 2000 and 2020 on the top of machine learning and related topics (Baier *et al.*, 2019).

Pre-Deployment	Deployment	Non-Technical
Data structure <ul style="list-style-type: none">++ Data quality++ Data quantity++ High dimensionality in data++ Imbalanced data♦ <i>Encrypted training data</i>	<ul style="list-style-type: none">++ High-frequency data++ Concept drift/data drift	
Implementation <ul style="list-style-type: none">++ Data collection++ Data preprocessing++ Transfer learning♦ <i>Technical debt</i>	<ul style="list-style-type: none">++ Ongoing data validation++ Ongoing result validation++ Robustness♦ <i>Automated model updates</i>	
Infrastructure <ul style="list-style-type: none">++ Computational effort++ Energy consumption	<ul style="list-style-type: none">++ Deployment infrastructure++ Scalability	<ul style="list-style-type: none">++ User-friendly tools
Governance <ul style="list-style-type: none">♦ <i>Data management</i>++ Data privacy protection++ Anti-discrimination validation for deep neural networks		<ul style="list-style-type: none">++ Legal requirements++ Result transparency and interpretability++ Trust
Customer relation <ul style="list-style-type: none">♦ <i>Domain knowledge</i>		<ul style="list-style-type: none">♦ <i>Expectation management</i>♦ <i>Customer/Result communication</i>♦ <i>Standardization of terminology</i>
Economic implication		<ul style="list-style-type: none">♦ <i>Real-world value of ML</i>♦ <i>Business impact of ML</i>♦ <i>Creating digital services with ML</i>++ Evaluation metrics

♦ Interviews ♦ Literature *italic* Challenges only mentioned one time (either literature or interviews)

Figure 1.2 Machine learning challenges identified in interviews and from literature.

A machine learning pipeline starts by defining business objectives of the exercise. For example, it is imperative to take a note of right questions to ask from data and agree on project success criteria. It is also important to select an infrastructure and architecture that can handle the scope of

the project and deal with multiple development languages and systems. Then we need to gather or create training data. This is followed by preparing and massaging the dataset to make it suitable for machine learning algorithms. We follow the pipeline by starting the experimentation stage. Experimentation refers to preparation of data, selection of algorithm and validation and verification of models. Once we have a model with an acceptable performance, we need to deploy it. Operationalization is concerned with streamlining the process of productionizing models. The focus is on processes used for deploying models and the subsequent utilization and monitoring of measurable services that are efficient and fault-tolerant. This is, most often, the challenging part of a machine learning project in industrial settings, and where most companies fail in delivering value for the money spent on data science initiatives.

A modern machine learning pipeline, as illustrated in Figure 1.3, differs from a traditional one by adding machine learning operations (MLOps) and Continuous Integration Continuous Delivery (CICD) processes. MLOps is an engineering discipline that aims to make the process of building, versioning, deploying, monitoring, and updating ML models, fast, reliable and reproducible. As you can see in the figure below, traditionally machine learning projects follow a linear pipeline. They start by defining objectives and move through a series of experiments to fine-tune the algorithms. The exercise is concluded by putting models to production, albeit without a clearly defined and automated mechanism to make future adjustments. MLOps pipeline resolves this issue by providing an opportunity to monitor the system and make changes in data collection, experimentation, and operationalization stages. For instance, this is useful for solving issues such as concept drift, where performance of a machine learning model deteriorates as statistical properties of incoming data and relationships between input and output parameters change. MLOps adds value by systemically organizing the life cycle of machine learning and reducing risks to AI-driven enterprises. It facilitates the collaboration

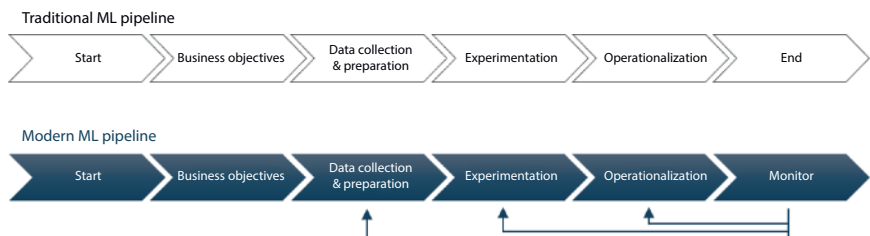


Figure 1.3 Modern machine learning pipeline (Hajizadeh, 2019).

between data scientists by having a unified model versioning and control story. It also tracks the performance of models after deployment, alerting and assisting data scientists in making necessary adjustments.

1.2 Introduction to Machine Learning

Arthur Samuel states “Machine Learning is a field of study that gives computers ability to learn without explicitly being programmed.” In 1959 Samuel wrote a Checker playing program which could learn over time (Sharm, 2019). This was the first defined and applied less formal machine learning program in the history. Four decades after Samuel, Tom Michel defined a formal version of the definition in 1999. He said, “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E ” (Sharm, 2019).

Machine learning, as a subset of AI (Figure 1.4), is a collection of algorithms that can transform data into intelligence and insights by learning through experiences and automated analytical model building. Machine learning is closely related to fields such as statistics, probability theory, data mining, pattern recognition, artificial intelligence, adaptive control, and theoretical computer science. They belong to a class of methods wherein solutions are primarily inferred from data rather than physics-based models (Data Flair Company, 2020; Michie *et al.*, 1994; Harrington, 2012; Alpaydin, 2014).

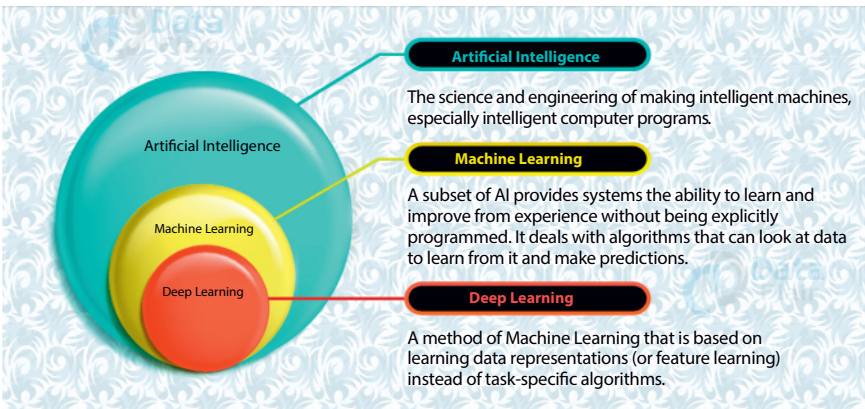


Figure 1.4 Machine learning infographic, data flair company (2020).

Over the last decade, computer processing power and machine intelligence advancements have resulted in a larger wave of emerging big data applications in industry including self-driving cars, facial recognition, text prediction, and natural language processing. A new age of information has required machine learning and artificial intelligence solutions to effectively utilize the massive streams of data accessible today (e.g., Smola and Vishwanathan, 2008; Le *et al.*, 2008).

As we see in Figure 1.5, machine learning algorithms can be generally categorized into supervised learning (i.e., regression, decision tree, random forest, and gradient boosting) [Singh *et al.*, 2016; Dey, 2016; Osisanwo *et al.*, 2017], unsupervised learning (i.e., k-Means, Hierarchical Clustering) [Kassambara, 2017; Chen, 2018; Usama *et al.*, 2019], semi-supervised learning (i.e., Deep Belief Networks (DBN) [Oliver *et al.*, 2018; Leslie *et al.*, 2018], Reinforcement learning (rewards from a sequence of actions) [Osa *et al.*, 2018; Kulkarni *et al.*, 2016], ensemble learning [Zhang and Ma, 2012], and instance-based methods [Aha *et al.*, 1991]. In this figure, we have included neural networks as a separate category due to their increasing popularity and applications in other ML classes such as supervised learning and reinforcement learning. In addition, Inductive Learning is a function in the form of data (x) and the output of the function ($f(x)$) [Snoek *et al.*, 2012; Madala, 2018].



Figure 1.5 Summary of machine learning and algorithms.

There are three key elements for machine learning models. These are representation (i.e., decision trees, neural networks, support vector machines, and model ensembles) [Seko *et al.*, 2017; Wöber *et al.*, 2018;], evaluation (i.e., accuracy, prediction, cost, margin, and entropy k-L divergence). [Gilpin *et al.*, 2018; Cai *et al.*, 2018], and optimization (i.e., combinatorial, and convex) [Luo *et al.*, 2018; Olson and Moore, 2019]. We included a few algorithmic examples in each category of machine learning in Figure 1.5. However, there are many more algorithms available, and new ones are being developed every year. Figure 1.6 includes more than 60 examples of machine learning algorithms. Choosing the best algorithm for machine learning systems depends on the type of a problem which experts are solving, available computer sources, and the nature of the data [Brownlee, 2019].

One of the supervised learning problems is a classification problem. In this problem, two or more classes can be chosen for providing probabilities for each class (Sen *et al.*, 2020). A clustering problem is a model to find the groups of the similar data points under an unsupervised learning problem which can be seen in Figure 1.7 (Oliver *et al.*, 1996).

As can be seen in Figure 1.8, two regression problems (Linear and Logistic) are used to predict a number in supervised learning problem [Joshi, 2020; Knowledgehut.com, 2020]. The outcome of dependent



Figure 1.6 A framework for understanding machine learning algorithms (Brownlee, 2019).

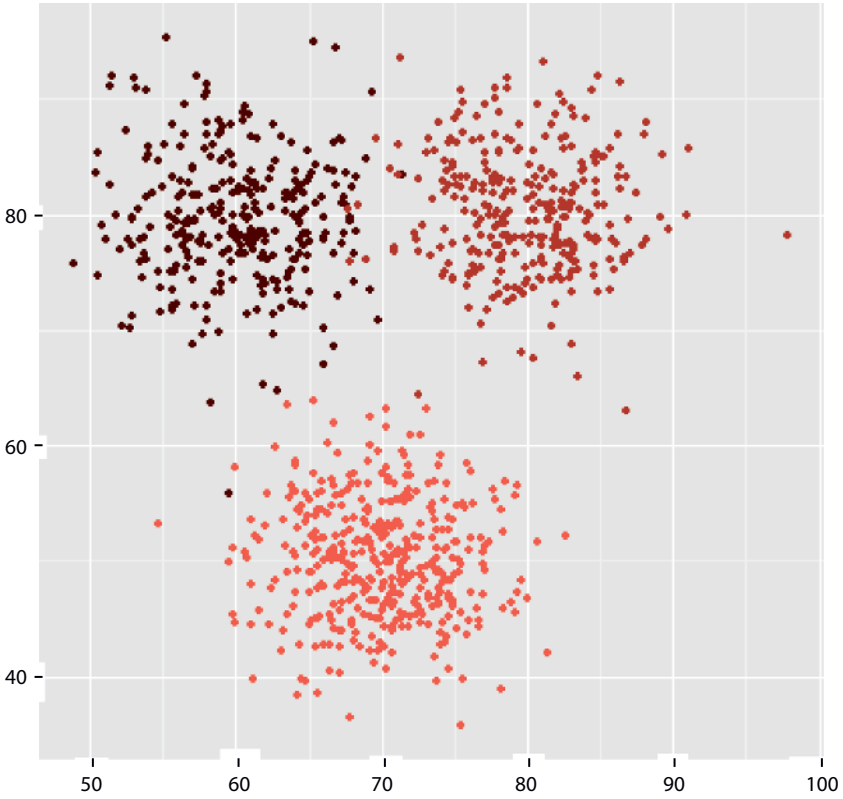


Figure 1.7 Machine learning clustering (Data Flair Training, 2020).

Linear Regression vs Logistic Regression

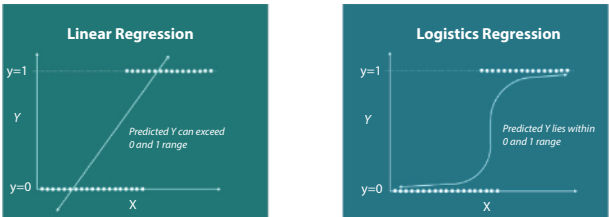


Figure 1.8 Comparison of linear and logistic regression methods (Knowledgehut.com, 2020).

variable is continuous in linear regression problems and has any one of an infinite number of possible values [Joshi, 2020].

1.3 General Example of Machine Learning

In the last decade machine learning has been used to increase productivity and reduce time [Witten *et al.*, 2005; Kashyap, 2017]. The exponential growth of technology required the use of better tools to understand the data that we had in the past and will have in the future. For this reason, an intelligent machine needs to be built to use AI technology. Software experts can write a code to perform certain tasks; however, hardwiring intelligence is the difficult part. The best way to handle this issue is to use technology to create machines to learn things themselves. This self-learning ability mechanism from an input can do the hard work for us. Nowadays, it is easy to hear about newer machine learning applications such as self-driving cars, game-playing machines, automatic language translation machines, and automatic speech to text systems (Price, 2019; pwc.com, 2020) (See Figure 1.9).

A few examples of machine learning applications are provided in Kubat *et al.*, 1998; Abdullah and Zhang, 2010; Hajizadeh, 2019; Gabralla *et al.*, 2013; Langley and Simon, 1995; and Provost, 2011. For example, mining of databases to increase automation may include web-click data for user experiences such as medical records and biological data (Das *et al.*, 2015; Kargupta *et al.*, 2008). Other examples include nonprogrammable

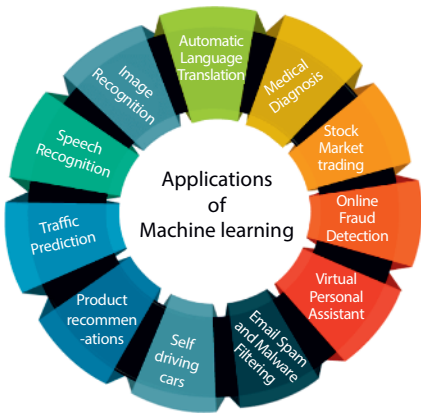


Figure 1.9 Application of machine learning and various applications (pwc.com - A, 2020).

application models in which tasks cannot be programmed such as autonomous driving, recognition tasks, natural language processing and computer vision (Wah *et al.*, 1993). The recent advances in the computerized systems will allow reservoir engineers to analyze the data that are collected from all the wells and help with the necessary predictions (Abiodun *et al.*, 2018).

1.4 E&P Examples of Machine Learning

Machine learning applications has played a major role in the oil-gas industry. Such applications date back at least two decades. Among them are: Aminzadeh *et al.* (2000), Poulton (2001), Wong *et al.* (2002), Nikravesh *et al.* (2003), Sandham and Leggett (2003), Russel *et al.* (2003), Devi *et al.* (2004) and Russell (2004), Aminzadeh (2005), Aminzadeh and de Groot (2006) and Mohaghegh (2011). Subsequently many other investigations helped contribute to the discovery of new reservoirs with high predictive analytics and soft computing. As an example, reservoir characterization, accurate prediction of reservoir behavior is an important aspect of reservoir optimization and management to change the control parameters. Before machine learning technology, computing resources were used to run simulators for significant data, even though this significant output limits the

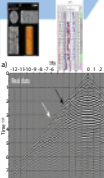
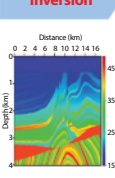
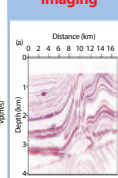
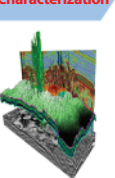
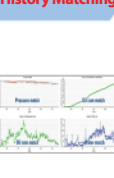
	3D Seismic/Well Logging/				Reservoir Simulation	Surveillance/4D Seis
Work	Measurement	Synthetic model Inversion	Subsurface Imaging	Interpretation/ Characterization	Simulation/ History Matching	Infill Drilling Production/EOR
						
Application	Compressive sensing	Elastic/Anisotropic Full Wave Inversion	Reverse Time Migration	Data Integration/Digital Twins/Visualization	Production forecasting Economics	Reservoir Monitoring Real Time Dec. Making
Challenge	Expensive	5 days with CPU Memory limitations	Takes time and effort	Nonunique Solutions	Trial and error	Optimize well location increase recovery factor
HPC Value	Realtime	Less than a day	Seamless	Collaborative Intr. Char.	Efficiency	Real time feedback
AI/DA Value add	Intelligent sensing smart processing Utilize experience	Pattern Recognition Ability to create alternative modles fast	Image updating with Data Mining Deep Learning	Data Mining method Big Data / 4v concept	Model Validation Digital Twins, Proxy models	Reservoir surveillance/ Visualization

Figure 1.10 Value addition of AI and HPC in different stages of exploration and production (Aminzadeh and Ikuta, 2021).

simulators' applicability for problems which are required more evaluations (Chiazor, 2018). Memon *et al.* (2014) calculated bottom hole pressure in reservoirs by using radial basis neural networks, and more examples can be found in Larue and Hovadik (2006). Also, see Shahkarami and Mohaghegh (2020) on applications of smart proxies for subsurface modeling.

In general, AI in conjunction with data analytics and high-performance computing adds value to the exploration and production (E&P) process. Figure 1.10 highlights this assertion with specific applications and challenges at all stages including data acquisition, modeling, imaging, interpretation/characterization, simulation/history matching and infill drilling/production/EOR. In separate rows some of the value additions associated with AI/DA as well as HPC are included.

1.5 Objectives of the Book

This book is aimed at covering both fundamentals and application domains of intelligence data analysis and machine learning algorithms and workflows. It provides an end-to-end story about where and how machine learning can be used to augment the knowledge of engineers working in the oil and gas industry and increase their productivity.

- 1) Learn the basics about advanced analytics and common machine learning algorithms
- 2) Get familiar with application domains in the entire value chain of oil and gas operations

1.6 Outline of Chapters

In Chapter 2, we discuss the human computer interaction and its role in machine learning. We look at different visualization and interaction frameworks to better understand the algorithms used for data analytics and how to utilize them to interactively work with these methods. The tools can be used for teaching concepts, assessing architectures, debugging and improving models, and visual explanation. Visualization and interaction tools must be tailored to user roles and experience levels in machine learning. As shown in Figure 1.11, users can interact with all stages of a machine learning project: initial data preparation and analysis, experimentation, and operationalization. For instance, during the initial data preparation, a user may visualize the correlation between input variables and eliminate

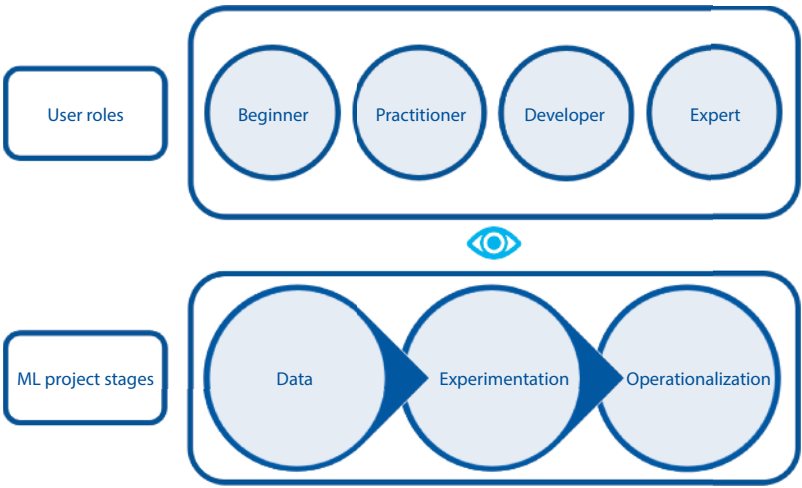


Figure 1.11 User roles and interaction with various stages of a machine learning project.

some of these variables. During the experimentation phase, the interaction may focus on modifying the tuning parameters of the ML algorithm based on visualization of performance metrics.

Chapter 3 focuses on neural networks, their similarities with biological neurons, their formulation, and various types of this machine learning algorithm. Figure 1.12 shows a simple classification of neural network architectures that differ in the way artificial neurons connect and link inputs and outputs.

We give a brief explanation of the structure of biological neurons. We introduce the main concepts behind artificial neurons, and different types of summation functions and activation functions. We continue with various types and structures of artificial neural networks, and different learning mechanisms. This chapter also discusses the pre-processing of ANN input data, and dimensionality reduction. We also provide several examples of

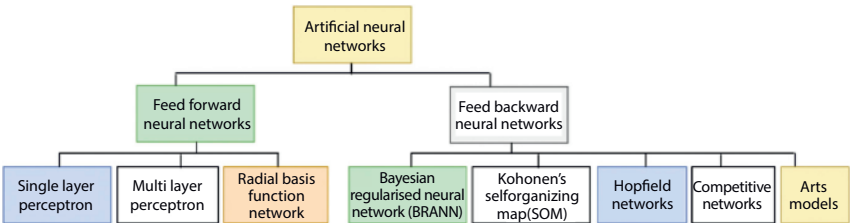


Figure 1.12 Review framework for ANNs (Abiodun *et al.*, 2018).

the ANN applications in various oil and gas applications, with many more to come in the later chapters that focus on many case histories involving ANN and other AI techniques.

Moving to Chapter 4, we introduce the fuzzy logic concept and how it can be used to represent incomplete and uncertain information. Concepts such as fuzzy arithmetic, fuzzy clustering, fuzzy kriging, and fuzzy systems are discussed in this chapter. The last part of this chapter focuses on building expert systems with fuzzy rules and demonstrates how it can be used to build a practical system for segmentation of a seismic section. Figure 1.13,

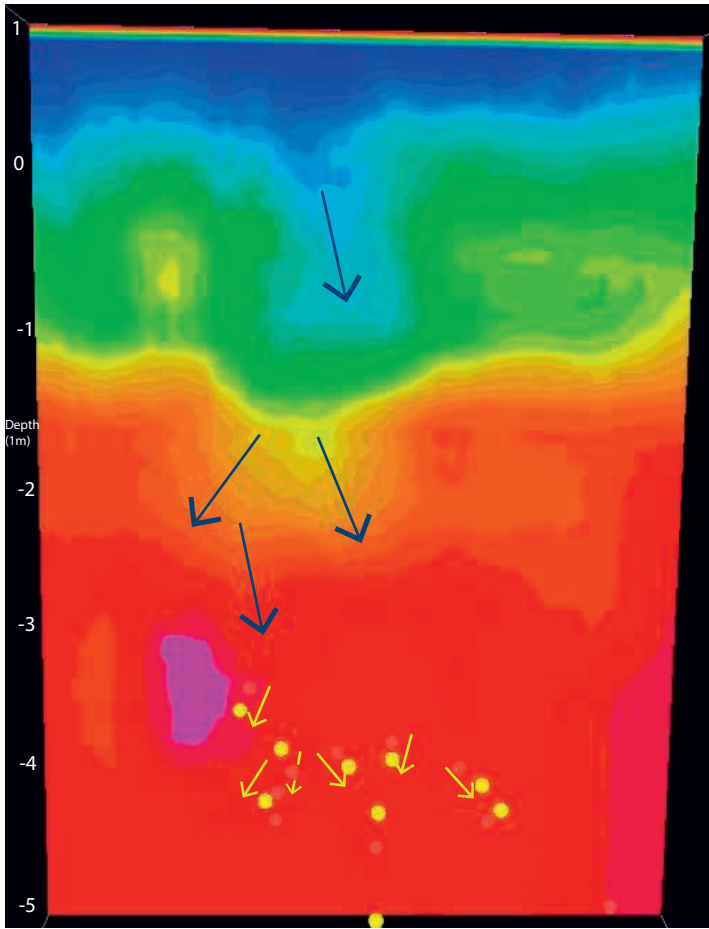


Figure 1.13 Correlation between micro-seismic cluster movement and velocity anomaly direction, pink and yellow circles are micro-seismic cluster center in 2006 and 2009 respectively (Aminzadeh, 2013).

taken from Aminzadeh *et al.* (2013), shows the use of fuzzy clustering in determining the changes in the high temperature zone of a geothermal reservoir from the movement of the cluster centers of fracture centers. More details can be found in Aminzadeh (2015).

After the introduction to neural networks and fuzzy logic in Chapters 3 and 4, Chapter 5 shows integration of conventional and unconventional methods for hybrid machine learning. In hybrid systems, illustrated in Figure 1.14, we can represent high-level and top-down hybrid systems involving any two or more “Soft Competing” or SC Techniques involving fuzzy logic, neural networks, genetic algorithms. Examples of such hybrid systems could be found, for example, at Lakhmi *et al.* (1998). SC methods each separately or their combinations can also be integrated with Conventional Statistical (CS) Techniques.

We demonstrate strength and weakness of different computing techniques based on the nature of the problem, using a “score-card” concept first introduced by Aminzadeh and de Groot (2006). We also show how hybrid systems such as neuro-fuzzy, neuro-genetic, and fuzzy-genetic algorithms overcome specific challenges by combining the respective strength of different methods.

Chapter 6 discusses natural language processing (NLP) to interpret, manipulate, understand, and drive meaning from human languages. This is particularly useful in man-machine interfacing where humans interact with computers to streamline user input process. The field of NLP has made considerable progress in recent years. Initially, NLP systems relied on word-based techniques as the basic unit of linguistic structure. However, this will gradually move towards exploiting semantics, as shown in Figure 1.15.

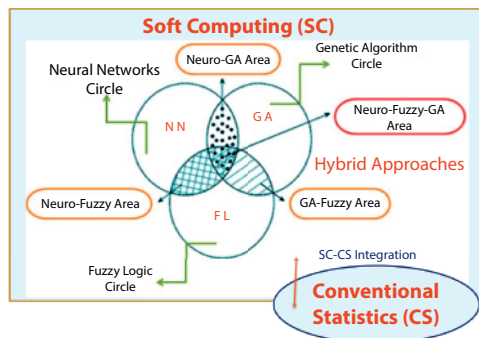


Figure 1.14 Schematic of hybrid intelligent systems depicting possible integration of different soft computing (SC) techniques NN-GA-FL as well as integration of SC with Conventional Statistical (CS) Techniques (Aminzadeh, 2021).

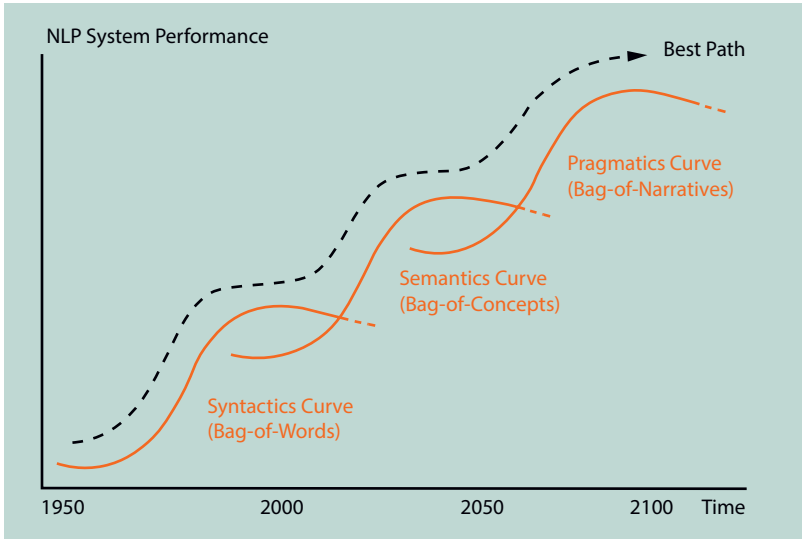


Figure 1.15 Evolution of NLP research (Cambria and White, 2014).

The progress will ultimately lead to natural language understanding systems that are intent-driven, context-aware and open-domain.

We briefly review the history and basic building blocks of NLP such as sentence segmentation, tokenization, dependency parsing, and entity recognition. We end the chapter by showcasing example use cases of NLP, including the ones in the oil and gas industry. One of the subsections of artificial intelligence (AI) is natural language processing (NLP), which analyzes human languages. In recent years, NLP applications have been used in many areas, including the oil and gas industry, in different topics such as sentence segmentation, tokenization, and lemmatization. For instance, the daily drilling reports (DDR) application is very important for monitoring field operations, staff, and processes in all oil-gas sectors. Besides DDR application, several different applications are also being used in the oil-gas industry, such as non-productive time (NPT), deep natural learning processing (DNLP), and convolutional neural network (CNN).

In Chapter 7, we discuss data science and big data analytics and three stages of a data science project: big data, algorithms and models, and infrastructure and tooling. We talk about data properties such as volume, velocity, veracity, and variety. We continue the discussion by introducing model life cycle management, automated machine learning and interpretable, explainable, and privacy-preserving machine learning. We finish the chapter by introducing some tools (e.g., R, Python, MATLAB and WEKA) and data repositories (e.g., CalGEM, Texas Railroad Commission, EDX) available for performing data science and machine learning projects. These tools, in conjunction with other workflows for coding, sharing, and communication, enable a collaborative approach for data science projects. Figure 1.16 shows the proportion of different types of tools used in various stages of a data science project based on a survey of 183 participants (Zhang *et al.*, 2020). For example, we can see that the proportion is heavily skewed having discussions initially during project scoping and creation of measurement plans. One can also appreciate the increased role of coding (e.g., in Python, R, Java) and code management (e.g., GitHub) in feature engineering, training, and evaluation of outcomes.

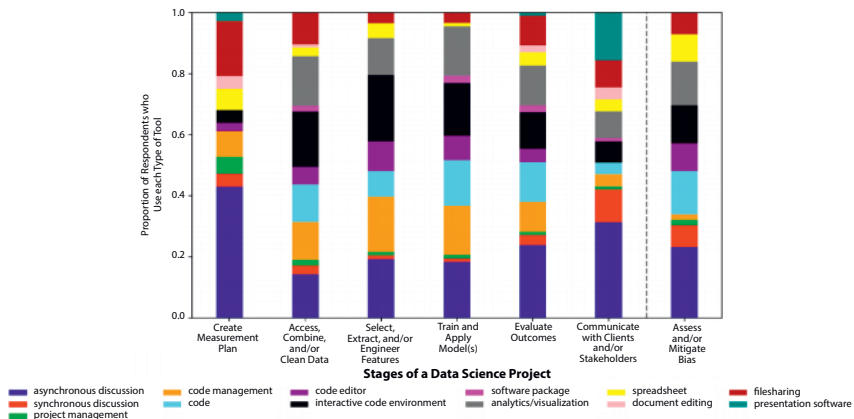


Figure 1.16 Proportion of respondents who use each type of tools for a data science project in various stages (Zhang *et al.*, 2020).

The next eight chapters of the book further highlight specific aspects of the exploration and production process, where machine learning and data analytic methods have shown promising results. These examples will hopefully pave the way for even more expanded use of intelligent data analysis and artificial intelligence in different E&P problems.

Chapter 8 considers exploration of oil and gas fields. We briefly overview petroleum systems and exploration risk factors, talk about data acquisition, processing and integration in exploration, and risk assessment of petroleum systems. We then follow by an example of using machine learning to characterize a given prospect in the Level of Knowledge (LoK) and experience using available data and geological models. Other examples introduced use artificial neural networks for auto-picking of micro-seismic data, supervised CNN and semi-supervised GAN networks for classification of 3D seismic facies, multi-layer perception ANN for generating gas chimney cubes, fuzzy kriging for geostatistical estimation, and ANN for identifying fracture zones. As an example, Figure 1.17, from Bouma (2020) shows how the use of machine learning technique yields a much clearer definition of the fault structure that in turn assists in the exploration process.

Chapter 9 focuses on application of machine learning in oil and gas drilling. Some example applications of AI in drilling system design and operations are shown in Figure 1.18 (Bello *et al.*, 2015). We illustrate how these methods can be used to improve drilling operation and detect various events in drilling. This would include use of real-time drilling data such as borehole and bit depth, mud weight, ROP, and RPM. Another example in this chapter is estimation of bottom-hole and formation temperature as well as various geosteering applications.

This chapter examines methods for improving EOR in the oil and gas industry, particularly the increasing role of machine learning (ML) to improve the efficiency of oil production in current and new reservoirs. Some algorithmic methods integrated in the sector, like Bayesian neural

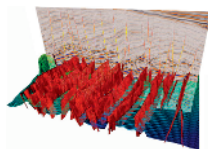


Figure 1.17 A seismic section from a 3-D seismic volume with a deep learning (machine learning)-based fault attribute blend. 3D network (red faults) showing the clarity and quality of fault information obtained from the process. Bouma (2020).

networks (BNN), are explored. This technique simulates human brain reasoning, generating a tree of potential outcomes based on prior information from well databases and adapting its parameters with changing data. Methods like these are part of a new wave of procedures to improve the recovery of oil, which tackle the current environmental challenges while providing a better efficiency for current and new wells.

Chapter 10 deals with artificial intelligence applications in reservoir characterization, with many recent applications and case histories. The process of reservoir characterization consists of the development of spatial numerical models based on the properties and the large quantities of data from a reservoir, with the goal of properly characterizing different reservoir properties such as porosity, permeability, saturation, and pressure among other properties. It leads to analyzing and exploiting the hydrocarbon inside the well. This chapter presents a series of computational models, based on artificial intelligence (AI), that improve the characterization process. Particularly, this section assesses hybrid computational intelligence (HCI), a framework that combines structures, allowing the user to combine techniques that would be incomplete when used separately. This framework can successfully work with large amounts of different types of data, and with a list of parameters that may not have complete information for a given well.



Figure 1.18 Some application of AI in drilling system design and operations – from Bello *et al.* (2015).

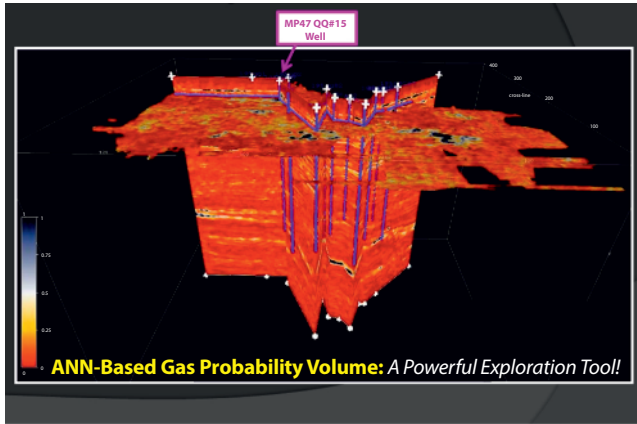


Figure 1.19 Characterization of a gas reservoir in Grand Bay Field, Gulf of Mexico, using a combination of absorption and amplitude versus offset attributes with ANN (Clifford and Aminzadeh, 2011).

Figure 1.19 shows a case where several seismic attributes were combined to characterize a Gulf of Mexico reservoir.

An important aspect of reservoir characterization is to monitor changes in the reservoir as we produce from it (primary recovery) or the various tools we use to stimulate the reservoir. This is referred to as dynamic reservoir characterization or DRC. As we note in the highlights of the respective chapter (15) on EOR quantifying changes in reservoir properties is accomplished by DRC in Figure 1.24).

In Chapter 11, we discuss the application of machine learning in production forecasting. We overview the limitations of various methods such as analytical, numerical and decline curve analysis (DCA). Conventional DCA is used to deterministically estimates future performance and remaining reserves. But several assumptions and restrictions apply, and the deterministic prediction of future declines is often far from actual production trends, or indication of true reserves. Here, we show deep learning is well suited to assimilate data from hundreds of variables to find correlations and patterns, with data-driven models that continuously learn and improve from new data. By using machine learning to automate the entire modeling and forecasting process, reservoir engineering teams can integrate day-to-day field conditions, real-time data from production facilities, and field operational constraints – like offset frac – to conduct what-if scenarios that affect production forecasting. With machine learning, the models can consider a larger and more diverse input dataset, making the decline curve analysis process more accurate and realistic.

One such example is highlighted by Shuang *et al.* (2017). As shown in Figure 1.20, an ANN with 6 input parameters including 4 logs, production data and completion data was used. The output of the ANN was the 3- and 6-month production rate. Unlike another method that creates a smoothed-out version of the production profile, the ANN-based method prediction follows the actual production values accurately.

Chapter 12 showcases the application of machine learning algorithms for production optimization, well completion, and stimulation. Examples include combining geological and completion features to optimize production, selecting optimal artificial lift strategy, ranking candidates for

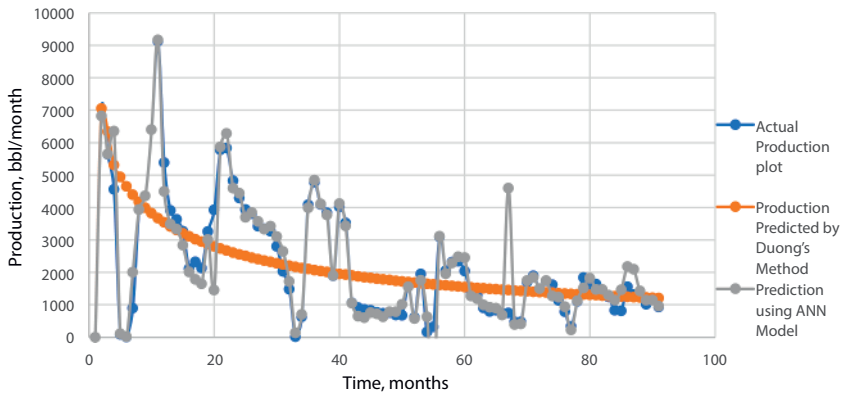


Figure 1.20 Neural network based production predictions versus Duong's method and the actual productions (Shuang *et al.*, 2017).

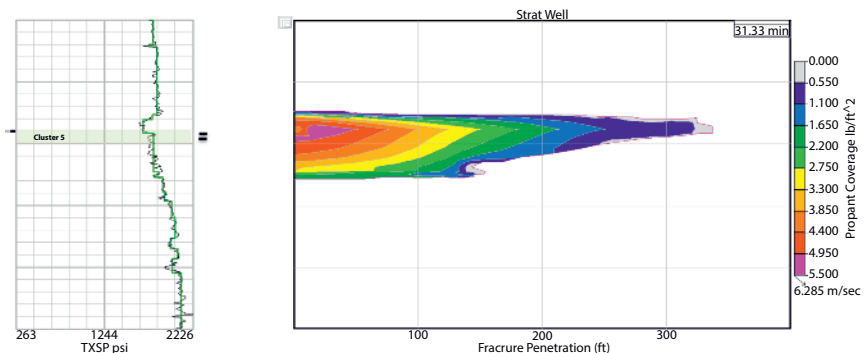


Figure 1.21 Fracture modeling. Proppant coverage at the end of pumping for perforations at 2680 ft. in Lost Hills Field in California (Popa *et al.*, 2019).

acid treatment, predicting hydraulic fracture pressure, and recommending completion methods.

With recent emphasis on shale resources, production optimization has through well stimulation focused on hydraulic fracturing as a primary tool. For example, see Aminzadeh (2019). Various machine learning approaches have been proposed to improve the process. One such example is shown by Popa *et al.* (2019), where fuzzy c- mans (FCM) clustering is used to model fractures that may be impacted by different hydraulic fracturing operational parameters. Figure 1.21 shows the results of FCM clustering-based fracture modeling for optimal perforation in a field in Bakersfield, California.

Chapter 13 discusses the reservoir engineering and simulation use cases of machine learning with examples around fluid properties estimation, proxy modeling of reservoir simulators, validation of changes in fracture

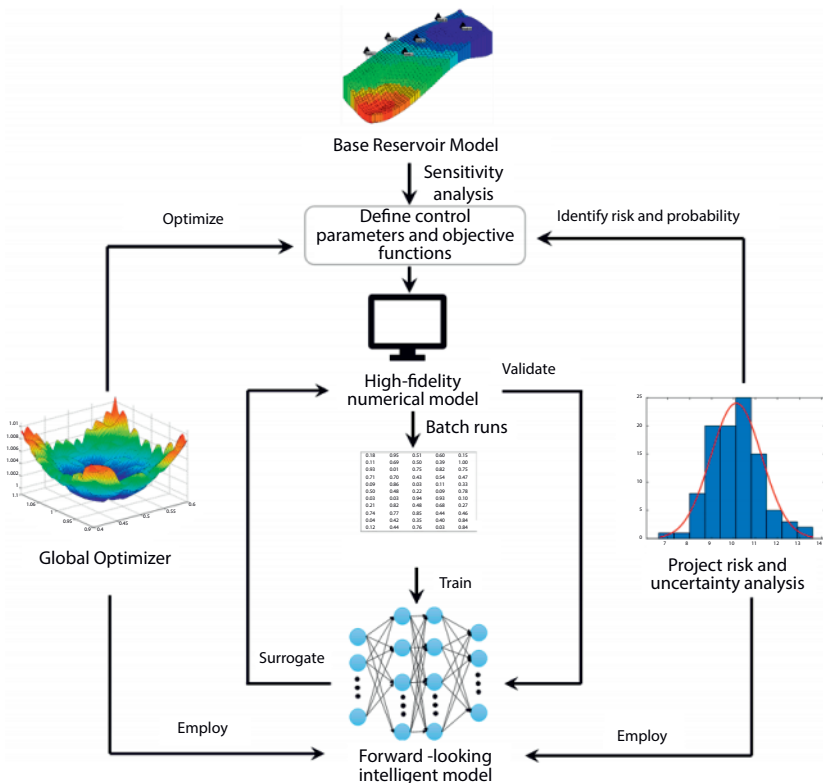


Figure 1.22 A general workflow to use ML-based models as an alternative to computationally heavy reservoir simulation models (Ertekin and Sun, 2019).

Factors of consideration in the processes of selecting artificial lift systems**Well Parameters**

- Depth
- Tortuosity
- Tubulars

Production Conditions

- Flow Rates
- Decline Rates
- GOR
- Hydraulic Fracture Operation

Surface Facilities & Supplier Factors

- Power Grid
- Pipeline
- Capacity
- Availability
- Maintenance & Spare Parts

**Fluid Properties & Reservoir Parameters**

- Density
- Viscosity
- Solid Production and Additives
- Scales
- Pressure
- Temperature
- Productivity

HSR Considerations

- Noise Level
- CO₂ Emission
- Well Integrity
- Staff Competency

Profitability Factors

- Service Life
- CAPEX
- OPEX
- Production Deferral

Figure 1.23 An example application (Syed *et al.*, 2020).

conductivity parameters, prediction of non-condensable gases in geothermal reservoirs, and well testing. An example of such workflows is illustrated in Figure 1.22 where ANN models are used to speed up the optimization and uncertainty analysis tasks (Ertekin and Sun, 2019).

Chapter 14 focuses on artificial lift and application of machine learning in this discipline to reduce equipment cost and improve efficiency, as well as keeping the system functional. As shown in Figure 1.23, several factors must be considered in selecting the most applicable lift systems (Syed *et al.*, 2020). With some of these factors being time dependent, and therefore requiring frequent predictions and updates, ML can add value in augmenting the knowledge of an engineer who is in charge of multiple wells.

Artificial lift is the process of pumping oil up the surface of the well through means that go beyond the natural reservoir pressure. This chapter reviews several artificial lift techniques powered by machine learning (ML), a subset of artificial intelligence. The role of machines has expanded over the years, and we can now rely on them to provide predictions, reports, and estimations of all sorts. Artificial lift methods can be significantly benefitted with the development of “Internet of Things” (IoT) devices, which minimize the costs, and proactively maintain operation and optimization processes.

Chapter 15 discusses how enhanced oil recovery (EOR) can benefit from intelligent data analytics. After a brief introduction to EOR techniques, this includes use of different water flooding, steam injection, and CO₂ flooding as well as injection of chemicals. We show a few sample applications of machine learning in selecting injection rates, volumes, and distribution of injection versus production well.

The success of EOR operation can be quantified by monitoring the front of injected material (water, steam, CO₂, and chemicals). One of the key technologies that can improve monitoring movement of the front water, steam, plume is use of machine learning to analyze the 4D or time- lapse seismic data before or after each stage of injection. Figure 1.24 from Dramsch *et al.* (2019) shows a comparison of the results, using the simulation approach and conventional Bayesian approach against a neural networks approach. This is a good example of the use of “Dynamic Reservoir Characterization” or DRC to quantify changes in the reservoir properties (in this case porosity, water and gas saturation) discussed in the highlights of Chapter 10 to monitor the effectiveness of EOR.

Finally, in Chapter 16 we make our conclusions in connection with the materials covered in the entire book. We also highlight our assessment of the future directions of different artificial intelligence and data analytic technologies in connection with various application areas in exploration and production covered in this book.

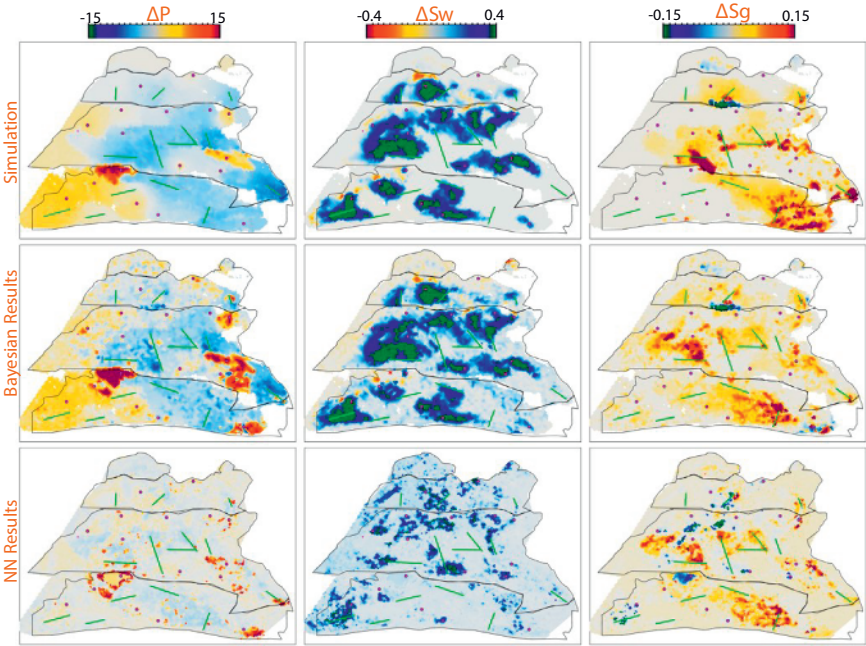


Figure 1.24 4D QI inversion results from Bayesian inversion and Neural Network inversion. Bayesian inversion closely resembles simulation output. NN result showing good coherency, consistent amplitudes, but problems in strong changes of gas saturation (from Dramsch *et al.*, 2019).