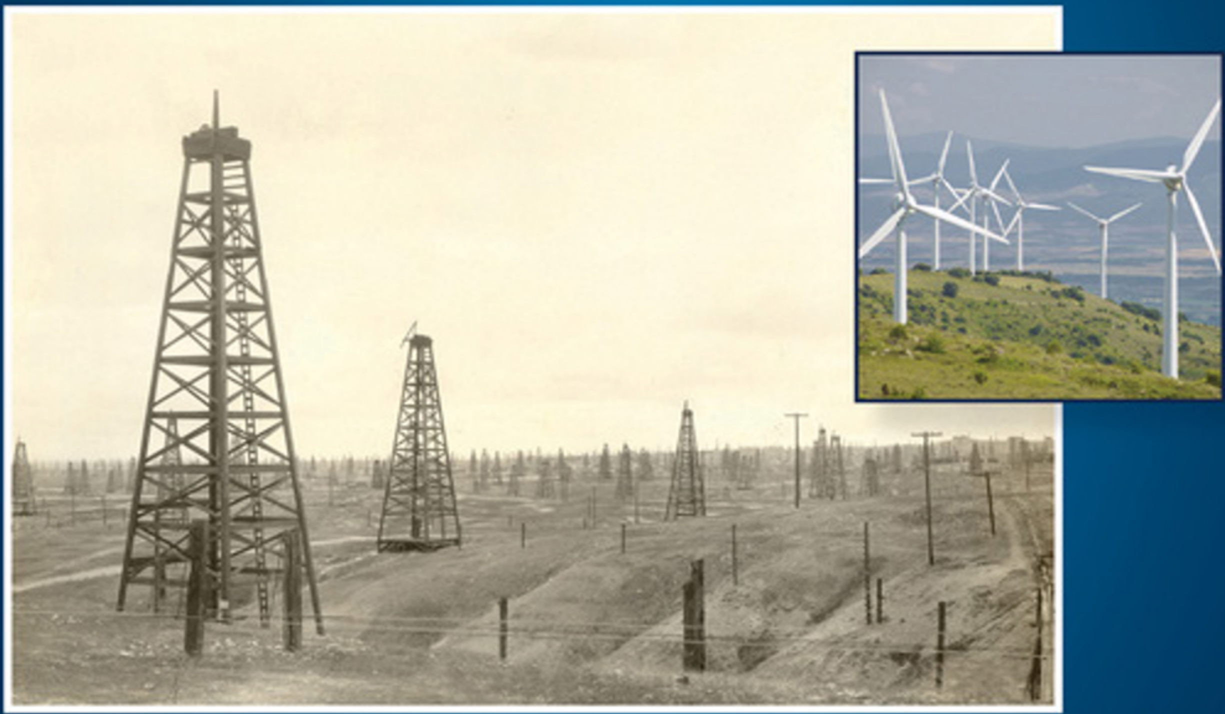
Mathematical Geoenergy

Discovery, Depletion, and Renewal



Paul Pukite, Dennis Coyne, and Daniel Challou

AGU100 ADVANCING EARTH AND SPACE SCIENCE

WILEY

Geophysical Monograph Series

Geophysical Monograph Series

189 **Climate Dynamics: Why Does Climate Vary?** *De-Zheng Sun and Frank Bryan (Eds.)*
190 **The Stratosphere: Dynamics, Transport, and Chemistry** *L. M. Polvani, A. H. Sobel, and D. W. Waugh (Eds.)*
191 **Rainfall: State of the Science** *Firat Y. Testik and Mekonnen Gebremichael (Eds.)*
192 **Antarctic Subglacial Aquatic Environments** *Martin J. Siegert, Mahlon C. Kennicut II, and Robert A. Bindschadler (Eds.)*
193 **Abrupt Climate Change: Mechanisms, Patterns, and Impacts** *Harunur Rashid, Leonid Polyak, and Ellen Mosley-Thompson (Eds.)*
194 **Stream Restoration in Dynamic Fluvial Systems: Scientific Approaches, Analyses, and Tools** *Andrew Simon, Sean J. Bennett, and Janine M. Castro (Eds.)*
195 **Monitoring and Modeling the Deepwater Horizon Oil Spill: A Record-Breaking Enterprise** *Yonggang Liu, Amy MacFadyen, Zhen-Gang Ji, and Robert H. Weisberg (Eds.)*
196 **Extreme Events and Natural Hazards: The Complexity Perspective** *A. Surjalal Sharma, Armin Bunde, Vijay P. Dimri, and Daniel N. Baker (Eds.)*
197 **Auroral Phenomenology and Magnetospheric Processes: Earth and Other Planets** *Andreas Keiling, Eric Donovan, Fran Bagenal, and Tomas Karlsson (Eds.)*
198 **Climates, Landscapes, and Civilizations** *Liviu Giosan, Dorian Q. Fuller, Kathleen Nicoll, Rowan K. Flad, and Peter D. Clift (Eds.)*
199 **Dynamics of the Earth's Radiation Belts and Inner Magnetosphere** *Danny Summers, Ian R. Mann, Daniel N. Baker, and Michael Schulz (Eds.)*
200 **Lagrangian Modeling of the Atmosphere** *John Lin (Ed.)*
201 **Modeling the Ionosphere-Thermosphere** *Jospeh D. Huba, Robert W. Schunk, and George V. Khazanov (Eds.)*
202 **The Mediterranean Sea: Temporal Variability and Spatial Patterns** *Gian Luca Eusebi Borzelli, Miroslav Gacic, Piero Lionello, and Paola Malanotte-Rizzoli (Eds.)*
203 **Future Earth – Advancing Civic Understanding of the Anthropocene** *Diana Dalbotten, Gillian Roehrig, and Patrick Hamilton (Eds.)*
204 **The Galápagos: A Natural Laboratory for the Earth Sciences** *Karen S. Harpp, Eric Mittelstaedt, Noémi d'Ozouville, and David W. Graham (Eds.)*
205 **Modeling Atmospheric and Oceanic Flows: Insightsfrom Laboratory Experiments and Numerical Simulations** *Thomas von Larcher and Paul D. Williams (Eds.)*
206 **Remote Sensing of the Terrestrial Water Cycle** *Venkat Lakshmi (Ed.)*
207 **Magnetotails in the Solar System** *Andreas Keiling, Caitríona Jackman, and Peter Delamere (Eds.)*
208 **Hawaiian Volcanoes: From Source to Surface** Rebecca *Carey, Valerie Cayol, Michael Poland, and Dominique Weis (Eds.)*
209 **Sea Ice: Physics, Mechanics, and Remote Sensing** *Mohammed Shokr and Nirmal Sinha (Eds.)*
210 **Fluid Dynamics in Complex Fractured-Porous Systems** *Boris Faybishenko, Sally M. Benson, and John E. Gale (Eds.)*
211 **Subduction Dynamics: From Mantle Flow to Mega Disasters** *Gabriele Morra, David A. Yuen, Scott King, Sang Mook Lee, and Seth Stein (Eds.)*
212 **The Early Earth: Accretion and Differentiation** *James Badro and Michael Walter (Eds.)*
213 **Global Vegetation Dynamics: Concepts and Applications in the MC1 Model** *Dominique Bachelet and David Turner (Eds.)*
214 **Extreme Events: Observations, Modeling and Economics** *Mario Chavez, Michael Ghil, and Jaime Urrutia-Fucugauchi (Eds.)*
215 **Auroral Dynamics and Space Weather** *Yongliang Zhang and Larry Paxton (Eds.)*
216 **Low-Frequency Waves in Space Plasmas** *Andreas Keiling, Dong-Hun Lee, and Valery Nakariakov (Eds.)*
217 **Deep Earth: Physics and Chemistry of the Lower Mantle and Core** *Hidenori Terasaki and Rebecca A. Fischer (Eds.)*
218 **Integrated Imaging of the Earth: Theory and Applications** *Max Moorkamp, Peter G. Lelievre, Niklas Linde, and Amir Khan (Eds.)*
219 **Plate Boundaries and Natural Hazards** *Joao Duarte and Wouter Schellart (Eds.)*
220 **Ionospheric Space Weather: Longitude and Hemispheric Dependences and Lower Atmosphere Forcing Timothy Fuller-Rowell**, *Endawoke Yizengaw, Patricia H. Doherty, and Sunanda Basu (Eds.)*
221 **Terrestrial Water Cycle and Climate Change Natural and Human-Induced Impacts** *Qiuhong Tang and Taikan Oki (Eds.)*
222 **Magnetosphere-Ionosphere Coupling in the Solar System** *Charles R. Chappell, Robert W. Schunk, Peter M. Banks, James L. Burch, and Richard M. Thorne (Eds.)*
223 **Natural Hazard Uncertainty Assessment: Modeling and Decision Support** *Karin Riley, Peter Webley, and Matthew Thompson (Eds.)*
224 **Hydrodynamics of Time-Periodic Groundwater Flow: Diffusion Waves in Porous Media** *Joe S. Depner and Todd C. Rasmussen (Auth.)*
225 **Active Global Seismology** *Ibrahim Cemen and Yucel Yilmaz (Eds.)*
226 **Climate Extremes** *Simon Wang (Ed.)*
227 **Fault Zone Dynamic Processes** *Marion Thomas (Ed.)*
228 **Flood Damage Survey and Assessment: New Insights from Research and Practice** *Daniela Molinari, Scira Menoni, and Francesco Ballio (Eds.)*
229 **Water-Energy-Food Nexus – Principles and Practices** *P. Abdul Salam, Sangam Shrestha, Vishnu Prasad Pandey, and Anil K Anal (Eds.)*
230 **Dawn–Dusk Asymmetries in Planetary Plasma Environments** *Stein Haaland, Andrei Rounov, and Colin Forsyth (Eds.)*
231 **Bioenergy and Land Use Change** *Zhangcai Qin, Umakant Mishra, and Astley Hastings (Eds.)*
232 **Microstructural Geochronology: Planetary Records Down to Atom Scale** *Desmond Moser, Fernando Corfu, James Darling, Steven Reddy, and Kimberly Tait (Eds.)*
233 **Global Flood Hazard: Applications in Modeling, Mapping and Forecasting** *Guy Schumann, Paul D. Bates, Giuseppe T. Aronica, and Heiko Apel (Eds.)*
234 **Pre-Earthquake Processes: A Multidisciplinary Approach to Earthquake Prediction Studies** *Dimitar Ouzounov, Sergey Pulinets, Katsumi Hattori, and Patrick Taylor (Eds.)*
235 **Electric Currents in Geospace and Beyond** *Andreas Keiling, Octav Marghitu, and Michael Wheatland (Eds.)*
236 **Quantifying Uncertainty in Subsurface Systems** *Céline Scheidt, Lewis Li and Jef Caers (Eds.)*
237 **Petroleum Engineering** *Moshood Sanni (Ed.)*
238 **Geological Carbon Storage: Subsurface Seals and Caprock Integrity** *Stéphanie Vialle, Jonathan Ajo-Franklin, and J. William Carey (Eds.)*
239 **Lithospheric Discontinuities** *Huaiyu Yuan and Barbara Romanowicz (Eds.)*
240 **Chemostratigraphy Across Major Chronological Eras** *Alcides N.Sial, Claudio Gaucher, Muthuvairavasamy Ramkumar, and Valderez Pinto Ferreira*

Geophysical Monograph 241

Mathematical Geoenergy
Discovery, Depletion, and Renewal

Paul Pukite
Dennis Coyne
Daniel Challou

This Work is a co-publication of the American Geophysical Union and John Wiley and Sons, Inc.

AGU 100
ADVANCING EARTH AND SPACE SCIENCE

WILEY

This Work is a co-publication between the American Geophysical Union and John Wiley & Sons, Inc.

This edition first published 2019 by John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA and the American Geophysical Union, 2000 Florida Avenue, N.W., Washington, D.C. 20009

Published under the aegis of the AGU Publications Committee

Brooks Hanson, Executive Vice President, Science
Lisa Tauxe, Chair, Publications Committee
For details about the American Geophysical Union visit us at www.agu.org.

Wiley Global Headquarters
111 River Street, Hoboken, NJ 07030, USA
For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Library of Congress Cataloging-in-Publication data is available

ISBN: 9781119434290

Cover image: Courtesy of Kern County Museum. Used with permission. © pedrosala/Shutterstock
Cover design by Wiley

Set in 10/12pt Times New Roman by SPi Global, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

CONTENTS

PREFACE

This book describes the mathematics and analytical tools behind analyzing the Earth's energy sources, in what we refer to as our geoenergy resources. A significant proportion of the Sun's energy is ultimately processed by the atmosphere, oceans, lakes, biological life (into fossil fuels), and land before being potentially used as energy resources. It was originally motivated by a shared interest in our global fossil fuel transition (Smalley, 2005) and in simplifying the models that we can use for engineering and scientific analysis. The adage that comes to mind is that "A complex system that works is invariably found to have evolved from a simple system that worked."

Because of that objective, many of the topics covered have the common theme that either the research is lacking in applying a mathematical approach (where instead heuristics are often used) or that there was significant potential for simplification in a specific domain. We have intentionally limited the scope to math and statistics that does not require enormous computational resources, in what is often referred to as first-order applied physics modeling. In that sense, the text is suitable for interdisciplinary applications where concise modeling approaches are favored.

The mathematics covers both deterministic and stochastic processes. As for the latter, several authors have tried to rationalize the utility of probability and statistics in larger contexts, which we have used for motivation:

1. *Dawning of the Age of Stochasticity and Pattern Theory*, David Mumford (Mumford, 2000; Mumford & Desolneux, 2010)
Mumford wrote a position paper on the prospects of using probability to solve problems in the future. From the introduction: "From its shady beginnings devising gambling strategies and counting corpses in medieval London, probability theory and statistical inference now emerge as better foundations for scientific models, especially those of the process of thinking and as essential ingredients of theoretical mathematics, even the foundations of mathematics itself." His book on pattern theory motivates the approach for finding patterns in real-world data, and in finding self-similarity among disparate natural phenomena (such as with fractals as described by Mandelbrot).

2. *Probability Theory: The Logic of Science*, Edwin T. Jaynes (Jaynes & Bretthorst, 2003)
Jaynes almost finished his treatise on probability as a unifying field, with his Maximum Entropy principle providing a recurring pattern of statistical similarity in many natural phenomena. From the body: "Our theme is simply: probability theory as extended logic. The 'new' perception amounts to the recognition that the mathematical rules of probability theory are not merely rules for calculating frequencies of 'random variables'; they are also the unique consistent rules for conducting inference (i.e. plausible reasoning) of any kind and we shall apply them in full generality to that end."

3. *On Thinking Probabilistically*, M.E. McIntyre (2007)
A white paper that provides a compatible view to Jaynes and Cox.

4. *The Black Swan and Fooled by Randomness*, N.N. Taleb (2010, 2005)
Popular books on probability in everyday life.

5. *Critical Phenomena in Natural Sciences*, Didier Sornette (2004)
The mathematical physics behind what Taleb discusses.

6. *Looking for New Problems to Solve? Consider the Climate*, Brad Marston (2011)
A suggestion to physicists that there are intellectual challenges in models for climate science.

The scope of the book is partitioned into two sections corresponding to each half of our energy transition, demarcated by the halfway point of "peak oil".

The first Part "Depletion" discusses aspects of oil depletion and fossil fuel energy availability where we try to go beyond the heuristics of classical projections and use more formal stochastic mathematical approaches.

The second Part "Renewal" discusses renewable energy and how we can harness our geophysical environment by finding patterns in available data derived from measured energy sources.

As a guideline, we tried to keep in mind that the utility and acceptance of a model depends as much on its plausibility and parsimony as its quality of fit or precision. Ultimately, the models presented here need to be evaluated with respect to other models of varying degrees of complexity. And also to remember that models are only as good as the data fed into the model (which in the case of the oil industry is often closely guarded either by corporations or by nation-states). Yet, even given poor data, part of the rationale of this book is providing approaches to deal with missing or uncertain information, where the models can help to interpolate or extrapolate and thus fill out that data.

An outgrowth of this work is that we will maintain an interactive web site GeoEnergyMath.com where models and mathematical formulations described herein will be organized for convenient access and other links to gray-literature and public data will be made available. As much of the data pertaining to energy usage is immediately obsolete once made publically available, it is important to provide continual updates to what is provided within this text. This is similar to what happens with weather forecasting (both historical information and updated forecasts are provided on a continual basis). As our goal is to provide an understanding of natural phenomena, the focus on actual forecasts within this book will be intentionally limited and readers will be encouraged to visit the web site for up-to-date data analyses. Further, as the data is often poor in quality or limited in extent, this will provide a means to validate or even invalidate the models over time. Since the earth sciences are primarily an observational and empirical discipline, and that controlled experiments are not often possible, it is largely an exercise in mathematically creative interpretations of the available data that enables progress.

Part of this work was originally funded through the Department of Interior as part of a DARPA-managed project, and as part of the contractual agreement, all the research work was approved for public release with unlimited distribution granted. Also, we would like to thank Sandra Pukite, Emil Moffa, and Jean Laherrere for detailed reviews, and Samuel Foucher for early collaborative research. In memoriam, we appreciate the valuable help and insight that Kevin O'Neill and Keith Pickering provided during this project but will not be able to share with. And thanks to the DJ.

Paul Pukite

REFERENCES

Jaynes, E. T., & Bretthorst, G. L. (2003). *Probability theory: The logic of science*. Cambridge: Cambridge University Press.

Marston, B. (2011). Looking for new problems to solve? Consider the climate. *Physics*, *4*, 20.

McIntyre, M. E. (2007). *On thinking probabilistically*. Paper presented at the Extreme Events: Proceedings of 15th 'Aha Huliko'a Workshop, 23–26 January 2007, Honolulu, HI, pp. 153–161. (http://www.soest.hawaii.edu/PubServices/AhaHulikoa.html; http://www.soest.hawaii.edu/PubServices/Aha_2007_final.pdf).

Mumford, D. (2000). The dawning of the age of stochasticity. In V. Arnold, M. Atiyah, P. Lax, & B. Mazur (Eds.), *Mathematics: Frontiers and perspectives* (pp. 197–218). Providence, RI: American Mathematical Society.

Mumford, D., & Desolneux, A. (2010). *Pattern theory: The stochastic analysis of real-world signals*. Natick, MA: A K Peters/CRC Press.

Smalley, R. E. (2005). Future global energy prosperity: The terawatt challenge. *Materials Research Society Bulletin*, *30*(6), 412–417.

Sornette, D. (2004). *Critical phenomena in natural sciences: chaos, fractals, selforganization, and disorder: Concepts and tools*. Berlin: Springer Verlag.

Taleb, N. N. (2005). *Fooled by randomness: The hidden role of chance in life and in the markets*. New York: Random House Inc.

Taleb, N. N. (2010). *The black swan: The impact of the highly improbable*. New York: Random House Inc.

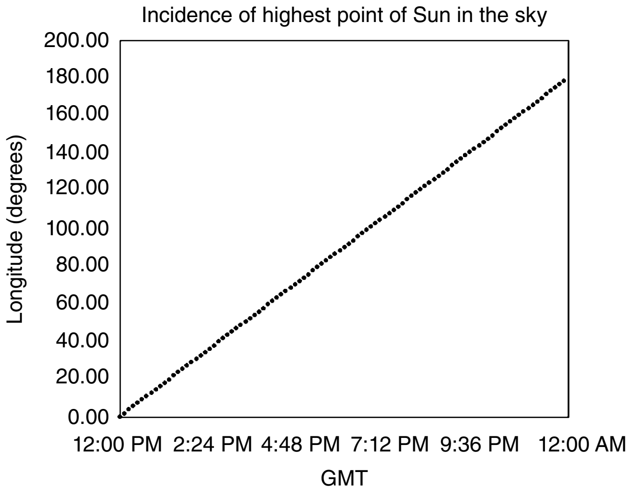
1
Introduction to Mathematical Geoenergy

ABSTRACT

In this introductory chapter, we relate the geophysics of the Earth and its atmosphere and of the influences of the sun and the moon and cast that into a geoenergy analysis. Geoenergy is energy derived from geological and geophysical processes and categorized according to its originating source. The sources are compartmentalized according to whether they are radiation-based (such as from sunlight via the photo-electric effect), gravitational (such as from the moon or terrain), geothermal (such as from volcanic sources), kinetic (from the rotation of the Earth and Coriolis forces), or chemical/nuclear (such as from fossil fuel and ion-based batteries). We use these models to project fossil fuel production and provide analysis tools for renewable technologies.

Our objective is to apply what we know about the geophysics of the Earth and its atmosphere and of the influences of the Sun and the Moon and cast that into a geoenergy analysis. As we define it, geoenergy is energy derived from geological and geophysical processes and categorized according to its originating source. Perhaps most convenient is to compartmentalize the sources according to whether they are radiation based (such as from sunlight via the photoelectric effect), gravitational (such as from the Moon or terrain), geothermal (such as from volcanic sources), kinetic (from the rotation of the Earth and Coriolis forces), or chemical/nuclear (such as from fossil fuel and ion-based batteries).

As the acquisition and use of energy is in essence an active process, geoenergy analysis becomes (1) a study of differentiating between *deterministic* and *stochastic* processes and (2) of applying *physics* or *heuristics* to come up with adequate models to aid in understanding and to perhaps improve the efficient use of our resources either statistically or with confidence based on sound physical models.



It really is not difficult to understand the first distinction (1), as the Sun rising in the morning and falling in the evening is an example of a deterministic process, while predicting cloud cover during the day is a stochastic process. This of course has impacts for predicting efficiencies

Mathematical Geoenergy: Discovery, Depletion, and Renewal, Geophysical Monograph 241, First Edition.
Paul Pukite, Dennis Coyne, and Daniel Challou.

in solar energy collection, as we know exactly when the Sun will be at its zenith in any geographic location (a deterministic process; see figure), yet we do not know if there will be significant cloud cover at any specific time (a stochastic process).

The second distinction (2), between physics and heuristics, is based on how well we scientifically understand a phenomenon. This becomes apparent when one realizes that many estimates of remaining fossil fuel reserves are heuristics (i.e., educated guesses), based many times only on historical trends. In neglecting a mathematical physics treatment, however, we unfortunately remain uncertain on projections as we cannot account for how the heuristic may fail. In general, we will have more confidence in a scientifically based physics model.

These distinctions can be combined to create four different basic categories.

	Stochastic	Deterministic
Physics	Weather	Tides
Heuristics	Hubbert curve	Sunspots

For example, stochastic physics would be represented by a detailed weather model which would include differential equations describing atmospheric flow and solved on a supercomputer. Different outcomes based on varying initial conditions would generate a statistical spread to be used in regional weather forecasting.

Stochastic heuristics typically apply to a situation that may be too complicated or detailed in scope, resulting in a model that may simply estimate a mean value and possibly a variance for some quantity. This would include our current best guess at predicting future oil production, which has typically applied the so-called Hubbert curve. But this may not be the best possible guess and explains why we have better and more physically oriented models as we will further detail.

On the deterministic side, a good example of a physics application is the theory of tides and tidal analysis. These have high precision and are routinely used for predicting tides down to the minute.

On the other hand, a deterministic heuristic is rare to come across. It is a behavior that appears very predictable yet one for which we lack a good physical model. For example, countering the easily predictable sunset and sunrise, which we physically understand, we have only a partial understanding with respect to solar sunspots. Sunspots appear to have an 11-year cycle, making them somewhat deterministic, yet we do not fully understand the mechanism. Thus, a heuristic is applied to the sunspot cycle describing an 11-year cycle.

1.1. NONRENEWABLE GEOENERGY

The comprehensive framework we will describe has aspects of probability-based forecasting (Limited by the psychology of collective human actions). The salient reason for using probabilistic-based models results from reasoning in the face of uncertainty. We never have had and probably never will have perfect and complete data to accurately analyze, much less predict, our current situation. Lacking this, imperfect probabilistic approaches serve us very well in our understanding of the fundamentals of oil depletion.

Concerning oil (defined as crude plus condensate) depletion, we know that three things will happen in sequence:

1. Oil output will peak.
2. Oil output will decline.
3. Extraction and use of oil will become counterproductive in terms of energy efficiency and the impact on the environment. This will occur for all sources of oil (such as shale oil, extra heavy oil, etc.).

The dates of these events remain unknown, but we have historical data and stochastic models to help guide us in understanding future energy resource availability.

1.2. RENEWABLE GEOENERGY

To understand how to harness renewable geoenergy, we need to model natural phenomena so that it becomes more predictable. In other contexts, we do that already. For example, for ocean tides, we create tidal tables that allow us to plan typical coastal activities. If we can do the same with related geophysical and climate phenomena, the benefits would be enormous.

We start with knowledge of the external energy sources, focusing on solar and gravitational, and find patterns that allow us to model these natural phenomena as both deterministic and stochastic processes. As of today, not any single one of these processes can take the place of fossil fuels in terms of efficiency, but taken together they may make a dent.

To that end, the scope of the analysis will include models of wind, climate cycles, solar energy conversion, battery technology, etc. The main idea in creating such models is that renewable energy is closely linked to efficiency, and the more we can wring out of these sources, the less the impact we will see during our energy transformation away from nonrenewable fuels to a renewable paradigm.

So, the main themes are to create deterministic and stochastic models of natural phenomena according to gathered empirical data using physics and heuristics where appropriate. The emphasis on mathematical physics is stressed because that has the potential for further insight. In several cases, we will show how machine learning models have uncovered patterns in the data leading directly to the applied physics mathematical models.

Models of the physical environment play an important role in supporting planning, analysis, and engineering. Fundamental principles of thermodynamics and statistical physics can be applied to create compact parameterized models capable of statistically capturing the patterns exhibited in a wide range of environmental contexts. Such models will allow more efficient and systematic assessment of the strengths and weaknesses of potential approaches to harnessing energy or efficiently working with the environment. Further, the models play an important role in computer simulations which can produce better designs of complex systems more quickly, affordably, and reliably.

In terms of renewable energy, models of the weather and climate are vital for planning, optimizing, and taking advantage of energy resources. Every aspect of the climate is important. For example, knowing the long-term climate forecast for the occurrence of El Niños will allow us to plan for hotter than average temperature extremes in certain parts of the world or to plan for droughts or floods. These climate behaviors are examples of geophysical fluid dynamics models (Vallis, 2016) where the distinction between stochastic and deterministic (and deterministically chaotic) causes is under intense research (Caprara & Vulpiani, 2016), and we will describe how we may be able to simplify the models.

From a computational perspective, there has been a steady increase of the use of machine learning to identify deterministic patterns (Jones, 2017; Karpatne & Kumar, 2017; Steinbach et al., 2002). For example, the quasi-biennial oscillation (QBO) behavior of stratospheric winds has long been speculated to be forced by the cyclic lunar tidal potential. A matching lunar pattern was discovered via a symbolic regression machine learning experiment and then verified by aliasing a strong seasonal (yearly) signal onto an empirical model of the lunar tidal potential (Pukite, 2016). We can expect more of these kinds of discoveries in the future, but appropriate mathematical and statistical physics will help guide this path.

REFERENCES

Caprara, S., & Vulpiani, A. (2016). Chaos and stochastic models in physics: Ontic and epistemic aspects. In E. Ippoliti, F. Sterpetti, & T. Nickles (Eds.), *Models and Inferences in Science* (pp. 133–146). Switzerland: Springer.

Jones, N. (2017). How machine learning could help to improve climate forecasts. *Nature*, *548*, 379–380.

Karpatne, A., & Kumar, V. (2017). *Big data in climate: Opportunities and challenges for machine learning*. Presented at the Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, Canada, pp. 21–22.

Pukite, P. (2016). *Analytical formulation of equatorial standing wave phenomena: Application to QBO and ENSO*. Presented at the AGU Fall Meeting Abstracts, San Francisco, CA.

Steinbach, M., Tan, P.-N., Kumar, V., Potter, C., Klooster, S., & Torregrosa, A. (2002). *Data mining for the discovery of ocean climate indices*. Presented at the Proceedings of the Fifth Workshop on Scientific Data Mining at 2nd SIAM International Conference on Data Mining, Arlington, VA.

Vallis, G. K. (2016). Geophysical fluid dynamics: Whence, whither and why? *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, *472*, 20160140.

2
Stochastic Modeling

ABSTRACT

We often see unrelated phenomenon that shows rather similar characteristics. In fact, the behaviors observed often have a common mathematical origin based on the properties of a *population* of observations. The effects of disorder and specifically that of entropy require us to use notions of probabilities to understand them. In this chapter, we provide some of the intuitive background to help guide us through a stochastic analysis.

The mathematics and probability and statistics behind stochastic models

We often see unrelated phenomenon that shows rather similar characteristics. In fact, the behaviors observed often have a common mathematical origin based on the properties of a *population* of observations. The effects of disorder and specifically that of entropy require us to use notions of probabilities to understand them. In this chapter, we provide some of the intuitive background to help guide us through a stochastic analysis.

2.1. ODDS AND UNCERTAINTY AND THE PRINCIPLE OF MAXIMUM ENTROPY

The scientist E.T. Jaynes was the originator of the principle of maximum entropy. Known best for relating entropy and probability to many areas of science and information technology, Jaynes provided an alternative Bayesian analytic framework to the classical statistics school, known as the *frequentists*.

The *probabilistic* school made great practical strides in solving many thorny physics problems, as Jaynes showed how ideas from probability could encompass some classical statistics ideas, going so far as to provocatively labeling probability theory as *the logic of science*. Similarly, the useful law known as Cox's theorem justified a *logical* interpretation of probability.

Jaynes described how the mathematician Laplace had worked out many of the fundamental probability ideas a couple of hundred years ago (Jaynes lived in the twentieth century and Laplace in the eighteenth century), yet became marginalized by a few (in retrospect) petty arguments. One of the infamous theories Laplace offered, the sunrise problem, has since supplied ammunition for critics of Bayesian ideas over the years. In this example, Laplace essentially placed into quantitative terms the probability that the Sun would rise tomorrow based on the count of how many times it had risen in the past. We can categorize this approach as Laplace's precursor of Bayes' rule, originally known as the rule of succession. In current parlance, we consider this a straightforward Bayesian (or Bayes-Laplace) update, a commonplace approach among empirical scientists and engineers who want to discern or predict trends.

Mathematical Geoenergy: Discovery, Depletion, and Renewal, Geophysical Monograph 241, First Edition.
Paul Pukite, Dennis Coyne, and Daniel Challou.

Yet, legions of mathematicians disparaged Laplace for years since his rule did not promote much certainty in the fact that the Sun would indeed rise tomorrow if we input numbers naively. Instead of resulting in a probability of unity (i.e., absolute certainty), Laplace's law could give numbers such as 0.99 or 0.999 depending on the number of preceding days included in the prior observations. Many scientists scoffed at this notion because it certainly did not follow any physical principle, yet Laplace had also placed a firm warning to use strong scientific evidence when appropriate. In many of his writings, Jaynes has defended Laplace by pointing out this caveat and decried the fact that no one heeded Laplace's advice. As a result, for many years hence, science had missed out on some very important ideas relating to representing uncertainty in data.

Jaynes along with the physicist R.D. Cox have had a significant impact in demonstrating how to apply probability arguments. This is important in a world filled with uncertainty and disorder. In some cases, such as in the world of statistical mechanics, one finds that predictable behavior can arise out of a largely disordered state space; Jaynes essentially reinterpreted statistical mechanics as an inferencing argument, basing it on incomplete information on the amount of order within a system.

In the oil depletion analysis covered in the first part of this book, we will see how effectively models of randomness play into the behavior. Missing pieces of data together with the lack of a good quantitative understanding motivate our attempts at arriving at some fundamental depletion models.

Jaynes spent much time understanding how to apply the maximum entropy principle (MaxEnt) to various problems. We applied the MaxEnt principle with regard to oil because much oil production and discovery numbers are not readily available. Unsurprisingly, that approach works quite effectively in other application areas as well and perhaps in many future situations. As Jaynes had suggested, the duality of its use for both statistics and statistical physics makes it a solid analysis approach:

> Any success that the theory has, makes it useful in an engineering sense, as an instrument for prediction. But any failures which we might find would be far more valuable to us, because they would disclose new laws of physics. You can't lose either way. — E.T. Jaynes

The oil industry has actually used the MaxEnt principle quite heavily over the years. Mobil Oil published one of the early classic Jaynes texts based on a symposium they funded under the banner of their research laboratory. Also during this era, academic geophysicists such as J.P. Burg had used Jaynes ideas to great effect. Burg essentially derived the approach known as maximum entropy spectral analysis. Not limited to geophysics, this technique for uncovering a signal buried in noise has become quite generally applied. The reliability researcher Myron Tribus pointed out this early success, demonstrating Burg's own personal victory whereby he applied his own algorithm at an abandoned oil field he christened *Rock Entropy #1*. The profits he made from the oil he extracted helped to fund his own research (Levine & Tribus, 1979).

So, given that the petroleum and geology fields contributed a significant early interest in the field of MaxEnt, we carried this approach forward with our depletion models. Jaynes has often pointed out that some of the applications work out so straightforwardly that an automaton, given only the fundamental probability rules, could figure out the solution to many of these problems:

> We're not going to ask the theory to predict everything a system could do. We're going to ask, is it possible that this theory might predict experimentally reproducible phenomena. — E.T. Jaynes

Jaynes has said that thinking about maximizing entropy parallels the idea that you place your bets on the situation that can happen in the greatest number of ways. Then because enough events and situations occur over the course of time, we end up with something that closely emulates what we observe:

> Entropy is the amount of uncertainty in a probability distribution. — E.T. Jaynes

This involves estimating the underlying probability distribution. This sounds hard to do, but the basic rules for maximizing entropy only assume the constraints; so, that includes things such as assuming the mean or the data interval:

> No matter how profound your mathematics is, if you hope to come out with a probability distribution, then some place you have to put in a probability distribution. — E.T. Jaynes

Given all that as motivation and noting how well MaxEnt works at estimating oil reservoir field sizes and other measures, we can see what other ideas shake out. We can start out with the context of oil reservoirs. So, based on a MaxEnt of the aggregation of oil reservoir sizes over time, we can foreshadow how we came up with the following cumulative probability distribution for field sizes:

$$P(\text{size}) = \frac{1}{1 + \dfrac{C}{\text{Size}}} \tag{2.1}$$

First, consider that most people have an intuitive understanding of gambling, especially in the form of sports betting, where a sports fan comprehends how the

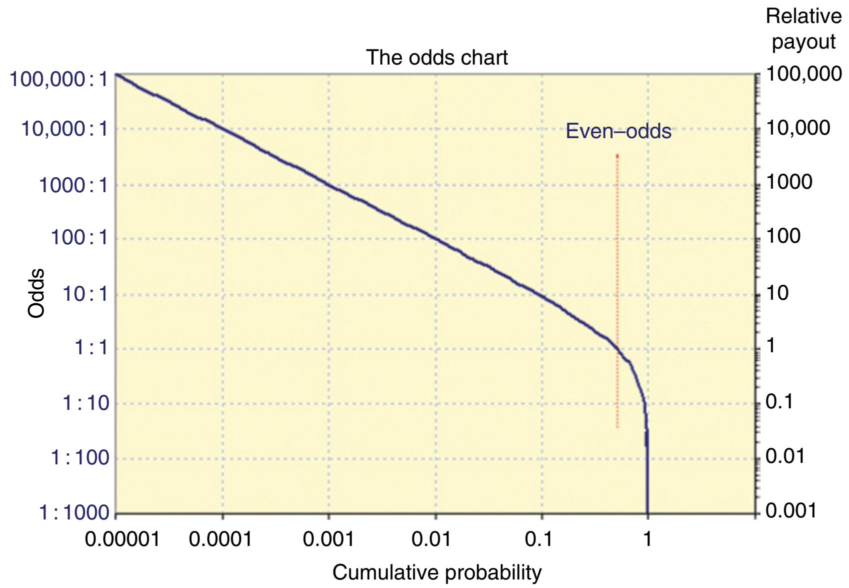


Figure 2.1 ☼ A plot of the cumulative probability for the odds function. This is transformed to a rank histogram if the *x* axis value of 1 is considered the max rank of a sampled population and the *y* axis consists of the sampled payouts. The sampled points will statistically lie along the line if the sample size is large enough.

odds function works. Odds against for some competitor to win is essentially cast in terms of the probability P:

$$\text{Odds} = \frac{1-P}{P} \tag{2.2}$$

So, in terms of odds, we can rearrange the first equation into the odds formulation by using either the definition of odds for, Odds $= P/(1-P)$, or odds against, Odds $= (1-P)/P$.

When plotted, the odds distribution appears like the curve in Figure 2.1.

When algebraically rearranged with the first equation, the odds of finding a reservoir larger than a certain size, assuming we randomly pick from the sample population, comes out to be

$$\text{Odds}(\text{size}) = \frac{C}{\text{Size}} \tag{2.3}$$

where C is a constant of proportionality. So, we can give the odds of discovering a size of a certain reservoir in comparison to the median characteristic value just by taking the ratio between the two values. This equates well to the relative payout of somebody who beats the odds and thus beats the house median.

This becomes even more obvious when we compare with Figure 2.2.

This simple result gives us some great insight. It essentially tells us that the greater the size of the reservoir desired, the progressively smaller the odds that we would come across at least that size. For the United States, the value of C comes out less than 1 million barrels (Mb), so finding a field of at least 10,000 Mb is 1 : 10,000. This assumes that we randomly draw from a sample of newly discovered fields.

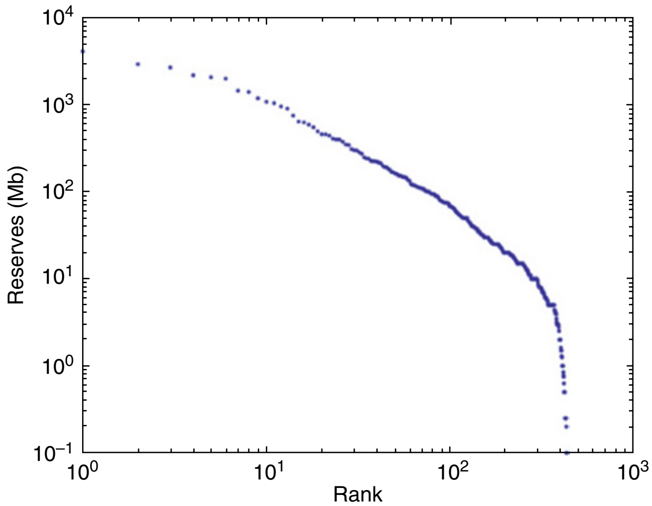


Figure 2.2 ☼ A plot of a cumulative rank histogram of a reservoir size distribution function. Note how closely it resembles the odds function of the previous figure. Higher odds against relate directly to the rare large size discovery. Other probability density functions such as (parabolic) fractal, lognormal, or stretched exponential may also fit this general profile [*Laherrère*, 2000]. The knee of the profile can change over time as many small fields, being at one time uneconomical, are now included as discoveries. Reproduced with permission of Elsevier.

On the other hand, if we want the odds of drawing from the sample and expecting at least a 1 Mb field, we put in the formula and obtain 1 : 1, or basically even odds. So, if we want to somehow maintain our current rate of domestic production by placing safe bets, we must find many small reservoirs.

We could also place our bets on the long payoff but need to realize that the probability size distribution starts to asymptotically limit for large sizes (due to a constrained limit L) and the odds factor blows up

$$P(\text{size}) = \frac{\text{Size}\cdot(L+C)}{(\text{Size}+C)\cdot L} \tag{2.4}$$

as the odds does this

$$\text{Odds}(\text{size}) = \frac{(\text{Size}+C)\cdot L}{\text{Size}\cdot(L+C)} - 1 \tag{2.5}$$

This gives similar odds for a small reservoir, still close to 1:1, but the odds for getting a large reservoir no longer scale. For example, if we use a max size L of 20,000 Mb, then the odds of a size of 10,000 Mb is one half the odds without the maximum size. And the odds for getting anything bigger than 20,000 Mb become essentially 1 in infinity.

This all comes about from assuming a maximum entropy distribution on the accumulation of the reservoirs and then applying a constraint on the time that these reservoirs accumulate. As Jaynes said, we can do much with incomplete information.

The same arguments apply to more elaborate discovery models which place fixed limits on the cumulative production based on similar incomplete information.

The geologist and oil analyst M. King Hubbert, who studied oil depletion, likely never applied any of Jaynes' principles, except perhaps at some deep intuitive level. But as Jaynes himself might have concluded, that would have worked out just as well since one intent of probability theory has always tried to attach quantitative terms to human insight, the so-called subjective probability approach. So, Hubbert gave us some of the insight in predicting future oil supplies, and the rest of the probability-based models provide the mathematical foundation.

2.2. DISPERSION

The odds function is a good starting point, as it has intuition behind it. As Jaynes would suggest, this has become part of our Bayesian conditioned belief system. But many other processes obey a similar *dispersive* effect.

One of the constraints is that the cumulative probability sums to one over the rank histogram, which is another intuitive aspect, in that probabilistically, every event must eventually be covered.

In many areas of applied mathematics, one can often find a purely analytic result solved strictly by equations of probability. Often one will find references to a Monte Carlo simulation. This actually results from an inversion of the analytical function, simply run through a random number variate generator. Many mathematicians do a Monte Carlo analysis to check their work and for generating statistical margins.

Finding *outlier* data in a simulation is important as this can reveal important *fat-tail* behaviors. Moreover, often these outliers do not show up in Monte Carlo runs unless the rest of the histogram gets sufficiently smoothed out by executing a large sample space.

Consider that the human mobility plot has an exceedingly simple rationalization given there is dispersion in human behaviors. We have the following derived equation that gives the probabilities of how far a sample population has moved in a certain time based on the dispersion principle:

$$P(x,t) = \frac{\beta}{\beta + \frac{x}{t}} \tag{2.6}$$

To simulate this behavior, we need to take a few straightforward steps. Firstly, we simply draw from a uniform random distribution for distance (x), and secondly, we draw another number for a random time span (t). Or we can do it from two inverted maximum entropy exponential draws (does not really matter to achieve the fat-tail statistics). You then divide the two to arrive at a random velocity, that is, $v = x/t$.

We need nothing simpler than this formula or formulation. The ranked histogram for the Monte Carlo simulation of 10,000,000 trials of independent draws appears like the data points shown in Figure 2.3 with the dispersion formula as the solid line.

The random draws converge to the derived maximum entropy dispersion derivation.

2.3. APPLICATION OF THE MAXIMUM ENTROPY IDEAS

These general techniques can be applied across many domains. The usual problem remains that different application domains use different terminology. For example, the topic of breakthrough analysis in contaminant dispersal has many similarities to carrier transport in semiconductors. So, the overriding dispersion analysis is a general concept, and one can apply the same technique in oil depletion by making the analogy to dispersion in human-aided discovery search rates. The fact that it also occurs for physical processes such as contaminant flow in groundwater, carrier transport in amorphous semiconductors, or heat dispersion should

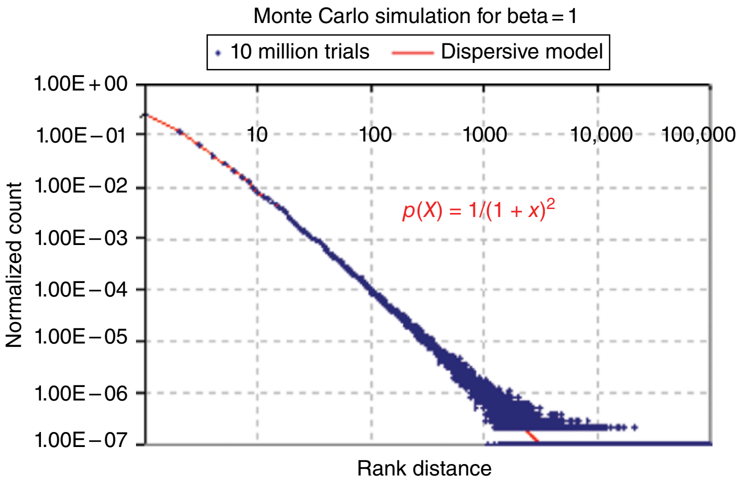


Figure 2.3 ☼ A Monte Carlo simulation of dispersion generates dispersion rates by, for example, deltaA/deltaB MaxEnt variates. The rarer events demonstrate noise from the counting statistics of a finite set of events. These exist at the low probability end of the scale.

not be surprising. These in fact all derive from some aspects of the field known as *statistical physics* or *statistical mechanics*.

It takes some intuition to determine the situations where disorder rules and where it does not. In much the same way that we can understand the dynamics of the Hubbert oil depletion curve via dispersion, so too can we understand the transient of an amorphous semiconductor time-of-flight experiment by applying dispersion. As a bottom line, we often can use fundamental concepts of disorder to understand the dynamics of these behaviors.

REFERENCES

Laherrère, J. (2000). Distribution of field sizes in a petroleum system: Parabolic fractal, lognormal or stretched exponential? *Marine and Petroleum Geology*, *4*, 17.

Levine, R. D., & Tribus, M. (1979). *The maximum entropy formalism*. A conference held at the Massachusetts Institute of Technology on 2–4 May, 1978. Ed. by Raphael D. Levine, Myron Tribus. Mass-London the MIT P XII.

Part I: Depletion

3

Fossil Fuel Depletion Modeling

ABSTRACT

We have a relatively poor understanding for oil production trends, with much of the analysis based on heuristics and empirical data. The root of the problem resides in a simple observation: most forecasts for oil production derive from predictions for demand. In other words, industry analysts often use demand as a projection yardstick. Where we encounter elastic supplies, an increase in demand is usually met with an increase in supply. And so we see the typical expectation level of an increasing demand met by an increasing supply of oil. In this chapter, we introduce how to create simple models for fossil fuel depletion where increasing demand is limited by a finite supply of oil.

Background on the stochastic mechanisms describing our consumption of oil

We have a relatively poor understanding for oil production trends, with much of the analysis based on heuristics and empirical data. The root of the problem resides in a simple observation: most forecasts for oil production derive from predictions for demand. In other words, industry analysts often use demand as a projection yardstick. Where we encounter elastic supplies, an increase in demand usually gets met with an increase in supply. And so we see the typical expectation level of an increasing demand met by an increasing supply of oil.

However, this does not hold for the case of a finite resource.

Engineers such as the geologist M. King Hubbert have sought to explain the finite nature of the resource. However, Hubbert's explanations universally lacked a real quantitative flavor, and he ended up guiding much of the work via intuition and the use of heuristics. Laherrère and Deffeyes have also done much work, essentially picking up where Hubbert left off.

These are the set of premises that we apply:

1. Accelerating growth in technology and human consumption. The growth in consumption in the face of a finite supply eventually leads to diminishing returns of supplies. We can explain this in terms of both micro and macro effects.
2. We sweep the volume of the Earth's crust to explain past and future oil discoveries and the possibility of reserve growth. We can understand the problem by incorporating the concept of dispersion, which amounts to varying the rates of search with time and geographical region.
3. Production dynamics and the effects of perturbations. We present the shock model as an intuitive way of analyzing the situation.

If we want to understand the problem, we need to bridge the separation between the mathematical abstraction of economic/human flows and that of geology. Ideally, we could discuss the flow of an arbitrary material, and the math concepts would remain nearly the same as modeling oil extraction.

Mathematical Geoenergy: Discovery, Depletion, and Renewal, Geophysical Monograph 241, First Edition.
Paul Pukite, Dennis Coyne, and Daniel Challou.

To derive the entire life cycle of oil, we break it down into three components that can be handled individually: *growth*, *discovery*, and *extraction*.

The first component drives the whole process. Intuitively, we need a rate function that describes how fast technology and consumption pressure stimulate the search and extraction process (and thus potentially reduce the cost). This rate can either accelerate in cases of restraint-free growth or perhaps decelerate if we hit the law of diminishing returns.

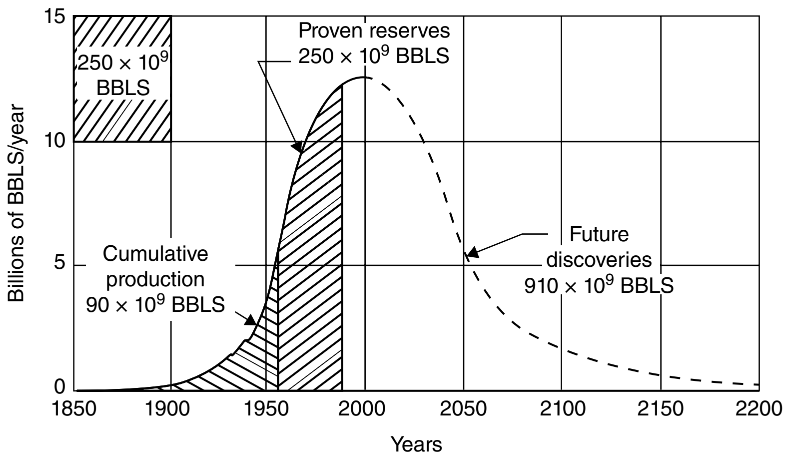
The second component describes the search for undiscovered resources. The basic premise describes a hard constraint: that we have a finite search space to deal with. There are approximations that we can make given that the volume is not precisely known, but this sets a target and foundation to work from.

The last of the three components describes the extraction process. Extraction can only happen after a discovery occurs. We separate extraction sequentially from the discovery process and so incorporate it as an independent process. The implication is that that we can create models of these processes independently and then combine them later.

Although the components of growth, discovery, and extraction remain largely independent (mathematicians refer to these as *orthogonal* components), we suggest that the role of basic human need for energy provides a consistent stimulus, thereby projecting the model forward from the basic premise. The human need for energy creates an unrelenting consumer demand which provides the impetus for growth. So, the outputs of extraction (i.e., cheap energy) form a *forcing function* for a supply and demand model.

For modeling, it is important to establish that demand remains consistent over time. So, even though conservation and efficient uses of oil has improved since the oil shock of the 1970s and 1980s, the fact that inflation-adjusted prices have remained relatively constant implies that demand of the same order of magnitude exists (For the 1930–1970 period, inflation-adjusted prices were relatively constant; from 1984 to 2002, this was also the case. The excursions from these are in part due to oil shocks due to wars, embargos, or shortages.). And a cheap supply spurring technology growth has in turn consistently promoted demand. For the United States, oil still does not cost us much to use in practical terms. We still see the same price-based consumer buying decisions that we always have, which means that we can draw a straight line from the early days of oil consumption until now and use essentially similar math. Therefore, demand becomes an *invariant* as described in mathematical terminology. The lack of a universal paradigm shift in fossil fuel consumption practices allows us to apply perturbations to the analysis to discover how scenarios such as current and future oil price shocks might play out.

This set of premises may exist as tacit knowledge, but we spell them out to establish our working model. By pulling together a model describing the entire life cycle, we hope to codify and make explicit this perhaps implicit information that may not be widely known. As an oil industry optimist had to say, “Here it is pertinent to note that peak oil forecasters do not enjoy an undiluted view of the state or corporate portfolios that contain these internal and hidden assessments which their models logically require” (Clarke, 2007).



3.1. PEAK OIL

The conventional wisdom holds that an oil production peak looks symmetric on the way up as well as on the way down. This leads to the familiar *bell-shaped* curve of peak oil. Shell Oil geologist M. King Hubbert in the late 1950s sketched a largely or nearly symmetric bell curve to reinforce the idea that maximum oil production would naturally and inevitably occur. Hubbert later presented this information (see figure inset) at a US congressional hearing (Hubbert, 1974a, 1974b).

By inspection, one can see that the 50% consumption point obviously occurs at the peak for purely symmetric profiles such as the logistic curve and Gaussian normal. Our probabilistic model will make no such assumption.

It may turn out that under certain circumstances, we can obtain the idealized symmetric curve. In the subsequent chapters, as we introduce the basic concepts for analyzing oil depletion dynamics, the presence or absence of symmetry will become apparent and thus better understood.

Figure 3.1 lays out a flowchart of the current understanding of the different phases of the oil production life cycle. The heading row provides short names for the life cycle phases. The top row lists some of the conventional practices used to describe the phases, primarily as a set of heuristics. Below this row appear the model interconnections, which establish the architectural foundations of the comprehensive model. Distinct stages which traverse the conventional phases of the life cycle draw from elements of probability theory which we use to model the behavior of the phases. Several surprising outcomes derive from the application of a probabilistic model. For example, we can derive the well-known logistic heuristic and explain the field size distributions observed. Further, we can use the dynamic elements to track shocks in the production process and extrapolate into the future.

The two darkened bubbles at the baseline contain the essential probability ideas that we use for aspects of the discovery process along with the oil production model. The combined model replaces a long-standing set of heuristics that in the past have been used in analyses of the aggregated oil production life cycle.

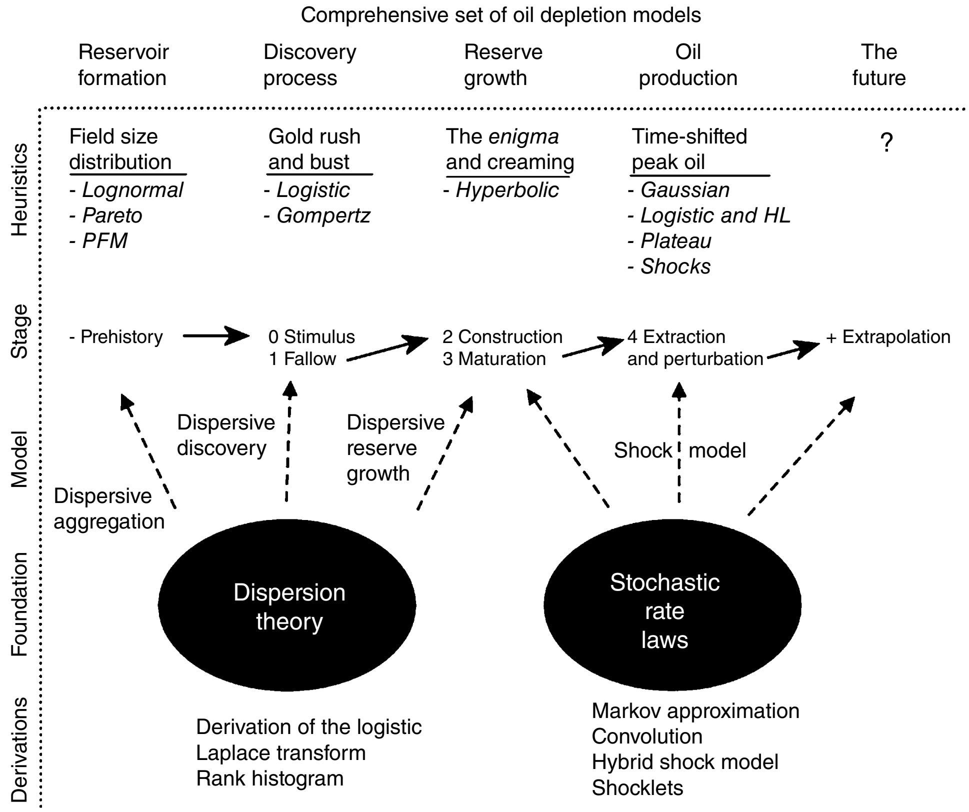


Figure 3.1 ☼ The roadmap for the analysis includes reuse of simple laws.

REFERENCES

Clarke, D. (2007). *The battle for barrels: Peak oil myths & world oil futures*. (pp. 130). London: Profile Books.

Hubbert, M. K. (1974a). On the nature of growth. *Testimony Hear. Natl. Energy Conserv. Policy Act 1974*.

Hubbert, M. K. (1974b). *US energy resources: A review as of 1972*. A background Paper Prepared at the Request of Henry M. Jackson, Chairman, Committee on Interior and Insular Affairs, United States Senate pursuant to S. Res. 45, A National Fuels and Energy Policy Study.

4
Discovering Oil Reserves

ABSTRACT

This chapter considers first the geological process that leads to oil formations. A good model of this will effectively describe the randomness of oil sources, both in size and location. A fraction of oil that gets created eventually becomes trapped in natural occurring reservoirs that have laid dormant for many years. The places that oil can congregate and are readily accessible occur sporadically, which makes it amenable to models based on stochastic considerations of diffusive and dispersive aggregation. These stochastic models of reservoir sizes are combined with dispersive discovery dynamics and compared to available data.

Where do we find oil reservoirs and how do we discover oil?

Let us consider first the geological process that leads to oil formations. A good model of this will effectively describe the randomness of oil sources, both in size and in location. A fraction of oil that gets created eventually becomes trapped in natural occurring reservoirs that have laid dormant for many years. The locations and sizes of individual reservoirs probably have some form of pattern but for the most part have huge elements of randomness to it. The places that oil can congregate and are readily accessible occur sporadically. Structurally, the Earth must provide trapping layers; otherwise, the oil becomes too dispersed within the Earth's crust. In this case, we end up with oil shale or oil sands which contain a suspension of oil that becomes much more difficult to extract. Ideally, the *best* traps occur in structural layers that may lie along fault lines (similar to those that can cause earthquakes). See Figure 4.1.

In terms of a timeline, a portion of the oil that initially gets formed in huge beds of dead biological material subsequently migrates from a dispersed state through porous rock until it ultimately reaches these semipermanent traps lined by impermeable rock, such as a salt dome. In John McPhee's book *Basin and Range*, he described the effect of geological forces that can move continents over the course of millions of years:

> Oil also moves after it forms. You never find it where God put it. It moves great distances through permeable rock. Unless something traps it, it will move on upward until it reaches daylight and turns into tar. You don't run a limousine service on tar, let alone a military-industrial complex. If, however, the oil moves upward through inclined sandstone and then hits a wall of salt, it stops, and stays — trapped. (McPhee, 1982)

As fault lines and similar structural anomalies (such as with Saudi Arabia's Ghawar field) and strata-related defects do not occur uniformly, we will not necessarily find oil wherever we decide to look. Adding to that the fact that oil itself did not form in every geographic region, we are left with a sporadic set of needles in a haystack to look through. So, instead of concentrating on the exact physical mechanisms, we can treat it as a probability and statistics problem.

One can analogize the distribution of these structural traps with the number of defects in so-called perfect crystals, such as a gem-quality diamond. Although occurring

Mathematical Geoenergy: Discovery, Depletion, and Renewal, Geophysical Monograph 241, First Edition.
Paul Pukite, Dennis Coyne, and Daniel Challou.

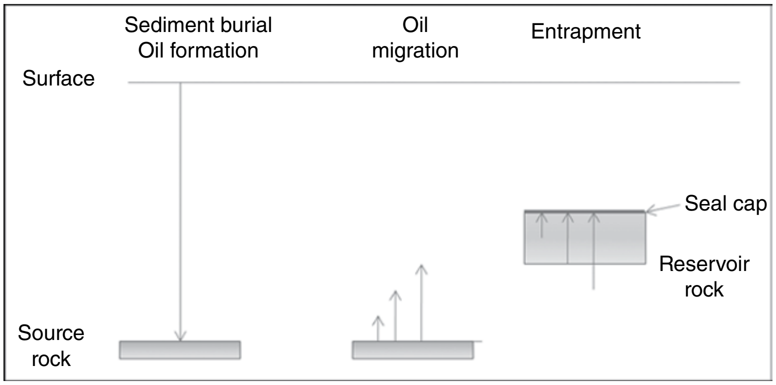


Figure 4.1 ☼ The effect of structural formations with natural caps tends to trap migrating oil into reservoirs. Not shown is the cooking of the biomass needed to form the kerogen and then oil.

on a microscopic scale, the crystalline faults share much in common with their macroscopic cousins. Although it is difficult to find the defects through a microscopic scan, we have indirect means to quantitatively characterize their density. This becomes the problem that oil prospectors face: that of adequately characterizing the number of oil-bearing faults that the Earth's crust contains. However, unlike the semiconductor industry, we want to seek out and maximize the number of these defects. Only by finding more structural anomalies do we have hope in finding more oil.

The other aspect one must deal with is the sizing of individual reservoirs. Through years of discovering and then estimating reservoir sizes, analysts have empirically guessed at various probability distributions for reservoir sizes. In general, we unsurprisingly reach the conclusion that smaller reservoirs occur much more frequently than larger reservoirs. And the largest, the so-called supergiants, occur very rarely.

With the empirical evidence for reservoir size distributions at hand, we can justify statistically to some degree how it came about. A few parameters have important consideration to how the sizes of reservoirs evolved: migration rate, available supply, and time. A concept that we will revisit several times involves the randomizing factor of dispersion.

4.1. FILLING THE RESERVOIRS

If not for these structural traps, we may never have had the chance to even encounter reservoirs of oil. The natural driving force of entropy tends to mix materials to a uniform consistency over time, and only the addition of energy or the formation of enclosures with a sufficient energy barrier allows some sort of homogeneity of matter such as we find with oil reservoirs. We can understand how the oil forms with some clarity.

Assume that the formation of oil over millions of years involves the following basic steps:

1. Formation of a layer of organic material (mainly prehistoric zooplankton, algae) at the bottom of a lake or ocean under anoxic conditions (no oxygen).
2. Sediment burial and diagenesis: The rise in pressure and temperature is transforming the organic materials into kerogen.
3. Catagenesis (or cracking): Organic kerogen transforms into lighter hydrocarbons.
4. Migration: Because hydrocarbons are less dense and more mobile than their surroundings, they can migrate into adjacent porous rock layers, with gravitationally caused pressure providing an assist.
5. Entrapment: Eventually, the oil is collected within a reservoir rock below a seal or cap rock, with low permeability that impedes the escape of hydrocarbons from the reservoir rock.

From considerations of steps 4 and 5 and drawing a parallel analogy to material nucleation and growth processes (There are many practical similarities between the two processes. For example, instead of individual atoms and molecules, we deal with quantities on the order of million barrels of oil; yet the fundamental processes remain the same: diffusion, drift, conservation of matter, rate equations, etc.), one can grasp the fundamentals that go into oil reservoir size distributions. Deep physical processes go into the distribution of field sizes, yet some basic statistical ideas surrounding kinetic growth laws may prove more useful than understanding the fundamental physics of the process. To make the case even stronger, we can use the same ideas from a model of dispersive discovery to demonstrate how us humans sweep through a volume searching for oil leading to oil discoveries, so too can oil diffuse or migrate to *discover* pockets that lead to larger reservoirs.

The premise that varying rates of advance can disperse the ultimate observable measure leads to the distribution we see. For oil discovery, the amount discovered gets dispersed with time, while with field sizes, the dispersion occurs with time as well, but in a much slower glacially paced geological time. For the latter, we will never see any changes in our lifetime, but much like tree rings and glacial cores can tell us about past Earth climates, the statistics of the size distribution can tell us about the past field size growth dynamics.

In 2006, Laherrère estimates that worldwide we have had on the order of 11,500 crude oil discoveries outside of onshore United States (Laherrere, 2006); other estimates range up to 50,000 (Robelius, 2007; Sorrell et al., 2012) since the United States alone has 31,000 as of 1989 (Ivanhoe & Leckie, 1993). If this is considered over a range of 100 years, there is a relatively small sample size to deal with per year. This small sample number over a reservoir size distribution has traditionally followed a lognormal function (Another distribution often cited to describe reservoir sizes is called the Pareto distribution, aka Zipf's law (Deffeyes & Silverman, 2004). This uses hyperbolic curves so it has convergence problems, so a truncation is usually applied.) (Smith, 1980), which has the property of preventing negative sizes by transforming the independent variable by its logarithm (i.e., logs of the values follow a normal distribution).

As the variance tightens around the mean, the shape of the curve peaks away from zero. But importantly, a large variance allows the larger than average sizes (the *supergiants*) to appear.

4.2. DISPERSIVE AGGREGATION MODEL OF RESERVOIR SIZES

From consideration of the fundamental process, we can assert that a peaked distribution (away from small sizes) likely arises from coalescence and agglomeration of deposits. Much like cloud droplet and aerosol particulate distribution (which also show a definite peak in average size due to coalescence), oil deposits have a similarity in structure, if not scale, that we can likely trace to similar fundamental processes.

The model derived next seems to work better than conventional heuristic models (such as the Pareto, lognormal, and fractal), and it derives in a similar manner, but in reverse, to the discovery process itself. If oil can tend to seek out itself or cluster via settling in low energy states and by increasing entropy via diffusing from regions of high concentration, we can consider this as a discovery process. So, as an analogy, we assume that oil can essentially *find* itself and thus pool up to some degree. By the same token, the ancient biological matter tended to accumulate in a similar way. In either case, this process has taken place over the span of millions of years. After this *discovery* or aggregation takes place, the oil does not get extracted like it would in a human-accelerated discovery process, but it gets stored in situ, ready to be rediscovered by humans and of course consumed in a much shorter time, by many orders of magnitude, than it took to generate.

We first assume that oil does indeed migrate from its original creation point through permeable rock to such traps. The buried organic material exists at great depth where it transforms into lighter hydrocarbons by heat and pressure. Then the hydrocarbons eventually start migrating from the source rock to adjacent rock layers. We treat the rate r at which it does this as a stochastic variable with a probability density (oil migration acts as a random process whereby the combined drift and diffusion rate variation follows an exponential law, so the variation equals the mean according to max entropy):

$$p(r|g) = \frac{1}{g} e^{-\frac{r}{g}} \tag{4.1}$$

This introduces two concepts at once: the idea that we do not assume a single rate (i.e., assume instead dispersion) together with the idea that we can only assume at best a mean (as the growth rate g) and treat the standard deviation as equivalent to the mean. This type of assumption makes the least presuppositions as to what has happened: we know we have a mean value, but beyond that, the rate can vary to the maximum entropy limit. To put a label on it, we will refer to this mechanism as *entropic dispersion*.

If we next assume that a collection of these rates can act to sweep out a selected volume of somewhat uniformly deposited oil, then over time we can imagine that a structural trap can collect this migrating oil. Intuitively, we can imagine, since these formed over many different timescales of the Earth's history, that we will obtain a distribution of partially filled reservoir sizes according to how long they have collected migrating oil.

Let us say that the oil diffuses upwardly from the source rock, so for a given time period t, oil will diffuse over a distance $x = rt$; a simple variable change gives (By nomenclature convention, we define two classes of probabilities, which differ by how the probability densities normalize: Conditional probability: p(random variable | parameters). Joint probability: p(random var1, random var2).)

$$p(r|g) = \frac{1}{gt} e^{-\frac{x}{gt}} \tag{4.2}$$

Over time, the probability that some oil will migrate at least a x_0 distance is

$$p(x > x_0|g,t) = \int_0^{\infty} p(x|g,t)\,dx = e^{-\frac{x_0}{gt}} \tag{4.3}$$

Alternatively, the following relation tells us the cumulative probability of the distance covered by material after time t. This again assumes a distance traveled $x = rt$.

$$P(x_0|g) = \int_{r=\frac{x_0}{t}}^{\infty} p(r)\,dr = e^{-\frac{x_0}{gt}} \tag{4.4}$$

This relation also crops up in terms of the population balance equation. It basically relates a conservation of particles law, in that we do not lose track of any material due to a flow. If no oil trap (or seal cap) exists, all the migrating oil will ultimately dissipate and disappear.

So next, we must accumulate this over a volume or depth at which we think the oil exists within. Let us assume that a seal cap exists at some depth x. The simplest approximation assumes that the oil gets distributed to a mean depth (L) with a similar exponential distribution:

$$f(x|L) = \frac{1}{L}\cdot e^{-\frac{x}{L}} \tag{4.5}$$

Combining the two relations turns into an a priori probability for the expected cumulative transfer after time t through the volume. Integrating over the entire Earth's crust column (this vertical column has a horizontal cross section of unity) gives the average oil trapped, U, at a mean depth:

$$\bar{U}(t|L) = \int_0^{\infty} f(x|L)\cdot P(x|g)\,dx = \int_0^{\infty} f(x|L)\cdot e^{-\frac{x}{gt}}dx$$

$$\bar{U}(t|L) = \frac{1}{1+\frac{L}{gt}} \tag{4.6}$$

For the last assumption, we note that if t gets evenly spread from the start of prehistory, some hundreds of millions of years ago, then the value $g\cdot t$ becomes the effective collected thickness W of a distribution of reservoirs by $W = k\cdot g\cdot t$, where we add a factor k to indicate collection efficiency. The collection or trap efficiency factor works in conjunction with the migration drift factor g (understood as some product of reservoir rock porosity, oil saturation, formation factor, and seal impermeability factor). Alternatively, we can interpret the stochastic variable W as the maximum net reservoir thickness that would develop over a diffusion time t if a perfect seal cap is situated near the mean depth L (see Fig. 4.1). The term kL sets the potential maximum net thickness achievable if all the reservoir rock between the source rock and the seal cap saturate with oil, so it turns into a type of hyperbolic discounting probability distribution (derived from the odds function described earlier):

$$\bar{U}(t\,|\,L) = \frac{1}{1+\frac{k\cdot L}{W}} \tag{4.7}$$

This relation states that the cumulative probability of reservoirs of less than or equal to W starts at 0 for very small reservoirs and slowly approaches 1 (unity) for the largest possible reservoir. In practical terms, if L remains close to zero, nature has a greater chance to capture large amounts of migrating oil. On the contrary, if L takes on a large value, there will be no significant accumulation because of the large distance between the source rock and the reservoir rock.

From now on, we can work in terms of field size Size = $W\cdot A$ by integrating on a given geographical area A

$$\bar{U}(\text{Size}|L) = \frac{1}{1+\frac{k\cdot L\cdot A}{\text{Size}}} \tag{4.8}$$

Note that if we set <Size> to the characteristic median field size (defined by the cumulative distribution equaling 0.5), then the equation reduces to

$$\bar{U}(S < \text{Size}) = \frac{1}{1+\frac{\langle\text{Size}\rangle}{\text{Size}}} \tag{4.9}$$

This again describes the cumulative distribution of all reservoirs below a certain size. If we need to know the cumulative distribution above a certain size, we take the complement of this distribution, which results in the subtle difference of inverting the ratio in the denominator:

$$\bar{U}(S > \text{Size}) = \frac{1}{1+\frac{\text{Size}}{\langle\text{Size}\rangle}} \tag{4.10}$$

In a moment we will see how this gets compared to actual data in terms of a rank histogram, but we can glean some insight by looking at the probability density function (PDF) corresponding to the derivative of the cumulative:

$$p(\text{Size}) = \frac{d\bar{U}}{d\text{Size}} = \frac{\langle\text{Size}\rangle}{\left(\langle\text{Size}\rangle+\text{Size}\right)^2} \tag{4.11}$$