Data-Driven Reservoir Modeling

Data-Driven Reservoir Modeling

Top-Down Modeling (TDM)

A Paradigm Shift in Reservoir Modeling
The Art and Science of Building Reservoir Models
Based on Field Measurements

Shahab D. Mohaghegh
Intelligent Solutions Incorporated and West Virginia University
As of this printing:
http://intelligentsolutionsinc.com/Products/IMagine.shtml

2017

Society of Petroleum Engineers

Printed in the United States of America.

Disclaimer

This book was prepared by members of the Society of Petroleum Engineers and their well-qualified colleagues from material published in the recognized technical literature and from their own individual experience and expertise. While the material presented is believed to be based on sound technical knowledge, neither the Society of Petroleum Engineers nor any of the authors or editors herein provide a warranty either expressed or implied in its application. Correspondingly, the discussion of materials, methods, or techniques that may be covered by patents implies no freedom to use such materials, methods, or techniques without permission through appropriate licensing. Nothing described within this book should be construed to lessen the need to apply sound engineering judgment nor to carefully apply accepted engineering practices in the design, implementation, or application of the techniques described herein.

ISBN 978-1-61399-560-0

First Printing 2017

Society of Petroleum Engineers
222 Palisades Creek Drive
Richardson, TX 75080-2040 USA

http://www.spe.org/store
service@spe.org
1.972.952.9393

Dedication

This book is dedicated to Turgay Ertekin. He was, is, and will always be my mentor and role model. I consider myself fortunate to have known him and to have worked under his supervision as a graduate student. What I have learned from him has guided me not only in reservoir engineering, but in life.

Acknowledgments

I would like to acknowledge and thank my colleagues at Intelligent Solutions Incorporated, Razi Gaskari and Mohammad Maysami, for their invaluable contributions throughout the many years that we have worked together.

I would like to acknowledge and thank all my graduate students throughout the past 25 years. They have enriched my life, and many have become life-long friends.

I am thankful to Narges for her unwavering support and for putting up with my tough schedule, and I am thankful to Dorna for giving my life meaning.

Foreword

Data-Driven Reservoir Modeling is intended to introduce a technology that is relatively new to petroleum engineers and geoscientists whose day-to-day job responsibilities always bring them to junctures where critical technical decisions need to be made and strategies need to be established. The technology covered in this book adds another decision-making tool to the arsenal of upstream technologists of the petroleum industry. This book should also be useful to petroleum engineering and geosciences undergraduate students in their junior or senior year, as well as to graduate students with some degree of exposure to the principles of petroleum engineering field operations, petroleum geology, and petroleum geophysics.

The aim of this book is to present a methodology that is rather new to the petroleum engineering community and is particularly suited to the application of data analytics to physical problems of reservoir engineering for tracking the state of dynamics with the goal of strengthening the decision-making process. With the help of the pragmatic approach provided in this book, data-driven modeling can be effectively used in field planning and development studies.

In today's field practices, zillions of bytes of information are generated daily. Every piece of data carries key signatures about the physical properties of the system being studied and about the ongoing physical and thermodynamic processes. The collected data can be so massive that they overwhelm manpower, while the available computational power may not permit conducting a comprehensive analysis. In addition to these issues, if we are dealing with data that are generated by relationships that are not understood or, at best, vaguely understood (i.e., all the physics and thermodynamics of the ongoing processes are not well known), reservoir analysis becomes even more challenging. In times like these, machine-learning-based algorithmic protocols (intelligent systems) come to the rescue. These systems are knowledge-based intelligent systems that emulate not only human intelligence but also the reasoning and decision-making aspects of human intelligence. In our daily operations, we are always confronted with uncertain data; here, the strategy is to exploit imprecision and uncertainty to achieve tractable, robust, and low-cost solutions. This is why capturing associations and discovering existing regularities in a big data set can become a reality even if the diversity of the data is large and the relationships between independent and dependent variables are understood only dimly.

The book assumes some degree of familiarity with the upstream petroleum industry vocabulary and the physics of flow in porous media. In order to maximize the benefits from the book, one also needs to be knowledgeable about transport processes and the thermodynamics of oilfield fluids. With that knowledge bank in place, it will be possible to critically analyze the results generated. Being conversant with computers on various platforms will be helpful if the reader is interested not only in using this class of solutions but also in developing such a catalogue of solutions.

The author of the book, Shahab Mohaghegh, is a leading authority in the application of data analytics in petroleum engineering. His writing style is extremely lucid and informative. Each chapter of the book is well structured and possesses logical continuity, clarity, and thoroughness. In my view, the book brings a wonderful opportunity to explore new modeling frontiers that are applicable to hydrocarbon reservoirs. No matter what type of reservoir or production engineering problem you are working on, Big Data analytics, when applied properly, has the potential to guide and streamline the solution work flow. Here, I present a summary of how the author covers topical areas of data-driven reservoir modeling.

Chapter 1 briefly reviews the reservoir models that are currently used in studies related to reservoir management and discusses the challenges of history matching, which is a critically important and notoriously difficult application for reservoir characterization. The chapter continues with a discussion of top-down modeling (TDM) and discusses how physical processes and geological characteristics collectively play an important role in generating the response function from the field (typically, pressure-transient and/or rate-transient data).

Chapter 2 provides a succinct review of the theory of data-driven problem-solving methodology. This chapter stresses the importance of powerful domain expertise in applying machine learning and pattern recognition to the class of problems studied in the book.

Chapter 3 outlines the historical progression of reservoir modeling and discusses the critical juncture where a decision needs to be made on the levels of accuracy and computational overhead that are faced during a full-fledged simulation study. What makes the decision even more challenging is that engineers who are conducting reservoir simulation studies almost always find themselves in the middle of a sea of uncertainties.

Chapter 4 concisely covers Big Data analytics methodologies, including data mining, artificial intelligence, artificial neural networks, and fuzzy logic, with greater emphasis on the last two. This chapter should be especially useful for readers who do not have great familiarity with data analytics practices.

Chapter 5 accentuates the importance of good understanding of the conventional reservoir-modeling techniques and principles of machine learning to appreciate the advantages and disadvantages of each of these two broadly dissimilar approaches.

Chapter 6 reminds the reader that there are empirical models that use the data collected in analyzing the performance of a hydrocarbon reservoir (e.g., decline curve analysis). This chapter also discusses some of the weaknesses that we face in the application of such empirical methodologies.

Chapter 7 introduces the concept of TDM as a new work flow in data-driven reservoir-modeling applications. The principal strength of TDM is recognized in terms of its encompassing approach, such that all available field measurements can be integrated in a seamless manner. What is more striking here is that even if we do not have a complete understanding of the dynamics of the ongoing physical phenomena, the TDM protocol is capable of generating a representative comprehensive model that incorporates all of the available data.

Chapter 8 discusses the spatio-temporal nature of the data base that is inevitably faced in reservoir engineering studies. The nonlinear nature of the process dynamics and parameters that are involved in the processes undoubtedly makes reservoir modeling even more challenging. The matters that need to be addressed at this stage typically include further simplification of the model. However, most of the assumptions that are used in such simplifications may not be compatible with the nature of the data collected, because these data internally carry many critical implications of the nonlinearities. Along these lines, the impact of static and dynamic parameters on the response functions is discussed.

Chapter 9 addresses the nonunique nature of inverse solutions (in this case, history-matching protocols). Like any other inverse model, a model that matches the history successfully cannot guarantee the accuracy of the predictions; however, a good-quality history match increases the level of confidence about the performance of the model. In a history-matching application spatio-temporal properties are accommodated with the help of artificial neural networks. In this process, a critical step is the correct prioritization of the relevance of such parameters. This enhances the overall optimization of the process that is being studied.

Chapter 10 covers the importance of conducting a critical analysis of the results of the TDM. Such an analysis will provide an opportunity to optimize the process that is being modeled and at the same time will establish realistic bounds to the results that are being generated. These limitations can be accommodated by keeping realistic and practical bounds on the operational parameters. In view of the high computational speeds that can be achieved by the TDM approach, it will be possible to conduct an expansive Monte Carlo simulation study using a large number of scenarios. In this chapter, the dynamic nature of the TDM is reiterated, just to ensure that the user does not forget to update the overall structure of the TDM whenever necessary.

Chapter 11 presents a compendium of three case studies involving mature oil fields from different corners of the world.

Chapter 12 discusses the limitations of data-driven reservoir modeling, indicating the importance of the representative nature of the data available in developing the model. This becomes especially important because the data used also carry vital information about the physical processes within the system.

Chapter 13 provides a discussion based on the author's extensive experience about what one might expect to see in the future concerning the use of data-driven reservoir modeling. One potential area where this type of modeling will be handy is in the analysis of fiber-optic data-collection systems that are finding applications in long horizontal wellbores. For example, it is believed that even small temperature variations captured in the horizontal well will provide critically important information about the identification of the producing and nonproducing zones.

Finally, extensive references are provided for any reader who is interested to learn more about data-driven reservoir modeling.

We hope you now have a good idea of what this volume is all about and what it can do for the problems that you are working on. I am confident that this book will equip you with what you need to know in order to develop realistic solutions for problems you may have thought that you would not be able solve. Here is your opportunity.

Turgay Ertekin
Professor of Petroleum and Natural Gas Engineering
Pennsylvania State University
University Park, Pennsylvania, USA
26 November 2016

Table of Contents

Chapter 1

Introduction

We are living in an interesting time. Let us put the speed at which technology is changing our world in perspective. Take the example of the printing press vs. electronic mail or email. The printing press was invented in the fifteenth century and changed how people communicated. It was the most common mode of communication for more than 25 generations.[1] It still is an integral part of our lives. But no one doubts that many of its functions are now performed by new technologies such as email, the Internet, eBooks, eNews, and so on. Furthermore, email became popular only after the 1990s. It quickly grew to become the most used mode of communication among people of all persuasions. However, now, in less than only one generation, it is losing its prominent position in human communications to other newly invented communication modes. The new generation hardly uses email. Email is now being replaced by text messages and social network communications. Email, such a step-change in our communications, is ready to relinquish its prominence in less than one generation. This is the speed at which the technology is moving forward.

The modern oil industry is a bit more than one hundred years old.[2] Most technologies that are currently in use by petroleum engineers and geoscientists can be traced back to their development for the oil industry during the mid-twentieth century.[3] Some of our newer technologies (e.g., horizontal wells, seismic survey, measurement while drilling) are less than a few decades old. This book presents an example of a new and quite recent technology that is making its way into the oil and gas industry. This book is dedicated to application of artificial intelligence and data mining in the upstream oil and gas industry, and specifically to their application to build comprehensive reservoir models.

1.1 Reservoir Models for Reservoir Management

It has been said that all models are wrong, but some models are useful (Box 1976). One of the objectives of reservoir engineers is to build reliable reservoir models to be used by reservoir managers in order to make decisions. The uniqueness and complexity of each hydrocarbon reservoir make the accomplishment of this objective quite challenging. For the purposes of this book, we define the problem as follows:

Building full-field models that have ***practical*** *utility for managing complex reservoirs. In order to fit this mold, the full-field reservoir model must be* ***accurate*** *and have a* ***small computational footprint****.*

Complementary stipulations for such a model include incorporation of all available information about the reservoir, including detailed information about all wells. The two attributes that are emphasized in the above definition are accuracy and small computational footprint. The model has to be

[1]Each century can be counted as four to five generations.

[2]Actual oil production for commercial use goes back to the mid-1820s in Imperial Russia.

[3]Darcy's law goes back to 1856, and Arp's decline curves were introduced in 1945.

accurate so that it can honor the past (history matched) before any comments can be made regarding its predictive capabilities. Furthermore, to be considered a viable reservoir management tool, the model's computational footprint should be small enough to warrant sensitivity analyses, quantification of uncertainties, and exploration of large solution spaces for field development planning.

The importance of accuracy of a full-field reservoir model should be obvious. Nevertheless, it must be emphasized that by history matching the past performance of the reservoir, which is usually preserved in the form of production history, one develops confidence in the utility of the model for further analyses. As the number of dynamic measurements in the field increases, so too does the complexity of the history-matching process. For example, when flowing bottomhole pressure or production rates are the only measured dynamic properties that are to be history matched (usually one of the two is used as the constraint, while the other one is calculated), the reservoir modeler has a much easier task to accomplish (and a better chance of success) as compared to the cases where other dynamic data such as Gas Oil Ratio (GOR), Water Cut (WC), time-lapsed saturation and static reservoir pressure (as a function of time) have been measured (and are present) in the history and need to be simultaneously history matched.[4]

The importance of the computational footprint cannot be overemphasized. It strikes at the heart of the problem: the model utility. This is a practical issue. No matter how accurate a full-field reservoir model is, a large computational footprint can make it useless in practice.[5] Reducing the computational footprint of numerical models is the main reason companies have moved toward supercomputers and clusters of parallel central processing units. The never-ending race between larger models (high resolution in space and time) and faster computers continues to preoccupy many of our colleagues in the larger companies.

Tasks such as sensitivity analysis and quantification of uncertainties associated with the geological (static) model (the backbone of any reservoir model) are complex and time-consuming undertakings. As the run time (time for execution after the development is completed) of a model starts increasing to more than a few hours, performing such analyses becomes impractical. The solution is either to cut corners or to come up with tricks to perform fewer runs and make the most out of the results. This last exercise (reducing the number of runs) starts introducing limitations in the analyses. We pay a price by reducing the number of runs. Sometimes unjustifiable assumptions of linearity must be made in order to make certain approaches work. There is very little we can do. It is simply a conundrum that we cannot simplify our way out of. As long as we live by the laws of the current paradigm,[6] there is only so much that we can do.

Equally complex and puzzling are field development and planning problems. As the number of wells (both producers and injectors) and the type of constraints increase, so too does the solution space that needs to be searched for optimal (or near-optimal) solutions. *Solution space* in this context is defined as the number of possible combinations of wells and their associated constraints that can form a solution (combinatorial explosion). Any optimization routine, no matter how smart it may be, requires examining a large number of solutions in order to find the optimal or the near-optimal solutions. Each solution in the case of a development plan means at least a single run of the reservoir model. It is easy to see that as the execution time (computational footprint) of a reservoir model increases, its utility as an objective function for planning is compromised. Again, many

[4]Needless to say, just because a model can history match the past does not mean that its predictive capabilities are guaranteed. Those familiar with the history-matching process are well aware that history matching is an art as much as a science. Unfortunately, performing unreasonable tuning and modification of parameters to reach a history match is not an uncommon practice in our industry. A history-matched model is a nonunique model, by definition.

[5]Especially in the eyes of reservoir managers, who most of the time are the primary users of reservoir simulation models as a tool.

[6]The current paradigm, known as the computational paradigm, states that building models includes development of governing equations using first-principles physics and then using discrete mathematics to numerically solve the governing equations.

reservoir engineers continue their quest to build proxy models that are simplified versions of the more-complex models, so that they can be used for such purposes, but there is always a price to pay. Sometimes the price we end up paying is so severe that it undermines the original efforts of building such complex and detailed numerical reservoir models from the very start.

Are there solutions for problems such as those mentioned above? Yes and no. For as long as we are sticking to the traditional paradigm of building reservoir models, there are no solutions. Here, by the traditional paradigm, we are referring to the well-known sequence of building the geological (static) model and then using the principles of fluid flow in porous media to develop a dynamic model based on the numerical solutions of the partial-differential equation that governs fluid flow in porous media, the so-called numerical reservoir simulation and modeling, and finally modifying the static model in order to history match the dynamic model.

As long as we adhere to these core principles, we are simply pushing the envelope. There are, and continue to be, successes and failures. Incremental gains are made here and there. But if we want to remove this serious practical shortcoming altogether, we need to move beyond the traditional paradigm of building reservoir models. The solution requires a paradigm shift. This paradigm shift and its manifestation in reservoir modeling are the subjects of this book.

This is not a general book that discusses the virtues and capabilities of data-driven analytics and indicates areas of the upstream oil and gas industry that can benefit from machine learning and data mining. This book is very specific. It has identified one specific area in our industry, namely reservoir modeling, and provides ample details on how this new and exciting technology (artificial intelligence and data mining) can be used to build new reservoir models. From that point of view, it is the only book of its kind in the oil and gas industry that provides step-by-step details in how to build a data-driven reservoir model, also known as a Top-Down Model (TDM). The following section is a summary of what can be expected in this book.

1.2 What Is Top-Down Modeling?

To efficiently develop and operate a petroleum reservoir, it is important to have a model. Currently, numerical reservoir simulation is the accepted and widely used technology for this purpose. Data-driven reservoir modeling (also known as top-down modeling or TDM[7]) is an alternative (or a complement) to numerical reservoir simulation. TDM uses a Big Data solution (machine learning and pattern recognition) to develop (train, calibrate, and validate) full-field reservoir models based on field measurements (facts) rather than mathematical formulations of our current understanding of the physics of the fluid flow through porous media.

Unlike other empirical technologies, which only use production data as a tool to forecast production,[8] or only use production/injection data for its analysis,[9] TDM integrates all available field measurements in order to forecast production from every single well in a field with multiple wells. The field measurements that are used by TDM to build a full-field reservoir model include well locations and trajectories, completions and stimulations, well logs, core data, well tests, seismic, and production/injection history (including wellhead pressure and choke setting). TDM combines all the information from the sources mentioned above into a cohesive, comprehensive, full-field reservoir model using artificial intelligence technologies. A top-down model is defined as a full-field model within which production [including gas/oil ratio (GOR) and water cut (WC)] is conditioned to all the measured reservoir characteristics and operational constraints. TDM matches the historical production (and is validated through blind history matching) and is capable of forecasting a field's future behavior on a well-by-well basis. Imagine a decline curve analysis technique that covers the entire field, well by well, and incorporates reservoir characteristics and operational constraints and

[7]Throughout this book "data-driven reservoir modeling" and "top-down modeling" are used interchangeably.

[8]Decline curve analysis

[9]Capacitance/resistance model

accounts for interaction between wells whether they are only producers or a combination of producers and injectors. TDM is such a tool.

The novelty of data-driven reservoir modeling stems from the fact that it is a complete departure from traditional approaches to reservoir modeling. Fact-based, data-driven reservoir modeling manifests a paradigm shift in how reservoir engineers and geoscientists model fluid flow through porous media. In this new paradigm, current understanding of physics and geology in a given reservoir is replaced by facts (data/field measurements) as the foundation of the model. This characteristic of TDM makes it a viable modeling technology for unconventional (shale) assets where the physics of the hydrocarbon production (in the presence of massive hydraulic fractures) is not yet well understood.

1.2.1 Role of Physics and Geology. Although it does not start from the first principles of physics, a top-down model is very much a physics-based reservoir mode. The incorporation of physics in TDM is quite non-traditional. Reservoir characteristics and geological aspects are incorporated in the model insofar as they are measured. Although interpretations are intentionally left out during the model development, reservoir engineering knowledge plays a vital role in the construction of the top-down model. Furthermore, expert knowledge and interpretation are extensively used during the analysis of model results. Although fluid flow through porous media is not explicitly (mathematically) formulated during the development of data-driven reservoir models, successful development of such models requires a solid understanding and experience in reservoir engineering and geosciences. Physics and geology are the foundation and the framework for the assimilation of the data set that is used to develop the top-down model. The diffusivity equation has inspired the invention of this technology and was the blueprint upon which TDM was developed.

1.2.2 Formulation and Computational Footprint. The top-down model is built by correlating[10] flow rate at each well and at each timestep[11] to a set of measured static and dynamic variables. The static variables include reservoir characteristics such as well logs (e.g., gamma ray, sonic, density, resistivity), porosity, formation tops and thickness, and others at the following locations:

1. At and around each well
2. The average from the drainage area of each well
3. The average from the drainage area of the offset producers
4. The average from the drainage area of the offset injectors

The dynamic variables include operational constraints and production/injection characteristics at the appropriate timestep when production is being calculated (estimated), such as

1. Wellhead or bottomhole pressure, or choke size, at timestep t
2. Completion modification (e.g., operation of inflow-control valve, squeeze off) at timestep t
3. Number of days of production at timestep t
4. GOR, WC, and oil production volume at timestep t–1
5. Water and/or gas injection at timestep t
6. Well stimulation details
7. Production characteristics of the offset producers at timestep t–1

[10]Correlation that is conditioned to causation will be discussed in more detail in subsequent sections of this book.

[11]The length of timesteps in TDM is a function of the production characteristics and the age of the reservoir as well as the availability of data. It is usually either daily, monthly, or annual.

The data (enumerated above) that are incorporated into the top-down model show its differences from other empirically formulated models. Once the development of the top-down model is completed, its deployment in forecast mode is computationally efficient. A single run of the top-down model is usually measured in seconds or in some cases in minutes. Size of a top-down model is determined by the number of producer and injector wells. The small computational footprint makes TDM an ideal tool for reservoir management, uncertainty quantification, and field development planning. Development and deployment costs of TDM are a small fraction of that for numerical reservoir simulation.

1.2.3 Expected Outcome of a Top-Down Model. Data-driven reservoir modeling can accurately model a mature hydrocarbon field and successfully forecast its future production behavior under a large variety of operational scenarios. Outcomes of TDM are forecast for oil production, GOR, and WC of existing wells as well as field development planning and infill drilling. When TDM is used to identify the communication between wells, it generates a map of reservoir conductivity that is defined as a composite variable that includes multiple geologic features and rock characteristics contributing to fluid flow in the reservoir. This is accomplished by deconvolving the impact of operational issues from reservoir characteristics on production.

1.2.4 Limitations of TDM. Data-driven reservoir modeling is applicable only to fields with a certain amount of production history; for this reason, TDM is not applicable to greenfields and fields with a small number of wells and short production history. Another limitation of TDM is that it is not valid once the physics of the fluid flow in a field goes through a complete and dramatic change. For example, once a top-down model is developed for a field under primary recovery, it cannot be applied to enhanced- recovery phases of the same field.

1.2.5 Software Tool for the Development of TDM. Top-Down Modeling was pioneered by Intelligent Solutions Incorporated (ISI). ISI has recently released a software product for the development and deployment of Top-Down Model, called IMagine™. At the time of publication of this book, no other company has announced a similar product for the development of top-down models.

1.3 Paradigm Shift

Paradigm shift, a term first coined by Thomas Kuhn (1996), constitutes a change in basic assumptions within the ruling theory of science. According to Kuhn, “A paradigm is what members of a scientific community, and they alone, share” (Kuhn 1977). Jim Gray, the American computer scientist who received the Turing Award for his seminal contribution to computer science, once said, “Originally, there was just experimental science, and then there was theoretical science, with Kepler’s Law, Newton’s Law of Motion, Maxwell’s Equations, and so on. Then for many problems, the theoretical models grew too complicated to solve analytically, and people had to start simulating. These simulations have carried us through much of the last half of the last millennium. At this point, these simulations are generating a whole lot of data, along with a huge increase in data from the experimental sciences. People now do not actually look through telescopes. Instead, they are ‘looking’ through large scale, complex instruments which relay data to datacenters, and only then look at the information on their computers.… The new model is for data to be captured by instruments or generated by simulations before being processed by software and for the resulting information or knowledge to be stored in computers.… The techniques and technologies for such data-intensive science are so different that it is worth distinguishing data intensive science from computational science as a new, *fourth paradigm* for scientific exploration” (Bell et al. 2009).

So how does this paradigm shift constitute itself in the exploration and production industry? Do we collect and/or generate (either from our simulators or from instruments) enough data to benefit from such a paradigm shift? In this book, we examine the reservoir-modeling discipline in the exploration and production industry in order to address these questions.

What is the current paradigm of building models that attempts to explore and explain fluid flow in porous media? We use analytical as well as numerical approaches both at the well level and at the reservoir level. This is the current paradigm. In our attempts at analytical solutions, where well testing is a good example, we approximate the problem in order to come to an exact analytical solution. Assumptions such as reservoir homogeneity, well-defined reservoir boundaries, and single-phase flow are among the basic assumptions that are required in order for the analytical solutions to be applicable. On the other hand, when we try to define the problem more realistically, by incorporating reservoir heterogeneity, irregular reservoir boundaries, multiple wells, and multiphase flow, then we are forced to discretize the problem in time and space and generate a large number of linear equations that can be solved numerically. Then, we use numerical solutions to solve the system of linear equations that was generated. Numerical solutions to the partial-differential equations are only approximate solutions. In other words, in attempts to minimize assumptions in the problem, we approximate the solutions. This paradigm has served our industry for decades and has resulted in many scientific and practical advances. We are not advocating that this be shelved. Data-driven reservoir modeling provides an alternative to (and in many cases a complement to) the traditional analytical and numerical approaches to modeling fluid flow in porous media. It is the paradigm shift as it is applied to reservoir modeling and reservoir management.

When it comes to reservoir modeling, we generate or collect massive amounts of data throughout the life of a hydrocarbon-producing field. Data-driven reservoir modeling proposes the use of this massive amount of data that is collected in the form of drilling characteristics, well construction and trajectories, well logs of all different natures, core data, well tests, seismic surveys, and finally production and injection histories along with pressure measurements, in order to build a reservoir model that is entirely based on these field measurements and minimizes the incorporation of our interpretations and biases into the resulting model. Let us examine examples where large amounts of data are collected during oil and gas exploration and production.

1.3.1 Drilling Operation. The modern drilling operation generates hundreds of gigabytes of data on a daily basis. Measurement while drilling (MWD) and logging while drilling (LWD), which have been around for years, generate considerable amounts of data in real time while the drilling operation is ongoing. Complementing these data with seismic surveys and geological models that are developed for a given field includes incredible amounts of information about the field that can help increase drilling efficiencies and eventually move the industry toward completely autonomous drilling operation.

1.3.2 Mature Fields. Mature fields around the world are sources of vast amounts of data and information that have been collected over decades. Mature fields usually include a large number of wells that have been drilled throughout its history. The fact that the wells have been drilled at different time periods provides valuable insight into the fluid flow as well as pressure and saturation distribution throughout the field. This includes large amounts of historical production and injection data usually with the associated wellhead pressure or choke settings. Most of the wells, if not all of them, have the basic set of well logs. Several wells will have been cored, and therefore some core analyses are also available. Usually, well tests are available, and sometimes seismic surveys have been performed (sometimes more than once).

In many cases, the size of the mature field determines the amount of data that can be expected to be available. More-prolific fields usually are blessed with larger amounts of data and more- diverse types of data. The amount and the variety of the data available on some prolific mature fields can be staggering. So much so, that it overwhelms reservoir engineers and reservoir managers. Many times in such cases large amounts of data will go unused and unanalyzed.

1.3.3 Smart Completions, Smart Wells, and Smart Fields. Smart fields have two major characteristics. First, they include smart completions with controls and measurements taking place at different locations along the completion, and, second, installation of permanent downhole gauges provides high-resolution data streams. Even haphazardly designed smart fields have generated massive amounts of data that hardly ever are looked at, even offline.

Smart completions let engineers intervene with details of wells' operations from a distance. Smart wells transmit nearly continuous (real-time) data streams (e.g., pressure, flow rate) to the remote office providing immediate feedback on the consequences of recently made decisions and actions taken. Smart fields include multiple smart wells providing the possibility of managing the entire reservoir remotely and in real time. Smart fields generate terabytes[12] and petabytes[13] of data that are good examples of Big Data in the upstream oil and gas industry.

1.3.4 Production From Shale Assets. We have been witnessing an incredible increase in hydrocarbon production from source rocks, such as shale, in recent years. Production from shale has been made possible by drilling long lateral wells and then stimulating them using massive, multiple stages of hydraulic fractures. During this process, operators are now collecting large amounts of data that include well-construction data, reservoir characteristics in the form of well logs, completion data including much detail about each cluster of hydraulic-fracturing procedures, and finally detailed production data. This is a massive amount of data, especially when we consider that the well count in shale assets is in the hundreds.

Furthermore, the introduction of distributed temperature sensing and distributed acoustic sensing systems is adding a whole new dimension to the important data that can be collected during oil and gas production in shale wells. Once the distributed temperature sensing and distributed acoustic sensing data can be coupled with microseismic and other reservoir and production-related data that have been collected from the shale wells, engineers and geoscientists will have a realistic shot at understanding all the complexities that are associated with hydrocarbon production from shale in the presence of coupled induced and natural fractures. There should be no doubt in anyone's mind that the only technology that has a realistic shot at making all this possible is advanced data-driven analytics, also known as "Shale Analytics."[14]

1.3.5 Reservoir Simulation Models. It is a well-known fact that reservoir simulation models generate massive amounts of data. Actually the amount of data that is generated by reservoir simulation models is so large that in order just to look at them for analysis, special visualization tools are required. Imagine that a large reservoir of tens or even hundreds of square miles that includes multiple distinct geological layers is being modeled. Furthermore, imagine that a large number of wells has been modeled in this reservoir. Having reservoir simulation models with tens and or even hundreds of millions of gridblocks is becoming standard in today's oil and gas industry.

Now imagine generating oil, gas, and water production from the wells. Such a model generates gigabytes or terabytes of data for each of its timesteps, which include not only the static characteristics of the reservoir that has been modeled but also the resulting pressure and saturation distributions throughout the time and space for each gridblock. If the simulation is compositional, the amount of generated data increases exponentially.

[12]A terabyte of data is 10^{12} bytes of data.

[13]A petabyte of data is 10^{15} bytes of data.

[14]A book by this name (*Shale Analytics*) has recently been published by Springer-Verlag.