SPE-195068-MS

When Petrophysics Meets Big Data: What can Machine Do?

Chicheng Xu, Aramco Services Company: Aramco Research Center; Siddharth Misra, The University of Oklahoma; Poorna Srinivasan, Aramco Services Company: Aramco Research Center; Shouxiang Ma, Saudi Aramco

This paper was prepared for presentation at the SPE Middle East Oil and Gas Show and Conference held in Manama, Bahrain, 18-21 March 2019.

Abstract

Petrophysics is a pivotal discipline that bridges engineering and geosciences for reservoir characterization and development. New sensor technologies have enabled real-time streaming of large-volume, multi-scale, and high-dimensional petrophysical data into our databases. Petrophysical data types are extremely diverse, and include numeric curves, arrays, waveforms, images, maps, 3-D volumes, and texts. All data can be indexed with depth (continuous or discrete) or time. Petrophysical data exhibits all the "7V" characteristics of big data, i.e., volume, velocity, variety, variability, veracity, visualization, and value. This paper will give an overview of both theories and applications of machine learning methods as applicable to petrophysical big data analysis.

Recent publications indicate that petrophysical data-driven analytics (PDDA) has been emerging as an active sub-discipline of petrophysics. Field examples from the petrophysics literature will be used to illustrate the advantages of machine learning in the following technical areas: (1) Geological facies classification or petrophysical rock typing; (2) Seismic rock properties or rock physics modeling; (3) Petrophysical/geochemical/geomechanical properties prediction; (3) Fast physical modeling of logging tools; (4) Well and reservoir surveillance; (6) Automated data quality control; (7) Pseudo data generation; and (8) Logging or coring operation guidance.

The paper will also review the major challenges that need to be overcome before the potentially game-changing value of machine learning for petrophysics discipline can be realized. First, a robust theoretical foundation to support the application of machine leaning to petrophysical interpretation should be established; second, the utility of existing machine learning algorithms must be evaluated and tested in different petrophysical tasks with different data scenarios; third, procedures to control the quality of data used in machine leaning algorithms need to be implemented and the associated uncertainties need to be appropriately addressed. The paper will outlook the future opportunities of enabling advanced data analytics to solve challenging oilfield problems in the era of the 4th industrial revolution (IR4.0).

Introduction

During the past decade, we have witnessed an unprecedented booming of real-world applications of big data analytics, machine learning (ML), and artificial intelligence (AI) in many industries. Autonomous vehicles or self-driving cars, a breakthrough in harnessing big data and AI for simplifying and improving human life, can serve as a good example among all the applications (Berger, 2014; Baruch, 2016). Not surprisingly, digital transformation, big data, and ML/AI have also become buzz words in the oil & gas industry with many field applications reported from both engineering and geosciences disciplines (Crnkovic-Friis and Erlandson, 2015; Gu et al., 2016; Sidahmed et al., 2017; Wu et al., 2018; Noshi et al., 2018; Odi and Nguyen, 2018; Pham et al., 2018; Li, 2018; Maniar et al., 2018). Since 2016, almost every conference has one or more dedicated plenary or technical sessions relevant to those topics. Meanwhile, a good number of articles addressing the technical progresses of big data analytics in the upstream oil and gas businesses can be found in the industrial flagship magazines such as *Journal of Petroleum Technology* and *the Leading Edge* (Mehta, 2016; Hall, 2016; Carpenter, 2016; Jacobs, 2018; Ma, 2018; Saputelli, 2018). In addition, many scholastic journals have dedicated special issues or editions to those topics that were applied in geosciences or petroleum engineering (see the Appendix for a collected list).

If we take a retrospective review of the related upstream applications of ML/AI in the literature, petroleum engineers and geoscientists have been harnessing machines and/or computers to dig into their databases for some time to accomplish complex and challenging tasks (sometimes regarded as "mission impossible"). Figure 1 shows the trend for the number of publications on ML/AI topics in the *SPE OnePetro* digital library. The peak of the first wave of AI applications can be dated back to the 1990s when old-generation predicting tools such as neural networks, expert systems, genetic algorithms, and fuzzy logic algorithms started to pick up their momentum in petroleum industrial applications (Dashevskiy et al., 1999; MacAllister et al., 1996; Fang et al., 1992; Rivera, 1994; Cuddy and Putnam, 1998). One may notice that the number of ML papers surpassed the number of AI papers near the year of 2014, which may indicate that ML has become a more focused area of interest than the general AI. Deep learning methods revitalized the ML/AI applications in many other industries and triggered another wave of AI applications in the petroleum industry in the past few years (Crnkovic-Friis and Erlandson, 2015; Korjani et al., 2016; Li and Misra, 2017; Sidahmed et al., 2017; Jobe et al., 2018; Odi and Nguyen, 2018; Pham et al., 2018; Li, 2018; Halpert, 2018; Li and Misra, 2018).

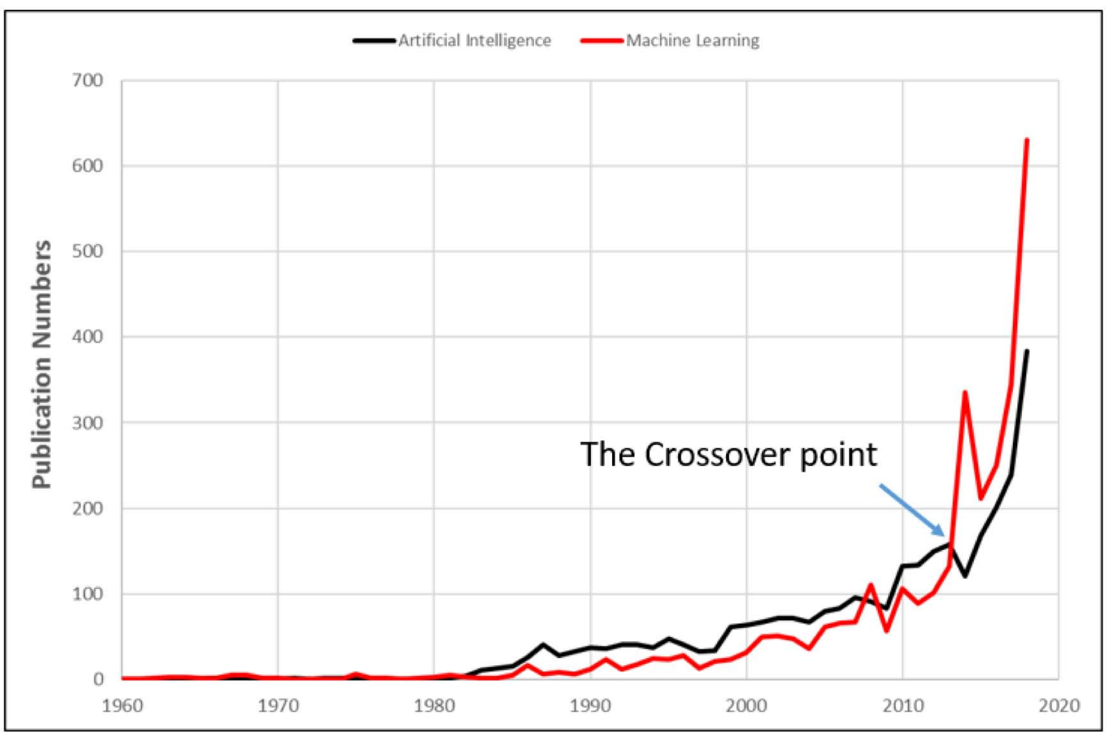

Figure 1—Trends of number of publications on "Machine Learning" and "Artificial Intelligence" in the SPE OnePetro digital library. There seems to be several waves of ML/AI applications in the petroleum industry since the 1980s. The most recent and significant wave (after 2010) is still climbing up and will likely boom in the next decade.

Significant investments from both operators and service companies are in place to drive this digital transformation revolution and IR4.0 (IKTVA 2017, Saudi Aramco Report). It is expected that the technical breakthrough in big data analytics and AI areas will generate a more significant impact in the next few decades to come. Many major and independent operators have established internal teams or departments working on ML/AI and Data Analytics. Meanwhile, numerous high-tech startups focusing on ML/AI and Data Analytics solutions or products have been established worldwide. ML, as a critical component of IR4.0, has huge potential in exploring the business value obtainable from the streaming petrophysical big data. Awareness of this potential and a high-level overview of data types, algorithms, and the existing field applications in various technical areas will significantly benefit the industry workers in selecting and applying machine learning algorithms to their own datasets to solve challenging field problems.

Emerging Trends of Petrophysical Data-Driven Analytics (PDDA)

In the petrophysics discipline, computer-aided formation evaluation (log analysis, interpretation, and integration) has become the mainstream since the 1980s with a few major industrial software platforms being developed (Doveton, 1986; Doveton, 1994). Popular data-driven methods in engineering include data mining, data visualization, and machine (deep) learning. ML/AI based petrophysical applications have also started to boom in the last decade. The confluence of distributed sensors, big data, internet of things, high-performance computing, and advanced data analytics have contributed to large-scale applications of petrophysical data driven analytics (PDDA) methods. Advanced implementations of PDDA methods are gaining acceptance in integrated formation evaluation workflows. PDDA, as a cross-discipline area, emerged from the Society of Petrophysicists and Well Log Analysts (SPWLA) in 2018. A Spring Topical Conference and a special issue of SPWLA Petrophysics Journal were dedicated to this new trend (Xu and Misra, 2018). Several areas in integrated formation evaluation showcased new advancements of data-driven methods.

In one implementation, deep learning models are used to integrate and interpret image-type datasets for geological-petrophysical analysis, such as thin sections, core/outcrop photos, image logs, seismic cross-sections, and maps (Jobe et al., 2018; Zhu et al., 2018). Another implementation involves geologically consistent integration of operational database and formation properties followed by the application of data analytics to perform more effective and efficient field operations, such as drilling and completion (Koryabkin et al., 2018; Al-Obaidi et al., 2018), fluid and rock sampling (Torlov et al., 2017), and data acquisition. Furthermore, data-driven methods have been implemented in formation evaluation for purposes of database reconstruction, noise filtering, and anomaly detection. Another popular application is to minimize tedious, repetitive human efforts required for tasks such as data cleaning, depth shifting, synthesizing missing data, and horizon picking (Zimmerman et al., 2018). In addition, data-driven analytics is particularly amenable to reservoir characterization tasks, such as facies classification and rock typing (Ozkan et al., 2011; Al-Mudhafar et al., 2017; Bize-Forest et al., 2018).

Notably, data mining and predictive modeling tasks on public databases can identify undiscovered and underutilized hydrocarbon assets. Efficacy of data-driven predictive models can be significantly improved by developing capabilities to transfer the statistical learning/mapping across: (1) reservoirs of varying rock/fluid types, (2) wells of varying orientation, geometry, and completion type, and (3) sensor types and configurations from various service providers (Bergman et al., 2017; Chen et al., 2018; Ismagilov et al., 2018). In addition, extensive multi-disciplinary and multi-organizational studies are needed to demonstrate the reliability of upstream oil and gas data analytics and data-driven predictive modeling in comparison to computationally expensive physics-based modeling methods (Xu et al., 2018; Shen et al., 2018).

To understand how machines can help with petrophysical data analysis, it is necessary to have a high-level overview of a generic framework that involves both diverse petrophysical data types and the existing pool of ML algorithms.

PDDA Generic Framework and Application Scenarios

Figure 2 shows a generic framework for utilizing ML algorithms to solve petrophysical problems. There are three key components in this framework: petrophysical data (raw data or derived attributes), ML algorithms, and output petrophysical properties. Petrophysics aims to solve formation evaluation and reservoir description problems based on multiscale subsurface data, including reservoir core and fluids, logs, well tests, seismic, and sometimes production data.

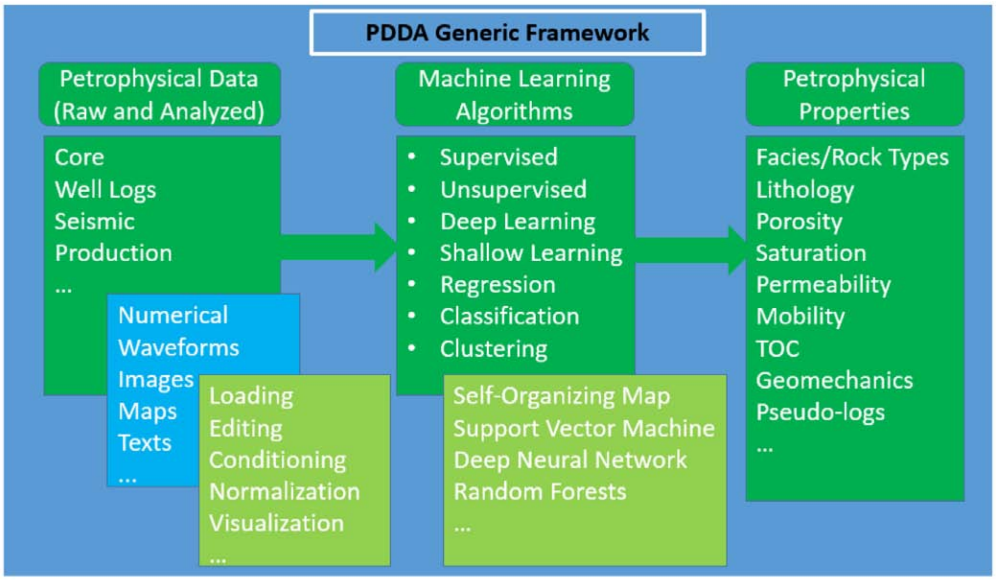

Figure 2—A generic framework for implementing PDDA to predict geological-petrophysical properties that feed into reservoir models.

There could be numerous variations of this generic model if we consider all the diverse petrophysical data types and the existing pool of algorithms. Here we list a few common petrophysical tasks that may utilize machine learning methods.

Task 1: Automated log data quality control and assurance such as bad-hole flagging, effects of mud and mud filtrate invasion, editing, and depth matching for a massive number of wells.

Task 2: Petrophysical rock typing and permeability prediction from core-log integration in a multi-well field study.

Task 3: Lithofacies, sedimentary facies, or Dunham texture classification from images logs and analyzing petrology and rock mechanics data (thin sections, SEMs, mineralogy, rock mechanical properties).

Task 4: Facies distribution in a 3D reservoir model based on core, logs, and seismic data.

Task 5: Utilizing a big operation database to analyze the root cause of a particular operational issue during logging or coring, and predicting the likelihood of having this issue given the future operation conditions.

The list will expand with more applications reported from the industry. A section highlighting the most recent PDDA applications will be presented later in this paper.

Petrophysical Data: BIG or NOT

Firstly, let's define the scope of big data analytics for petrophysical analysis. In a multi-disciplinary collaborative environment, there should not be any clear boundaries between data used by different teams such as geological, geophysical, petrophysical, or reservoir engineering. In fact, data acquisitions are often initiated by more than one discipline and should be shared across the whole asset team to maximize its value. In this paper, we generally and simply define the data commonly used by the petrophysics discipline as petrophysical data. Petrophysical data comes from various sources with different formats or types.

Petrophysical Data Sources

There are multiple sources of data that are commonly used by petrophysicists including core, log, well tests, seismic, and production. Table 1 shows some typically acquired data from various sources. As the sensor technologies are continuingly developing, the list will become longer, and should therefore not be regarded as thorough.

Table 1—Petrophysical data sources and their typical range of vertical measurement scales.

Core Measurements	Wellbore Measurements	Remotely Sourced Measurements	Reservoir Performance
nm – um – mm – cm scale	**cm – 10 m scale**	**10-100 m scale**	**10-1000 m scale**
Micro-computed tomographic images and digital rock physics Petrography (scanning electron microscope/thin sections/core photos) Mineralogy (x-ray diffraction/x-ray fluorescence) Routine core analyses (porosity/grain density/permeability/saturations) Special core analyses (Nuclear magnetic resonance/capillary pressure/electrical properties/relative permeability) Core Gamma Ray Facies description	Borehole Image (electrical and acoustic) Dielectric attributes Nuclear magnetic resonance Nuclear spectroscopy Formation testing and fluid sampling Sonic transmission and reflection attributes (various propagation modes) Cased hole nuclear attributes (various neutron interactions, spectroscopy, gravel pack evaluation) Cement evaluation Resistivity Fluid production attributes (velocity, capacitance, density, temperature) Mud log (sample description, drilling parameters, gas chromatography)	Through-casing resistivity Cross-well methods Microseismic Seismic reflection Controlled-source electromagnetic method	Drill stem test (inflow and fluid sampling) Pressure transient analysis Interference tests Decline curves Production data analyses

Petrophysical Data Types

Numerical (1-D curves): Most conventional well logs are recorded in 1-D numerical format such as gamma ray, neutron porosity, bulk density, and resistivity.

Arrays or Waveforms: Data from some advanced logging sensors are recorded in arrays or waveforms such as nuclear magnetic resonance, nuclear spectroscopy, formation testing, and dipole sonic waveforms.

Images or maps: Petrographic data such as thin sections and SEM images, image logs, cross sections (from multiple wells or seismic). 2-D maps can be generated from NMR and sonic logs.

3-D volumes: Several types of petrophysical data are spatially distributed such as x-ray computed tomography, seismic volumes, and formation testing.

Text: Lithology column and facies description from core and mud logging, annotations from drilling or interpreters.

Data can be depth indexed or time indexed. Much of the depth-indexed well logging data, considered a continuous format (i.e., wireline logs) is sampled based on time and mapped to depth, generally assuming a constant instrument velocity over short intervals. The depth index may also be discrete, e.g., core and fluid sampling. 3-D data sets are critical for understanding heterogeneity formation testing and include core x-ray tomography and remotely sourced measurements. Pressure transient testing is a unique 3-D data set, volumes of reservoir are sampled and discrete barriers may be diagnosed, but the exact spatial locations are not know.

Petrophysical Data Acquisition Scenarios

Petrophysical data acquisition for formation evaluation gets reduced with the reservoir exploration and development cycle as reservoir uncertainty level decreases (Fig. 3). Therefore, sometimes it is hard to find a common set of data in every well for a field.

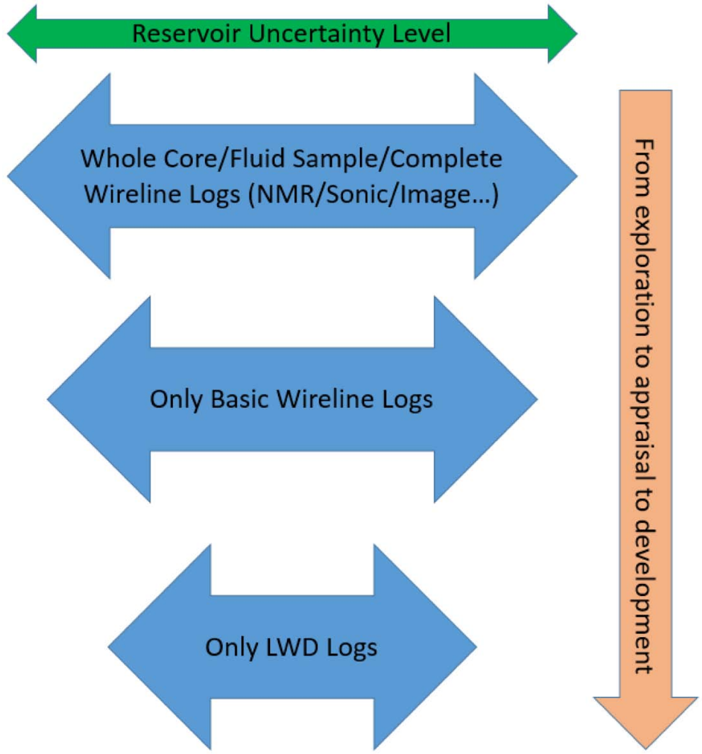

Figure 3—Petrophysical data acquisition for formation evaluation gets reduced with the reservoir exploration and development cycle as reservoir uncertainty level decreases.

Exploration & Appraisal Wells or Key Development Wells: Whole and sidewall core/fluid sampling/ complete logging suites are commonly acquired to get as much information as possible to reduce the uncertainty in the early phase of an E&P project.

Development Wells (non-key): In non-key vertical wells, typically only basic logging suites (gamma ray, neutron porosity, bulk density, and resistivity) are included in the data acquisition plan.

Development Wells (High-Angle or Horizontal): In high-angle or horizontal (HAHZ) wells, only a reduced set of formation evaluation logs are acquired by LWD, with advanced geosteering tools deployed in complex geological settings to maximize reservoir contact (Mudhhi et al., 2004). For unconventional reservoir development, it is common to only acquire gamma ray logs in many horizontal wells (Xu et al., 2016).

Is Petrophysical Data Big?

It is still debatable whether petrophysical data can be considered big or not, and it should be treated on a case-by-case basis. In general, petrophysical data meets the following "7V" characteristics that are common to big data.

Volume: The volume of core data and conventional numerical types of logs from a single well may not be considered as big. However, the data size of a field can become significantly large as the data acquisition expands to multiple well scenarios with few key wells having large datasets of advanced logs, such as sonic & NMR waveforms or borehole image logs.

Velocity: New sensor technologies, high-speed telemetries, and remote communications have enabled real-time streaming of large-volume, multi-scale, and high-dimensional petrophysical data into our databases.

Variety: Petrophysical data types are extremely diverse, and include images, waveforms, numeric values with continuous and discrete depth index (refer to Table 1 and Table 2 for a variety of petrophysical data).

Table 2—List of some commonly used petrophysical data and classification based on their attributes.

Data	Source	Type	Depth-index	Dimension
Routine Core Analysis	Core	Numerical	Discrete	Low
X-Ray Diffraction (XRD) Mineralogy	Core	Numerical	Discrete	High
SEM	Core	Image	Discrete	High
Thin Sections	Core	Image	Discrete	High
Core Photo	Core	Image	Discrete	High
Capillary Pressure	Core	Array	Discrete	High
Electrical properties	Core	Numerical	Discrete	High
Relative Permeability	Core	Array	Discrete	High
Facies Description	Core	Text	Discrete	High
Mud logging	Mud Log	Numerical and Text	Continuous	Low
Conventional Logs	Log	Numerical	Continuous	Low
Dielectric Log	Log	Numerical	Continuous	Low
Nuclear Spectroscopy	Log	Array	Continuous	High
NMR Log	Log	Waveforms/maps	Continuous	High
Image Log	Log	Image	Continuous	High
Sonic Log	Log	Waveforms/maps	Continuous	High
Formation Testing (pressure build up/draw down)	Log	Array (pressure vs. time)	Discrete	High
Pressure Transient	Production	3-D	Discrete/Time-index	High
Seismic Attributes	Seismic	3-D	Continuous	High

Variability: Sources of variability in petrophysical data includes geological controls such as heterogeneity, engineering factors such as operation environments (drilling/logging), and physical sensors (i.e., tools from different vendors).

Veracity: Petrophysical data are regarded as one of the most quantitative and reliable data sources for both geoscience and engineering disciplines to use. However, data quality can still be a challenge considering the tough logging environments and the potential physical sensor failures. It is very common to have bad and noisy data that need to undergo quality control before feeding it into any machines.

Visualization: Petrophysical data can be visualized in multiple ways to reveal the meaningful trends, patterns, and clusters. Histograms, cross-plots, logging track displays, and well correlation graphs are some typical methods to visualize petrophysical data.

Value: Petrophysical data plays an increasingly important role in modern reservoir modeling and characterization for estimating both reserve and production. It has big impact as well as value on business decision making.

Ad-hoc Challenges of Petrophysical Data

In addition to the commonly shared "7V" characteristics, petrophysical data has its own unique challenges in two aspects: scales and dimensions.

Multi-Scale Challenge. Petrophysical data comes from physical sensors that have different dimensions and resolutions, i.e., vertical, radial and azimuthal. From pore scale to core to logs to seismic to production history, the scale of measurements span many orders of magnitude. Core data can be measured on micrometer to centimeter resolution while log data are normally measured at centimeter to meter resolution.

Even in the log domain only, the scale of measurements is also a variable depending on the logging tool physics and vendors' tool designs. Figure 4 shows a wide range of logging tools and their resolutions as well as depth of investigation. As we can see, while the density logs can achieve a resolution less than 1 ft, the induction logs commonly have a lower resolution of nearly 3 ft. The azimuthal resolution varies from centimeter with focused mesaurements to full-circumference with omnidirectional measurements.

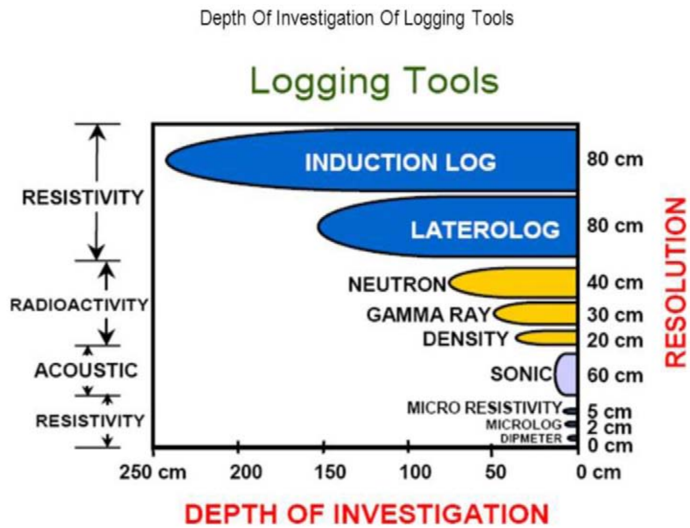

Figure 4—Comparison of vertical resolution (vertical arrows) and radial depth of investigation (horizontal arrows) of common well logs (Verdin, 2010).

Multi-Dimension Challenge. Petrophysical log data can be of different dimensions depending on the physical sensor measurements. Most conventional logs are 1D numerical values with specialty logs providing 2-D waveforms. Core tomography, production testing, and surface seismic are 3D. The dimensions of the data integrated also increases as more physical measurements become available. If we consider a single numerical log such as gamma ray as 1D, then the log data will become highly dimensional by appending more logging tracks. In fact, a petrophysical data set can be a mixture of data of different formats and dimensions. This is a unique challenge that petrophysicists need to handle.

Data: The bigger, the better?

There is a claim in the ML community that bigger data gives better models or results. However, this can be a misnomer in the petrophysics area. In petrophysical workflows, the rules governing the petrophysical estimations and interpretations are driven by the available data, so data quality and quantity holds critical importance. Having bad data is considered worse than having no data. Density, representativeness, and coverage are other parameters of the data besides data quality that are required for data-driven petrophysics (Ma, 2018). Therefore, we need to pay attention to several important aspects such as the quality and relevancy of our data before we can make this claim true.

Data Quality. "Garbage in, garbage out" is another voice often heard in geological modeling, reservoir dynamic simulation, and the data analytics community. Regardless of the source, raw petrophysical data is often "dirty" and needs rigorous and meticulous quality control by petrophysicists following protocols (Theys, 2011). Frost and Quinn (2018) discuss methods to ensure data quality and to correct environmentally-affected data, data reconstruction, and statistical correction and reconstruction processes for open hole and cased hole wireline logs, as well as logging-while-drilling data. If not adequately controlled for data quality, ML engines and AI workflows may homogenize and obscure relevant geologic and reservoir features. However, data quality control work is cumbersome, especially for large field projects that contain hundreds or even thousands of wells. ML has demonstrated some advantages over humans in delivering consistent data QC work such as depth matching. Zimmermann et al. (2018) tested ML based methods as a new direction to tackle depth matching issues with a fully automated solution.

Data Relevancy. Irrelevant data can also add confusion to the predictive model. Many tests have shown that use of irrelevant data will only deteriorate the accuracy and efficiency of the predictive models. The relevancy should be based on some physical or petrophysical principles. For example, gamma ray logs are relevant to permeability in a shaly sand reservoir because they are an indicator of total clay volume, which is a controlling factor of many rock properties such as effective porosity, pore throat size, and permeability. But for a carbonate reservoir with an almost flat gamma ray response, inclusion of gamma ray logs in the permeability prediction model may not help at all. Feature selection in many old-generation algorithms needs to be performed carefully by domain experts. Many new-generation algorithms have the capability to determine the relevancy of various features and put less weighting factor on features of low relevancy (Akande et al., 2015; Anifowose et al., 2016).

Petrophysical Model: Physics Based or Data-Driven

In general, petrophysical models are grounded in sound physics, referred as mechanistic models, or developed as empirical or phenomenological models. The petrophysical models are used to quantitatively derive various petrophysical properties by processing the physical measurements obtained from core or well logs. For example, Archie's model and its variations are empirical models that are commonly used to calculate water saturation by jointly processing porosity estimations and resistivity logs (Archie, 1942). However, Archie's model involves several assumptions, such as clay-free rocks, no significant invasion, and absence of complex pore systems. Archie's model is inadequate for clay-rich, highly tortuous, and thinly laminated reservoirs (Worthington, 2000). Similarly, other mechanistic, empirical, and phenomenological models involve various assumptions for the models to be valid. When the assumptions of these models cannot be met due to the complexity, heterogeneity, and multiscale nature of the physical processes, then these models become unsuitable for petrophysical interpretations and estimations.

In some cases, the models required for petrophysical calculations need to be extremely nonlinear and non-explicit. Mechanistic, empirical, and phenomenological models cannot account for such nonlinearity. ML methods can be used as an alternative approach to develop data-driven models for better characterization of petrophysical processes and systems. Data-driven models can provide a computationally cheaper surrogate model to substitute the costly physics-based model or provide an approximate statistical model from observations when there is no deterministic physics-based model (Aifa, 2014).

Machine Learning Algorithms Overview

ML extracts patterns and structures from historical data (training data for supervised or unsupervised learning) or by interacting with an environment (reinforcement learning) to characterize, identify, or predict the behavior of a system and/or process. Such systems can be massively multivariate involving thousands of features/attributes. ML algorithms use the data to learn the underlying behavior of the system without prior knowledge of the nature of relationships between data points. ML is well suited to address those problems where theoretical physics-based knowledge is still incomplete but for which we do have a significant number of measurements and observations.

ML techniques are broadly categorized into supervised learning, unsupervised learning, and reinforcement learning. Unlike unsupervised learning, supervised learning relates features/attributes to labels. Two popular supervised learning tasks are regression and classification. Regression generates continuous outputs/responses (e.g., porosity, saturation, permeability), whereas classification generates discrete output/classes/labels (e.g., lithology, facies, rock types). Unsupervised learning processes features/attributes to identify commonalities and differences to learn relationships and patterns in the training data. Clustering, anomaly detection, dimensionality reduction, feature ranking, and data visualization are few tasks based on unsupervised learning. In supervised and unsupervised learning, a comprehensive 'training dataset' of examples is constructed covering as much of the system parameter space as possible. For supervised learning, a random subset of the data is put aside for a completely independent testing. A glimpse at the pool of ML algorithms will help understand their applicability to our specific petrophysical data set.

Supervised Learning vs. Unsupervised Learning

In supervised learning, the data-driven model is built by processing a known labeled dataset that includes desired inputs (features) and outputs (labels/responses). Supervised learning finds the model that generates the outputs based on the inputs. A physics-driven model is a theory-based mathematical mapping that relates the input and output, whereas supervised learning identifies patterns in the available dataset, learns from observations, and makes the necessary predictions based on statistical mapping of inputs and outputs. During the process of building the supervised learning model, the predictions are compared to the output and the model is improved based on a loss function. This process continues until the data-driven model achieves a high level of accuracy and performance so the loss function can be minimized.

Unsupervised learning processes datasets to identify patterns, relations, and commonalities without using examples, labels, and human instruction. This ML technique organizes the data in a certain way that describes the structure, variance, density, distribution, etc. of the dataset. This might mean grouping the data into clusters or arranging it in a way that looks more organized or easy to visualize. Few examples of unsupervised learning: dimensionality reduction, data compression, manifold learning, and feature extraction. An inherent issue with unsupervised learning is to assess the efficacy and reliability of the model.

Shallow Learning vs. Deep learning

Deep learning is a subset of ML that uses multiple layers of non-linear processing units to generate multiple levels of representation of the data corresponding to various levels of abstraction. Deep learning is also referred as representational learning. Such an architecture can learn representations and features directly from the input with little to no prior knowledge, hand-coded rules, or engineered features. Recently, deep learning has delivered state-of-the-art accuracy in tasks such as object detection, speech recognition, and language translation. Their highly flexible architectures can learn directly from raw data, such that their predictive accuracy increases with a greater exposure to new data. Unlike deep learning, shallow learning is based on the user feeding prior knowledge and engineering features as additional inputs to aid the learning process. The emphasis in shallow learning is often (not always) on feature engineering and selection, whereas the emphasis in deep learning is on defining the most useful computational graph topology and finding optimal parameters/hyperparameters. A few examples of shallow learners are decision trees, support vector machines, and Naive Bayes. Multilayer feed forward neural networks, autoencoders, recurrent neural networks are examples of deep learning. Unlike shallow learning that plateau at a certain level of performance with the supply of additional examples and training data, deep learning models scale with data.

Regression vs. Classification

Supervised learning analyzes training data and produces inferred function that can map new examples. Depending on the type of output, supervised algorithms are broadly categorized into regression and classification. Regression generates continuous outputs/responses (e.g., porosity, saturation, permeability). It determines the contribution and correlation of features that produce a specific output. On the other hand, a classification generates discrete output/classes/labels (e.g., lithology, facies, and rock types). Unlike regression, classification is done when output is labelled into categories and the classifier assigns a label/class based on certain features.

Table 3 lists some commonly used ML algorithms in the Petrophysics literature. This is by no means a complete list and many algorithms can be adapted to fit for different purposes.

Table 3—Summary of ML algorithms commonly used for solving geological and petrophysical problems (Wikipedia: Outline of machine learning, 2018).

Algorithms	References
Support Vector Machine (SVM)	Nazari et al., 2011; Zhao et al., 2015B; Venna et al., 2018
Random Forest (RF)	Bestagini et al., 2017; Akinnikawe et al., 2018
Self-Organizing Map (SOM)	Zhao et al., 2015A
K-means Clustering	Gupta et al., 2018; Han et al., 2019
Hierarchical Clustering	Gupta et al., 2018; Han et al., 2019
Expectation-Maximization (EM)	Schlanser et al., 2014
Bayesian Network (BN)	Agar et al., 2018; Shen et al., 2018
Convolutional Neural Network (CNN)	Jobe et al., 2018; Zhu et al., 2018; Alqahtani et al., 2018; Li and Misra, 2018
Recurrent Neural Network (RNN)	Alfarraj and AlRegib, 2018
Autoencoder Neural Network	Li and Misra, 2017; Li and Misra, 2018

Highlights of the Most Recent ML Applications in Petrophysics

In Petrophysics literature, ML has demonstrated some advantages over the conventional methods in the following technical areas: (1) Geological facies classification or petrophysical rock typing; (2) Seismic rock properties or rock physics modeling; (3) Petrophysical/geochemical/geomechanical properties prediction; (4) Fast physical modeling of logging tools; (5) Well and reservoir surveillance; (6) Automated data quality control; (7) Pseudo data generation; (8) Logging or coring operation guidance.

A special issue of the Petrophysics Journal was released in Dec. 2018, highlighting PDDA and featuring eleven papers covering theories and applications of emerging data analytics and ML methods as applicable to petrophysical data analysis. Table 4 features those recent PDDA applications by summarizing the input data, algorithms, and tasks of each reference in all different technical areas.

Table 4—Summary of the input data, algorithms, and tasks of each reference in the special issue of SPWLA Petrophysics Journal (Dec. 2018).

Reference	Data Input	Algorithms	Achieved Tasks
Jobe et al., 2018	Thin section images	CNN – deep learning	Dunham texture classification
Zhu et al., 2018	Well logs	Wavelet decomposition and CNN – deep learning	Lithology classification
Gupta et al., 2018	Well logs K-means, hierarchical clustering, and SVM	Hydraulic fracturing location	
He et al., 2018	Well logs	OLS, PLS, LASSO, MARS, ANN.	Pseudo log generation
Tariq et al., 2018	Well logs	Functional Network	Static Poisson's ratio prediction
Xu et al., 2018	LWD logs	Neural network, Gradient-Boost Regression Tree, Gaussian Process Regression	LWD resistivity modeling
Shen et al., 2018	LWD logs	Transdimensional Bayesian method	Ultradeep azimuthal resistivity interpretation and uncertainty quantification
Venna et al., 2018	Acoustic signal	PCA & SVM	Flowing phase classification
Chen et al., 2018	Fluid optical data	Forward & inverse neural network	Missing fluid data reconstruction
Frost and Quinn, 2018	Logs in general	Automation workflows	Data quality control
Zimmermann et al., 2018	Logs in general	Fully connected neural network	Automatic depth matching

Geological facies classification or petrophysical rock typing

ML approaches have proven to be efficient and effective in automating facies classification with core images, well logs, and seismic attributes (Wang et al., 2014). Izadi et al. developed an intelligent system for mineral identification in thin section images based on both incremental color clustering (2015, overall accuracy 92.15%) and artificial neural network (2017, overall accuracy 93.81%). Cui et al. (2017) applied the principal component analysis method on well logs to classify four diagenetic facies pre-defined by samples from a tight sandstone reservoir in the Ordos Basin, Central China. Schlanser et al. (2016) tested a statistical clustering algorithm with petrophysical and elastic well logs for lithofacies classification in the Marcellus Shale. Zhao et al. (2015A) used different ML algorithms, such as self-organizing map and artificial neural networks to assist quantitative seismic facies interpretation that differentiated the architectural elements of a turbidite system in the Canterbury Basin, offshore New Zealand. Anifowose et al. (2016 & 2017) demonstrated integrated reservoir studies based on a hybrid computational intelligence model (2016) and ensemble ML (2017). Jobe et al. (2018) introduced two ML approaches to interpret thin section image data for geological-petrophysical analysis. The first approach uses pore-geometric analysis and four ML algorithms to train and classify samples by reservoir zone. The second approach uses a convolutional neural network to classify and predict carbonate Dunham textures of mudstone, wackestone, packstone and grainstone. (Fig. 5). Zhu et al. (2018) proposed a method based on wavelet decomposition to construct multilayer image-style input for each logging point, converting the problem of logging lithological interpretation into a supervised image recognition task. Real-world application in the Daqing oilfield demonstrates that the proposed method is able to perform more accurate lithological classification compared to the ordinary neural network, which provides a reliable alternative way for further logging lithological interpretation practices (Fig. 6).

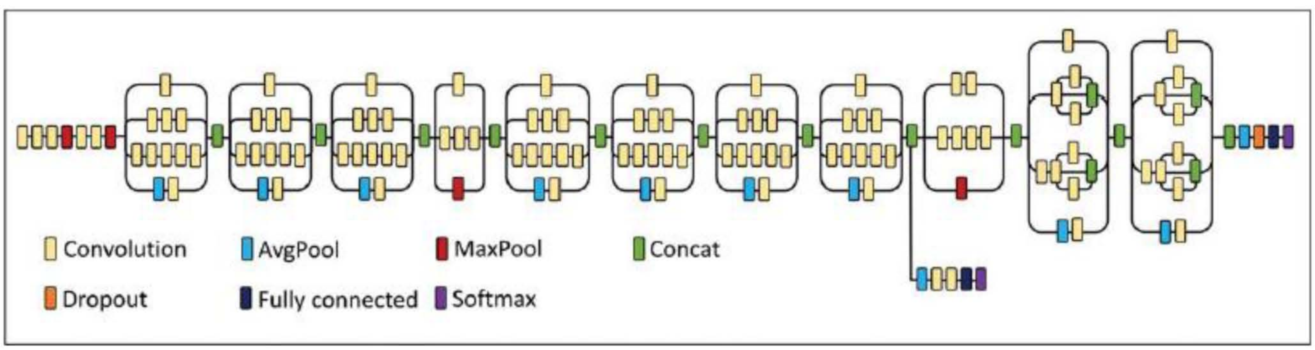

Figure 5—Computational graph of TensorFlow Inception-v3 convolutional neural network (image-based deep learning) for carbonate Dunham textures classification (Jobe et al. 2018).

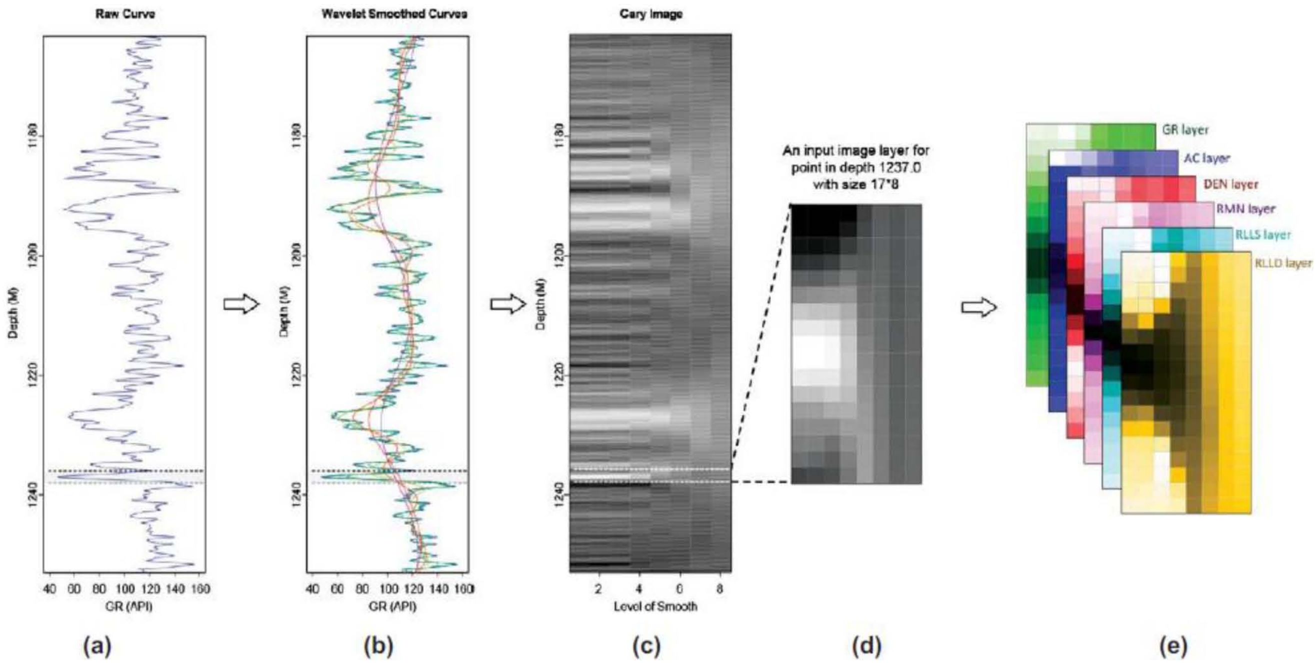

Figure 6—Construction of input image layers for the specified point by Zhu et al. 2018.

Seismic rock properties or rock physics modeling

Seismic rock properties and rock physics is a unique area integrating geophysics and petrophysics. Data-driven ML approaches have gained momentum in geological & geophysical interpretations (Zhao et al., 2015C; Hall, 2016; Zhang et al., 2018; Bandura et al., 2018; Alfarraj et al., 2018; Ma et al., 2018). Digging into the latest SEG literature, we can find many applications of using ML to do velocity autopicking (Smith, 2017), amplitude-variation-with-angle classification (Bougher and Hermann, 2016), seismic facies classification (Bestagini et al., 2017; Chopra and Marfurt, 2018), elastic facies classification (Schlanser et al. 2014), attenuation logs prediction and modeling (Parra et al. 2014), shear wave velocity prediction (Du et al., 2018), and geomechanical properties estimation (Gu et al., 2016). On the small scale end of rock physics, ML has been used in combination with digital rock physics for nanoscale pore-space characterization (Kazak, 2018), porous media properties prediction (Alqahtani et al., 2018), pore system classification (Serag et al., 2018), and rock mechanical properties characterization (Saad et al., 2018).

Petrophysical/Geochemical/Geomechanical properties prediction

As previously mentioned, most petrophysical models will become nonlinear or non-explicit when multiple factors contribute and only indirect physical measurements can be taken. Data-driven models can be used as an alternative approach to predict many petrophysical properties such as lithology (Shin et al., 2018), porosity (An et al., 2018), saturation or fluid volumes (Khan et al., 2018; Jain et al., 2015; Venkataramanan et al., 2018), permeability (Nazari et al., 2011; Anifowose et al., 2013; Shabab et al., 2016), total organic content (Zhao et al., 2015B; Mahmoud et al., 2017), kerogen properties (Craddock et al., 2018), net pay (Masoudi et al., 2012), and Poisson's ratio (Tariq et al., 2018). In particular, it is noted that machine learning has huge potential in processing and interpreting advanced well logs such as NMR echoes, image logs, sonic waveforms, nuclear spectroscopy which often demand more advanced algorithms to decipher the hidden petrophysics properties (Jain et al., 2015; Venkataramanan et al., 2018; Han et al., 2019).

Fast physical modeling of logging tools

Physical modeling and simulation of logging tools requires intensive computational time. This is particularly true for nuclear, resistivity, and sonic tools. Several applications of using ML to achieve fast physical modeling of logging tools were reported. Xu et al. (2018) introduced a variety of ML techniques into the electromagnetic modeling problem for LWD (Fig. 7). With their ML-based methods, modeling of LWD resistivity tool responses can be accelerated, with higher accuracy, to support real time drilling scenarios. The ML approach separates the training phase from the real-time prediction phase. This feature makes it a perfect tool to facilitate real-time workflow. Compared with the existing look-up table approach, the ML approach can be more accurate. Furthermore, it also has much smaller memory requirement, so it has the potential to support a larger training dataset and higher dimensional input parameter space.

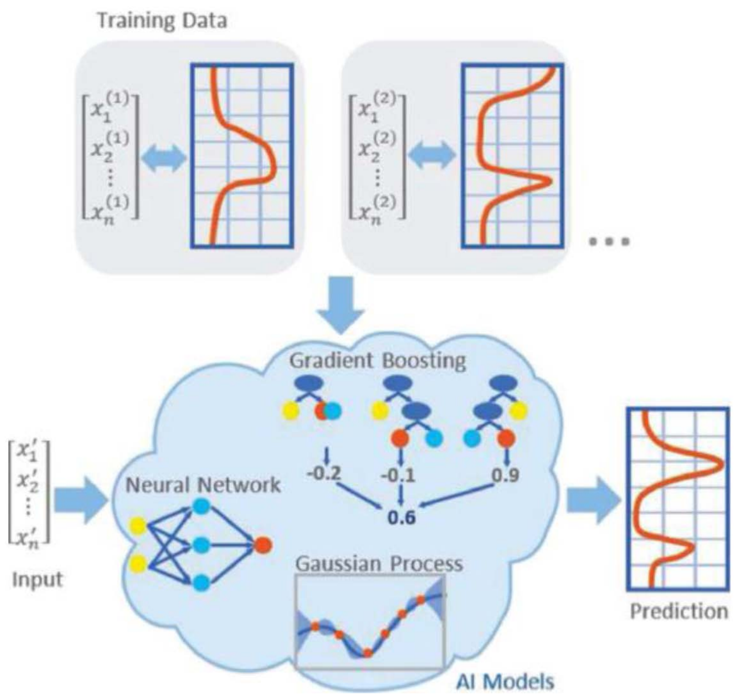

Figure 7—Use of ML for logging tool physical modeling (Xu et al. 2018).

Well and reservoir surveillance

Real-time reservoir monitoring or surveillance with distributed acoustic sensing (DAS) and distributed temperature sensing (DTS) by using fiber optics technology has been widely deployed in some major oilfields (Hveding and Bukhamsin, 2018). These distributed fiber-optic measurements enables nearly continuous monitoring of dynamic downhole environments changing with both space and time. Although continuous monitoring greatly facilitates asset management applications, it comes with its unique challenges associated with data transmission, management, and security. Cloud-based services for fiber-optic data management have been successfully deployed in the petroleum industry as an effective way to collect, transfer, store and visualize distributed measurement data from various downhole environments. Yang et al. (2018) presented their successful integration of an analytics library into the cloud-based fiber-optic data management system. This integration enables real-time, and in some cases near real-time, asset decision making. They also introduced a few application examples of the analytics integration using real-time data streamed directly from the field. Their applications included gas lift alerts, hydrate monitoring, and distributed strain alerting.

Automated data quality control

Data quality control includes data cleaning, data editing, data normalization, and depth shifting or matching. It is a critical step for reservoir description (Voleti et al., 2018) but can be very tedious and time-consuming if hundreds or thousands of wells (each with multiple logs) need to be analyzed. ML can be used to create automated data quality control workflow. Andrianova et al. (2018) used ML methods to automatically quality control large volumes of field data including well logs, fluid properties, and well production and interventions. Akkurt et al. (2018) utilized ML system to automate the workflow of identifying outliers in bulk-density and compressional slowness logs, and to reconstruct them using data-driven predictive models. Zimmermann et al. (2018) emphasize that depth matching is critical for integrated well log analysis but remains a long-standing challenge in the industry. Traditional methods do not guarantee convincing depth matches without user intervention. ML opens the door to leverage fully automated computer based solutions to tackles this tedious and challenging issue.

Pseudo data generation

Well logs acquired in bad-hole conditions often exhibit poor quality and need to be either removed or edited. A few instances require specific logs to be generated/acquired in a well or an interval to enable running a complete petrophysical or reservoir characterization workflow. For example, sonic logs are often missing in shallow zones or even not acquired in many development wells due to operational or financial constraints. But sonic logs are critical to time-to-depth conversion as well as seismic well tie workflow. He et al. (2018) performed a comparative study of various shallow learning methods to synthesize compressional and shear travel time logs in the absence of a sonic logging tool. The shallow learning models processed 13 conventional and easy-to-acquire logs, namely lithology, gamma ray, caliper, density porosity, neutron porosity, photoelectric factor, bulk density, and resistivity at six depths of investigation. The best performing methods for the sonic logs syntheses were Artificial Neural Networks (ANN) and Multivariate Adaptive Regression Splines (MARS) that synthesized the logs at a coefficient of determination (R^2) of 0.85. ANN and MARS performed better than ordinary least squares (OLS), partial least squares (PLS), ridge regression, LASSO, and Elastic Net. Li et al. (2018) extended the workflow proposed by He et al. (2018) that applied six clustering methods to identify formations that are not suitable for sonic log synthesis using the regression models.

Chen et al. (2018) presented a novel mutual-complementing method that enables the reconstruction of missing fluid optical measurement data in external public databases (Fig. 8). In this method, forward and inverse neural networks are used to build multivariate correlation models between optical measurements and fluid composition/property data in the in-house database. The neural networks can then be applied to an external petroleum fluid database by alternatively performing forward and inverse computation to evaluate the data consistency and to reconstruct the missing fluid optical measurement, therefore augmenting the in-house database with enhanced sample diversity and measurable improvements for downhole fluid analysis.

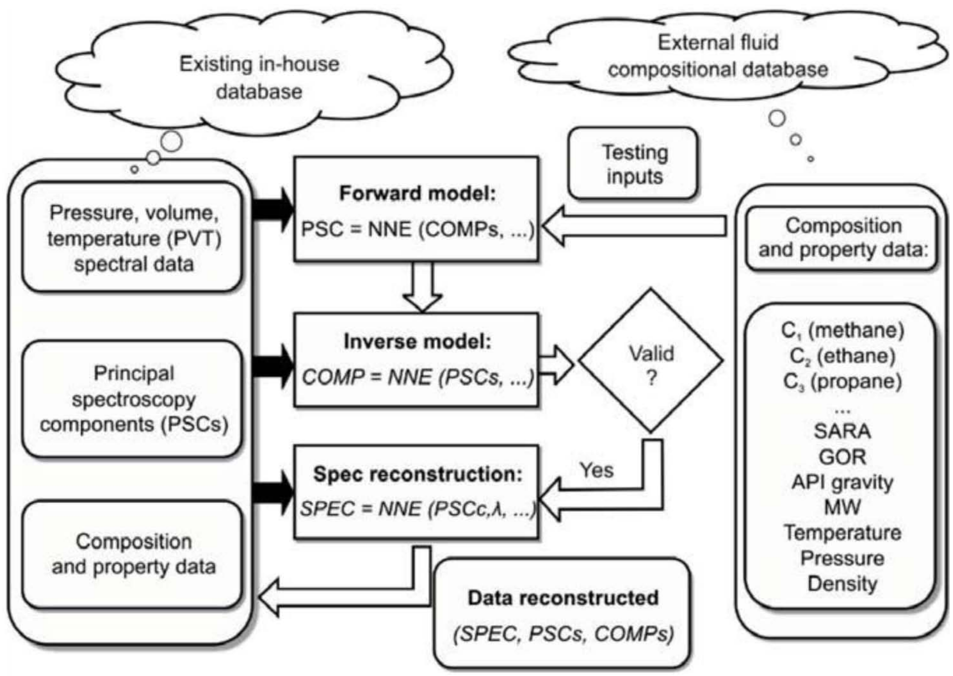

Figure 8—Workflow applied to reconstructing optical principal spectroscopy components (PSC) and spectral data from external sources using available fluid compositional inputs (Chen et al. 2018).

Operation guidance: logging, coring, fluid sampling, and fracturing

ML and data analytics also found their extensive applications in guiding wellsite operations such as logging, coring, fluid sampling, or even hydraulic fracturing. Large operators and service companies have accumulated multi-year experience in certain jobs from worldwide locations under different operation environments and reservoir conditions. Massive relevant data from their job history can be pooled together to form a database which can be used to build predictive models for diagnosing operation issues or guiding future operations. Xu et al. (2016) presented case histories and lessons learned from experience with 304 wireline formation pressure runs implemented with downhole automation controls. They used smart self-learning operating methods to achieve reliable performance by improving the accuracy of the pressure and mobility data as well as reducing the test duration. Torlov et al. (2017) studied many factors that impact rotary sidewall coring service quality from over 3000 attempted samples by 8 different types of rotary coring tools of 3 logging companies. They used statistical methods such as principal component analysis to make data-driven assessment of rotary sidewall coring performance. Gupta et al. (2018) used several clustering techniques, including *K*-means and hierarchical clustering, to identify the most suitable cluster locations for hydraulic fracturing. The results can be upscaled to triple combo logs by supervised clustering techniques such as decision trees, gradient boosting and support vector machines. The upscaled results can be used to select optimal completion zones. A multivariable statistical analysis was done with a large formation testing database comprising 344 fluid sampling stations from 59 wells located across several Gulf of Mexico fields, to determine the probability of fishing and the top five factors that cause fishing (Wu et al., 2015; Pineda et al., 2018).

Key Challenges for PDDA

Data-driven methods require exposure to large datasets for the purposes of extracting patterns and trends in the data to build predictive models and for testing the robustness and generalization capabilities of these models. A big challenge in upstream oil and gas is the inconsistent data quality and limited data quantity available for ML applications. Small oil and gas companies and academia often lack large and complete datasets. An industry-wide effort is necessary to generate and compile extensive high-quality datasets, possibly masked, that can foster development of data-driven methods. In 2018, Equinor and the Volve license partners decided to disclose all subsurface and operating data from Volve. This has been the most comprehensive oilfield data release so far, which is a milestone in oilfield data sharing.

Without rigorous data preprocessing and quality control, the quality and relevance of predictive models deteriorate with the increase of available data. Technical presentation, dissemination, and demonstration of ML implementations should emphasize data cleaning, standardization, normalization, and calibration. Data conditioning and feature selection improve the robustness of the data-driven models. A standardized quality control procedure for data-driven methods will significantly mitigate the aforementioned problems and improve computational efficiency as well as predictive capability (Proett et al., 2015). Lack of well-trained data technicians, investment in data management, interdisciplinary integration, industrial standards and best practices lead to poorly performing data-driven solutions. Despite ideal data quality and reliable data preprocessing, applications of data-driven methods are challenging in the presence of heterogeneous and multi-scale nature of rocks, limited volume of sensor investigations, real-time operational requirements, and high-dimensional data with large uncertainty.

In addition to the above-mentioned challenges, there remains a few other pertinent issues. Users of data-driven methods should evaluate the results by adopting the following techniques: blind source testing, offset well comparison, ensemble approach, and cross validation. From a petrophysical viewpoint, these methods need to be geologically consistent, which mandates extensive training with field analogs and geological rules followed by populating the reservoir model with all the constraints extracted from the training data. Efficacy of data-driven predictive models can be significantly improved by developing capabilities to transfer the statistical learning/mapping across: (1) reservoirs of varying rock/fluid types, (2) wells of varying orientation, geometry, and completion type, and (3) sensor types and configurations from various service providers.

Software packages for log interpretation and formation evaluation may consider providing modules that can integrate with open-source data analytics software and codes. The next generation of petrophysicists and log analysts should be trained through standardized courses, internships, software training, and good exposure to business cases. In the near future, logging tools and software development and design may be directly geared towards data-driven method implementations, which require a completely new way of thinking that builds on our experience with physics-centered models and tools. ML methods are still not intelligent enough to aggregate human experience and expertise. For the data-driven methods to become more intelligent, there is a need for collaboration among the petrophysics, tool physics, and data science domain experts. Such cross-disciplinary collaborations are necessary, but still remains a challenge.

Future Outlook and Summary

The new era of digital transformation has opened numerous opportunities to solve the challenging problems associated with the big data generated from the oil and gas industry. A glimpse of the recent applications in the petrophysics domain indicates that a fully blossomed digital age is yet to come. Several critical conditions have been met to trigger this revolution. First, there has been increasing awareness of the potential use of advanced data analytics in the petroleum industry; second, advanced algorithms (from other industries) and high-performance computing machines (GPUs) are maturing and ready to use; third, tough business environment necessitates better efficiency and lower cost which provide a strong business drive for new technology implementation; lastly, the oil and gas industry is training its technical staff to apply the right technology at the right time to solve the right problems using the right data.

The petrophysics discipline has abundant and quantitative data to leverage the emerging ML/AI technology to accomplish challenging tasks in almost every technical area. PDDA has become one of the most active sub-disciplines of petrophysics that triggered widespread interest from both operators and service companies. We need to keep in mind that there are still some unique and outstanding challenges that need to be resolved before another wave of industrial applications of ML/AI technology comes to make industry wide business impact. To go beyond mysterious black-box applications, fundamental research work must be conducted thoroughly to understand both data and algorithms so that ML/AI technology will have a solid foundation to expedite IR 4.0 in the petroleum industry.

Acknowledgement

The manuscript is restructured to highlight the new concept of Petrophysical Data-Driven Analytics (PDDA) based on a few previous non-peer-reviewed publications, including the meeting summary of 2018 SPWLA Spring Topical Conference (April 16-17, Houston, Texas) and the call-for-paper and introduction for the December 2018 special issue of the SPWLA Petrophysics Journal. I would like to thank all workers who participated in these two events for contributing their ideas and sharing their field applications to be referenced in this paper. A special note of thanks goes to Susan Agar, Jeffry Hamman, Dawn Jobe, Shuo Zhang, and Cristina Young for proofreading and improving the manuscript.

References

Agar, S., Goteti, R., Jobe, D., Zhang, S., and Zaretskiy, Y. (2018). Geologic Risking with Bayesian Methods: Fracture Case Study in an Unconventional Setting. GEO2018 Conference held in Manama, Bahrain.

Aifa, T. (2014). Neural network applications to reservoirs: physics-based models and data models. *Journal of Petroleum Science and Engineering*, **123**, 1-6. doi: 10.1016/j.petrol.2014.10.015.

Akande, K. O., Olatunji, S. O., Owolabi, T. O., and AbdulRaheem, A. (2015). Comparative analysis of feature selection-based machine learning techniques in reservoir characterization. Society of Petroleum Engineers. doi: 10.2118/178006-MS.

Akinnikawe, O., Lyne, S., and Roberts, J. (2018). Synthetic well log generation using machine learning techniques. *Unconventional Resources Technology Conference*. doi: 10.15530/URTEC-2018-2877021.

Akkurt, R., Conroy, T. T., Psaila, D., Paxton, A., Low, J., and Spaans, P. (2018). Accelerating and enhancing petrophysical analysis with Machine learning: A case study of an automated system for well log outlier detection and reconstruction. Society of Petrophysicists and Well-Log Analysts.

Alfarraj, M. and AlRegib, G. (2018). Petrophysical property estimation from seismic data using recurrent neural networks. Society of Exploration Geophysicists.

Al-Mudhafar, W. J. (2017). Integrating well log interpretations for lithofacies classification and permeability modeling through advanced machine learning algorithms. *Journal of Petroleum Exploration and Production Technology*. **7**(4), 1023-1033.

Al-Mudhhi, M. A., Ma, S. M., Al-Hajari, A. A., Lewis, K., Berberian, G., Butt, P. J., & Richter, P. (2005, January 1). Geo-Steering with Advanced LWD Technologies - Placement of Maximum Reservoir Contact Wells in a Thinly Layered Carbonate Reservoir. International Petroleum Technology Conference. doi: 10.2523/IPTC-10077-MS

Al-Obaidi, M., Heidari, Z., Casey, B., Williams, R., and Spath, J. (2018). Automatic well-log-based fabric-oriented rock classification for optimizing landing spots and completion intervals in the midland basin. Society of Petrophysicists and Well-Log Analysts.

Alqahtani, N., Armstrong, R. T., and Mostaghimi, P. (2018). Deep learning convolutional neural networks to predict porous media properties. Society of Petroleum Engineers. doi: 10.2118/191906-MS.

An, P., Yang, X., and Zhang, M. (2018). Porosity prediction and application with multiwell-logging curves based on deep neural network. *Society of Exploration Geophysicists*.

Andrianova, A., Simonov, M., Perets, D., Margarit, A., Serebryakova, D., Bogdanov, Y., Bukharev, A. (2018). Application of machine learning for oilfield data quality improvement (Russian). Society of Petroleum Engineers. doi: 10.2118/191601-18RPTC-RU.

Anifowose, F. A., Abdulraheem, A., Al-Shuhail, A. A., and Schmitt, D. P. (2013). Improved permeability prediction from seismic and log data using artificial intelligence techniques. Society of Petroleum Engineers. doi: 10.2118/164465-MS.

Anifowose, F. A., Labadin, J., and Abdulraheem, A. (2017). Ensemble machine learning: An untapped modeling paradigm for petroleum reservoir characterization. *Journal of Petroleum Science and Engineering*, **151**, 480-487.

Anifowose, F., Adeniye, S., Abdulraheem, A., and Al-Shuhail, A. (2016). Integrating seismic and log data for improved petroleum reservoir properties estimation using non-linear feature-selection based hybrid computational intelligence models. *Journal of Petroleum Science and Engineering*, **145**, 230-237.

Archie, G. E. (1942). The Electrical Resistivity Log as an Aid in Determining Some Reservoir Characteristics. Society of Petroleum Engineers. doi: 10.2118/942054-G.

Bandura, L., Halpert, A. D., and Zhang, Z. (2018). Machine learning in the interpreter's toolbox: Unsupervised, supervised, and deep-learning applications. Society of Exploration Geophysicists.

Baruch, J. (2016). Steer driverless cars towards full automation. *Nature*, **536**, 127.

Berger, C. (2014). From a competition for self-driving miniature cars to a standardized experimental platform: concept, models, architecture, and evaluation, *Journal of Software Engineering for Robotics*, **5**(1), 63-79.

Bergman, D. L., Henning, M. J., Sarma, P., and Hunt, I. (2017). Applying statistical learning to quantitative well log analysis. Society of Petroleum Engineers. doi: 10.2118/186044-MS.

Bestagini, P., Lipari, V., and Tubaro, S. (2017). A machine learning approach to facies classification using well logs. *Society of Exploration Geophysicists*.

Bize-Forest, N., Lima, L., Baines, V., Boyd, A., Abbots, F., and Barnett, A. (2018). Using machine-learning for depositional facies prediction in a complex carbonate reservoir. Society of Petrophysicists and Well-Log Analysts.

Carpenter, C. (2016). Geology-driven estimated-ultimate-recovery prediction with deep learning. Society of Petroleum Engineers. doi: 10.2118/0516-0074-JPT.

Chen, D., Jones, C., Dai, B. and van Zuilekom, A. (2018). Fluid optical database reconstruction with validated mapping from external oil and gas information source. Society of Petrophysicists and Well-Log Analysts.

Craddock, P. R., Mossé, L., Prioul, R., Miles, J., Loan, M. L., Pirie, I., Pomerantz, A. E. (2018). Integrating measured kerogen properties with log analysis for petrophysics and geomechanics in unconventional resources. Society of Petrophysicists and Well-Log Analysts.

Crnkovic-Friis, L., and Erlandson, M. (2015). Geology Driven EUR Prediction Using Deep Learning. Society of Petroleum Engineers. doi: 10.2118/174799-MS.

Cuddy, S. J., and Putnam, T. W. (1998). Litho-facies and permeability prediction from electrical logs using fuzzy logic. Society of Petroleum Engineers. doi: 10.2118/49470-MS.

Cui, Y., Wang, G., Jones, S. J., Zhou, Z., Ran, Y., Lai, J., Li, R., Deng, L. (2017). Prediction of diagenetic facies using well logs – A case study from the upper Triassic Yanchang Formation, Ordos Basin, China. *Marine and Petroleum Geology*, **81**, 50-65.

Dashevskiy, D., Dubinsky, V., & Macpherson, J. D. (1999). Application of Neural Networks for Predictive Control in Drilling Dynamics. Society of Petroleum Engineers. doi: 10.2118/56442-MS.

Doveton, J. H. (1986). Log Analysis of Subsurface Geology: Concepts and Computer Methods. John Wiley & Sons, Inc. New York, NY, USA.

Doveton, J. H. (1994). Geologic Log Analysis Using Computer Method: Application in Geology. Volume 2. American Association of Petroleum Geologists. Tulsa, OK, USA.

Du, Q., Yasin, Q., and Ismail, A. (2018). A comparative analysis of artificial neural network and rock physics for the estimation of shear wave velocity in a highly heterogeneous reservoir. Society of Exploration Geophysicists.

Fang, J. H., Karr, C. L., and Stanley, D. A. (1992). Genetic algorithm and its application to petrophysics. Society of Petroleum Engineers.

Frost, E. and Quinn, T. (2018). Data preconditioning for predictive and interpretative algorithms: Importance in data-driven analytics and methods for application. Society of Petrophysicists and Well-Log Analysts.

Gu, M., Gokaraju, D., Chen, D., and Quirein, J. (2016). Shale fracturing characterization and optimization by using anisotropic acoustic interpretation, 3D fracture modeling, and supervised machine learning. Society of Petrophysicists and Well-Log Analysts.

Gupta, I., Rai, C., Devegowda, D., and Sondergeld, C. (2018). Use of data analytics to optimize hydraulic fracture locations along borehole. Society of Petrophysicists and Well-Log Analysts.

Hall, B. (2016). Facies classification using machine learning. *The Leading Edge*, **35**(10), 906–909.

Halpert, A. D. (2018). Deep learning-enabled seismic image enhancement. Society of Exploration Geophysicists.

Han, J., Daigle, H., Tian, X., Pyrcz, M., Griffith, C., and Zhang, B. (2019). A Comparison of Clustering Algorithms applied to Fluid Characterzation using NMR $T1$-$T2$ Maps of Shale. *Computers & Geosciences*. In Press.

He, J., Misra, S., and Li, H. (2018). Comparative study of shallow learning models for generating compressional and shear traveltime logs. Society of Petrophysicists and Well-Log Analysts.

IKTVA. (2017). *Saudi Aramco Report*. http://iktva.sa/.

Ismagilov, A., Sudakov, V., Nurgaliev, D., Murtazin, T., Usmanov, S., and Nugumanova, N. (2018). Machine learning approach to open hole interpretation and static modelling applied to a giant field. Society of Petroleum Engineers. doi: 10.2118/191595-18RPTC-MS.

Izadi, H., Sadri, J., and Bayati, M. (2017). An intelligent system for mineral identification in thin sections based on a cascade approach. *Computers & Geosciences*, **99**, 37-49.

Izadi, H., Sadri, J., and Mehran, N.-A. (2015). A new intelligent method for minerals segmentation in thin sections based on a novel incremental color clustering. *Computers & Geosciences*, **81**, 38-52.

Jacobs, T. (2018). Pioneer's analytics project reveals the good and bad of machine learning. *Journal of Petroleum Technology*, **70**(9).

Jain, V., Gzara, K., Makarychev, G., Minh, C. C., and Heliot, D. (2015). Maximizing information through data driven analytics in petrophysical evaluation of well logs. Society of Petroleum Engineers. doi: 10.2118/174735-MS.

Jobe, T. D., Vital-Brazil, E., and Khaif, M. (2018). Geological feature prediction using image-based machine learning. Society of Petrophysicists and Well-Log Analysts.

Kazak, A., Chugunov, S., and Chashkov, A. (2018). Integration of large-area Scanning-Electron-Microscopy imaging and automated mineralogy/petrography data for selection of nanoscale pore-space characterization sites. Society of Petroleum Engineers. doi: 10.2118/191369-PA.

Khan, M. R., Tariq, Z., and Abdulraheem, A. (2018). Machine learning derived correlation to determine water saturation in complex lithologies. Society of Petroleum Engineers. doi: 10.2118/192307-MS.

Korjani, M. M., Popa, A. S., Grijalva, E., Cassidy, S., and Ershaghi, I. (2016). Reservoir characterization using fuzzy kriging and deep learning neural networks. Society of Petroleum Engineers. doi: 10.2118/181578-MS.

Koryabkin, V., Stishenko, S., Kolba, P., and Tatur, O. (2018). Application of the combined real-time petrophysical and geosteering model to increase drilling efficiency. Society of Petroleum Engineers. doi: 10.2118/191689-18RPTC-MS.

Li, H., & Misra, S. (2017). Prediction of subsurface NMR T2 distributions in a shale petroleum system using variational autoencoder-based neural networks. *IEEE Geoscience and Remote Sensing Letters*, **14**(12), 2395–2397.

Li, H., & Misra, S. (2018). Long Short-Term Memory and Variational Autoencoder With Convolutional Neural Networks for Generating NMR T2 Distributions. *IEEE Geoscience and Remote Sensing Letters*.

Li, H., He, J., & Misra, S. (2018). Data-Driven In-Situ Geomechanical Characterization in Shale Reservoirs. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Li, W. (2018). Classifying geological structure elements from seismic images using deep learning. Society of Exploration Geophysicists.

Ma, S. (Mark). (2018). Technology Focus: Formation Evaluation (August 2018). Society of Petroleum Engineers. doi: 10.2118/0818-0050-JPT.

Ma, Y., Ji, X., BenHassan, N. M., and Luo, Y. (2018). A deep learning method for automatic fault detection. Society of Exploration Geophysicists.

Mahmoud, A. A., ElKatatny, S., Abdulraheem, A., Mahmoud, M., Omar Ibrahim, M., and Ali, A. (2017). New technique to determine the total organic carbon based on well logs using artificial neural betwork (White Box). Society of Petroleum Engineers. doi: 10.2118/188016-MS.

Maniar, H., Ryali, S., Kulkarni, M. S., and Abubakar, A. (2018). Machine-learning methods in geoscience. Society of Exploration Geophysicists.

MacAllister, D. J., Day, R., & McCarmack, M. D. (1996). Expert Systems: A Five-Year Perspective. Society of Petroleum Engineers. doi: 10.2118/26247-PA.

Mehta, A. (2016). Tapping the value from big data analytics. *Journal of Petroleum Technology*, **68**(12). doi: 10.2118/1216-0040-JPT.

Nazari, S., Kuzma, H. A., and Rector, J. W. (2011). Predicting permeability from well log data and core measurements using support vector machines. Society of Exploration Geophysicists.

Negara, A., Jin, G., and Agrawal, G. (2016). Enhancing rock property prediction from conventional well logs using machine learning technique - case studies of conventional and unconventional reservoirs. Society of Petroleum Engineers. doi: 10.2118/183106-MS.

Noshi, C. I., Assem, A. I., and Schubert, J. J. (2018). The role of big data analytics in exploration and production: A review of benefits and applications. Society of Petroleum Engineers. doi: 10.2118/193776-MS.

Odi, U. and Nguyen, T. (2018). Geological facies prediction using computed tomography in a machine learning and deep learning environment. Unconventional Resources Technology Conference. doi: 10.15530/URTEC-2018-2901881.

Ozkan A., Cumella S.P., Milliken K.L., Laubach S.E. (2011) Prediction of lithofacies and reservoir quality using well logs, Late Cretaceous Williams Fork Formation, Mamm Creek field, Piceance Basin, Colorado. *AAPG Bulletin*, **95**(10), 1699-1723.

Parra, J. O., Iturrarán-Viveros, U., Parra, J. S., and Xu, P.-C. (2014). Neural network and rock physics for predicting and modeling attenuation logs. Society of Exploration Geophysicists.

Pham, N., Fomel, S., and Dunlap, D. (2018). Automatic channel detection using deep learning. Society of Exploration Geophysicists.

Pineda, W., Soza, E., Bergeron, J., Williams, J., & Gonzalez, D. (2018). Wireline Formation Fluid Sampling: From Making the Value Case, To Applying the Lessons Learned. A Guide To Improve Rate of Success While Taking Fluid Samples in the Lower for Longer Oil Price Environment. Society of Petrophysicists and Well-Log Analysts.

Proett, M., Musharfi, N., Gill, H., Ma, S., Meridji, Y., & Eyuboglu, S. (2015). Automatically Quantifying Wireline and LWD Pressure-Test Quality. Society of Petrophysicists and Well-Log Analysts.

Rivera, V. P. (1994). Fuzzy Logic Controls Pressure In Fracturing Fluid Characterization Facility. Society of Petroleum Engineers. doi: 10.2118/28239-MS.

Saad, B., Negara, A., and Syed Ali, S. (2018). Digital rock physics combined with machine learning for rock mechanical properties characterization. Society of Petroleum Engineers. doi: 10.2118/193269-MS.

Saputelli, L. (2018). Technology focus: Petroleum data analytics. *Journal of Petroleum Technology*, **70**(10).

Schlanser, K., Grana, D., and Campbell-Stone, E. (2016). Lithofacies classification in the Marcellus Shale by applying a statistical clustering algorithm to petrophysical and elastic well logs. *Interpretation*, **4**(2), SE31-SE49.

Schlanser, K., Grana, D., and Campbell, Erin. (2014). Petro-elastic facies classification in the Marcellus Shale by applying expectation maximization to measured well logs. 659-663. 10.1190/segam2014-0939.1.

Serag El Din, S., Clerke, E., and Badawi, A. (2018). Selected results from big carbonate pore system database: A case study from Saudi Aramco. Society of Petroleum Engineers. doi: 10.2118/193163-MS.

Shabab, M., Jin, G., Negara, A., and Agrawal, G. (2016). New data-driven method for predicting formation permeability using conventional well logs and limited core data. Society of Petroleum Engineers. doi: 10.2118/182826-MS.

Shen, Q., Chen, J., and Wang, H. (2018). Data-driven interpretation of ultradeep azimuthal propagation resistivity measurements: Transdimensional stochastic inversion and uncertainty quantification. Society of Petrophysicists and Well-Log Analysts.

Shin, S., Li, Y., Park, S., and Cho, S.-J. (2018). Supervised machine learning for lithology estimation using spectral induced polarization data. Society of Exploration Geophysicists.

Sidahmed, M., Roy, A., and Sayed, A. (2017). Streamline rock facies classification with deep learning cognitive process. Society of Petroleum Engineers. doi: 10.2118/187436-MS.

Tariq, Z., Abdulraheem, A., Mahmoud, M., and Ahmed, A. (2018). A rigorous data-driven approach to predict Poisson's ratio of carbonate rocks using a functional network. Society of Petrophysicists and Well-Log Analysts.

Theys, P. (2011). Quest for Quality Data. Editions Technip. Paris, France.

Torlov, V., Bonavides, C., & Belowi, A. (2017). Data Driven Assessment of Rotary Sidewall Coring Performance. Society of Petroleum Engineers. doi: 10.2118/187107-MS.

Venkataramanan, L., Evirgen, N., Allen, D. F., Mutina, A., Cai, Q., Johnson, A. C., Jiang, T. (2018). An unsupervised learning algorithm to compute fluid volumes from NMR T_1-T_2 Logs in unconventional reservoirs. Society of Petrophysicists and Well-Log Analysts.

Venna, A. R., Ang, Y. Y., Nguyen, N., Lu, Y., and Walters, D. (2018). Support-vector-machine phase classification of downhole leak flows based on acoustic signals. Society of Petrophysicists and Well-Log Analysts.

Verdin, C. (2010). Fundamental Well Logging Lecture Notes.

Voleti, D. K., Kundu, A., and Singh, M. (2017). Normalized depths as key input and detailed QC steps for improved permeability predictions using existing machine learning techniques. Society of Petroleum Engineers. doi: 10.2118/188493-MS.

Wang, G., Ju, Y., Carr, T. R., Li, C., and Cheng, G. (2014). Application of artificial intelligence on black shale lithofacies prediction in Marcellus Shale, Appalachian Basin. Unconventional Resources Technology Conference. doi: 10.15530/URTEC-2014-1935021

Whittaker, A. H. and Macpherson, J. D. (1986). Assistant expert log data analysis system: Applications of artificial intelligence programming techniques in formation evaluation. Society of Petrophysicists and Well-Log Analysts.

Wikipedia. (2018). Outline of machine learning. (30 November 2018 revision), https://en.wikipedia.org/wiki/Outline_of_machine_learning#Deep_learning (accessed 20 December 2018).

Worthington, P. F. (1997). Recognition and Development of Low-Resistivity Pay. Society of Petroleum Engineers. doi: 10.2118/38035-MS.

Wu, P. -Y., Jain, V., Kulkarni, M. S., and Abubakar, A. (2018). Machine learning-based method for automated well-log processing and interpretation. Society of Exploration Geophysicists.

Wu. S., Pineda W., Lu Z., Prewett H. and Fisher R., (2015). GoM wireline formation tester probability, using data to understand sticking and fishing in past wells and predict probability to fish in future wells: *BP Subsurface technical magazine*, Issue I, 10-11.

Xu, C., Bayer, W. S., Wunderle, M., & Bansal, A. (2016). Normalizing Gamma-Ray Logs Acquired from a Mixture of Vertical and Horizontal Wells in the Haynesville Shale. Society of Petrophysicists and Well-Log Analysts.

Xu, C., and Misra, S. (2018). Petrophysics Data-Driven Analytics: Theories and Applications. SPWLA 2018 STC Panel Discussions Summary. April 16-17, Houston TX. SPWLA Newsletter September 2018.

Xu, H., Nieuwoudt, H., Oloyede, P., & Di Miele, D. (2016). Artificial Intelligence Methods Enhance Formation Testing Accuracy and Efficiency: Challenges and Benefits. Society of Petrophysicists and Well-Log Analysts.

Xu, Y., Sun, K., Xie, H., Zhong, X., Mirto, E., Feng, Y., and Hong, X. (2018). Borehole resistivity measurement modeling using machine-learning techniques. Society of Petrophysicists and Well-Log Analysts.

Yang, L., Bale, D. S., Yang, D., Raum, M., Bello, O., Failla, R., Ye, S. (2018). Enabling Real-Time Asset Analytics for a Cloud-Based Fiber-Optic Data Management System. Society of Petroleum Engineers. doi: 10.2118/191592-MS.

Zhang, Z., Halpert, A. D., Bandura, L., and Coumont, A. D. (2018). Machine learning based technique for lithology and fluid content prediction — Case study from offshore West Africa. Society of Exploration Geophysicists.

Zhao, T., Jayaram, V., Roy, A., and Marfurt, K. J. (2015A). A comparison of classification techniques for seismic facies recognition. *Interpretation*, **3**(4), SAE29-SAE58.

Zhao, T., Verma, S., Devegowda, D., and Jayaram, V. (2015B). TOC estimation in the Barnett Shale from triple combo logs using support vector machine. Society of Exploration Geophysicists.

Zhao, T., Verma, S., Qi, J., and Marfurt, K. J. (2015C). Supervised and unsupervised learning: How machines can assist quantitative seismic interpretation. Society of Exploration Geophysicists.

Zhu, L., Li, H., Yang, Z., Li, C., and Ao, Y. (2018). Intelligent logging lithological interpretation with convolutional neural networks. Society of Petrophysicists and Well-Log Analysts.

Zimmermann, T., Liang, L., and Zeroug, S. (2018). Machine-learning-based automatic well-log depth matching. Society of Petrophysicists and Well-Log Analysts.

Appendix:

List of Special Issues on Machine Learning & Data Analytics Applied in Geosciences and Petroleum Engineering

Authors' Note: the list is ordered with publishing time – from the latest to the oldest.

Publishing Time	Journal	Special Issue Title	Editors	Volume (Issue)
Summer 2019	Mathematical Geosciences	Data Science in Geosciences	Vasily Demianov, Erwan Gloaguen, Mikhail Kanevski	--
Aug 2019	Interpretation	Insights into digital oil field data using artificial intelligence and big data analytics	Vikram Jayaram, Andrea Cortis, Bill Barna, Atish Roy, Deepak Devegowda, Jacqueline S. Floyd, Pradeepkumar Ashok, Satyam Priyadarshy, Aria Abubakar, Chiranth Hegde, Emmanuel Schnetzler	--
Aug 2019	Interpretation	Machine Learning in Seismic Data Analysis	Haibin Di, Lei Huang, Mauricio Araya-Polo, Youzuo Lin, Anne Solberg, Tao Zhao, Xinming Wu, Vikram Jayaram, Jun Cao, Satinder Chopra, Erwan Gloaguen, Saleh Al-Dossary, Ghassan AlRegib, Hongliu Zeng	--
Dec 2018	Petrophysics	Data-Driven Analytics in Logging and Petrophysics	Chicheng Xu, Jeffry Hamman, Jesús M. Salazar, Shouxiang Mark Ma, Susan Agar, Yang Liu, Carlos Torres-Verdin, Stephen Prensky	59 (6)
May 2018	Computers & Geosciences	Big Data and Natural Disasters: New Approaches for Spatial and Temporal Massive Data Analysis	Francisco Martínez-Álvarez, Antonio Morales-Esteban	Volume 110-1115
Aug 2017	Interpretation	Computer-Assisted Seismic Interpretation Methods	David H. Johnston, Geoffrey Dorn, Sergey Fomel, Jesse Lomask, Murray Roth, and Tracy Stark	5(3)
Jan 2016	Geoscience Frontiers	Progress of Machine Learning in Geosciences	Amir H. Alavi, Amir H. Gandomi, David J. Lary	7 (1)
Dec 2015	Computers & Geosciences	Statistical Learning in Geoscience Modelling: Novel	Vasily Demyanov, Mikhail Kanevski	85(B)

Publishing Time	Journal	Special Issue Title	Editors	Volume (Issue)
		Algorithms and Challenging Case Studies		
Nov 2015	Interpretation	Pattern Recognition and Machine Learning	Vikram Jayaram, Per Age Avseth, Kostia Azbel, Theirry Coléou, Deepak Devegowda, Paul de Groot, Dengliang Gao, Kurt Marfurt, Marcilio Matos, Tapan Mukerji, Manuel Poupon, Atish Roy, Brian Russell, Brad Wallet, and Vikas Kumar	3(4)
Nov 2014	Journal of Petroleum Science and Engineering	Neural network applications to reservoirs: Physics-based models and data models	Tahar Aifa	123
Nov 2009	Computers & Geosciences	Progressive Transmission of Spatial Datasets in the Web Environment	Bisheng Yang, Robert Weibel	35 (11)
May 2007	Neural Networks	Computational Intelligence in Earth and Environmental Science	Vladimir Cherkassy, William Hsieh, Vladimir Krasnopolsky, Dimitri Solomatine, Julio Valdes	20(4)
Mar 2006	Neural Networks	Earth Sciences and Environmental Applications of Computational Intelligence	V. Cherkassky, V. Krasnopolsky, D. Solomatine, J. Valdes	19(2)
Dec 2005	Journal of Petroleum Science and Engineering	An Introduction to Artificial Intelligence Applications in Petroleum Exploration and Production	Ridha B.C. Gharbi, G. Ali. Mansoori	49(3-4)
May 2005	Journal of Petroleum	Intelligent Computing	Leonid Sheremetov, Matías Alvarado, Fred Aminzadeh, René Bañares-Alcántara, G. Ali Mansoori	47(1-2)

Publishing Time	Journal	Special Issue Title	Editors	Volume (Issue)
	Science and Engineering	in Petroleum Engineering		
Apr 2003	Neural Networks	Neural Nework Analysis of Complex Scientific Data: Astronomy and Geosciences	Roberto Tagliaferri et al.	16(3-4)
Nov 2001	Journal of Petroleum Science and Engineering	Soft Computing and Earth Sciences (Part 2)	M. Nikravesh, F. Aminzadeh, L.A. Zadeh	31(2-4)
May 2001	Journal of Petroleum Science and Engineering	Soft Computing and Earth Sciences	M. Nikravesh, F. Aminzadeh, L.A. Zadeh	29(3-4)
Oct 2000	Computers & Geosciences	Applications of Virtual Intelligence to Petroleum Engineering	Shahab Mohaghegh	26(8)