

Chapter 27

Bayesianism in the Geosciences



Jef Caers

Abstract Bayesianism is currently one of the leading ways of scientific thinking. Due to its novelty, the paradigm still has many interpretations, in particular with regard to the notion of “prior distribution”. In this chapter, Bayesianism is introduced within the historical context of the evolving notions of scientific reasoning such as inductionism, deductions, falsificationism and paradigms. From these notions, the current use of Bayesianism in the geosciences is elaborated from the viewpoint of uncertainty quantification, which has considerable relevance to practical applications of geosciences such as in oil/gas, groundwater, geothermal energy or contamination. The chapter concludes with some future perspectives on building realistic prior distributions for such applications.

27.1 Introduction

Much of the topic of research within the IAMG community involves developing tools for prediction: what is the grade? The volume of Oil in Place? The spatio-temporal changes of a contaminant plume? Making realistic predictions, meaning providing realistic uncertainty quantification, is key to making informed decisions. Decisions and their consequences are what matters in the end, not the kriging map of gold, or simulated permeability, or hydraulic conductivity. These are only intermediate steps to decision-making. In this chapter, I focus on a fundamental discussion on how we make predictions in the Geosciences and about the current leading paradigm: Bayesianism. This chapter is a revised version of the book “Quantifying Uncertainty in Subsurface Systems”, Scheidt et al. Wiley Blackwell, 2018. The term UQ is therefore used for “Uncertainty Quantification”

Most of our applications involve three major components: data, a model and a decision. For example, in contaminant hydrology, we need to decide on a

J. Caers (✉)
Stanford University, Stanford, USA
e-mail: jcaers@stanford.edu

© The Author(s) 2018
B. S. Daya Sagar et al. (eds.), *Handbook of Mathematical Geosciences*,
https://doi.org/10.1007/978-3-319-78999-6_27

remediation strategy or simply a decision to clean or not. We collect data: geochemical samples, geological studies, possibly even some geophysical surveys. We build models: a reactive transport model, a geostatistical model of spatial properties, a geochemical model. How does this all come together? Bayesian modeling is usually invoked as a way of integrating all these components. But what really constitutes “Bayesian” modeling? Thomas Bayes did not write Bayes’ rule in the form we often see it in textbooks. However, after a long period of being mostly ignored in history, his idea of using a “prior” distribution heralded a new way of scientific reasoning which can be broadly classified as: Bayesianism. The aim of this chapter is to frame Bayesianism within the historical context of other forms of scientific reasoning such as induction, deduction, falsification, intuitionism and others. The application of Bayesianism is then discussed in the context of uncertainty quantification and specific to the Geosciences. This makes sense since quantifying uncertainty is about quantifying a lack of understanding or lack of knowledge. Science is all about creating knowledge. But then, what do we understand and what exactly is knowledge (the field of epistemology)? How can this ever be quantified with a consistent set of axioms and definitions, that is, if a mathematical approach is taken? Is such quantification unique? Is it rational at all to quantify uncertainty? Are we in agreement as to what Bayesianism really is?

These questions are not just practical questions towards engineering solutions, but to a deeper discussion around uncertainty. This discussion is philosophical, a discussion at the intersection of philosophy, science and mathematics. The science of studying knowledge and as a result, uncertainty. In many papers published journals that address uncertainty in subsurface systems, or in any system for that matter, philosophical views are rarely touched upon. Many such publications would start with the “we take the Bayesian approach...” or, “we take a fuzzy logic approach to...” But what entails making this decision? Papers quickly become about algebra and calculus. Bayes or any other way of inferential reasoning is simply seen as a set of methodologies, technical tools and computer programs. The emphasis lies on the beauty of the calculus, solving the puzzle, improving “accuracy” not on any desire of deeper understanding to what exactly one is quantifying. A pragmatic realist may state that in the end, the answer is provided by the computer codes, based on the developed calculus. Ultimately, everything is about bits and bytes and transistors amplifying or switching electronic signals; inputs and outputs. The debate is then which method is better, but such debate is only within the choices of the particular way of reasoning about uncertainty. That particular choice is rarely discussed. The paradigm is blindly accepted.

Bayes is like old medicine, we know how it works, what the side effects are and has been debated, tweaked, improved, discussed since Reverend Bayes’ account was published by Price (1763). Our discussion will start with a general overview of the scientific method and the philosophy of science. This discussion will be useful in the sense that it will help introduce Bayesianism, as a way of inductive reasoning, compared to very different ways of reasoning. Bayes is popular, but not accepted by all (Earman 1992; Wang 2004; Gelman 2008; Klir 1994).

27.2 A Historical Perspective

In the philosophy of sciences, fundamental questions are posed such as: what is a “law of nature”? How much evidence and what kind of evidence should we use to confirm a hypothesis? Can we ever confirm hypotheses as truths? What is truth? Why do we appear to rely on inaccurate theories (e.g. Newtonian physics) in the light of clear evidence that they are false and should be falsified? How does science and the scientific method work? What is science and what is not (the demarcation problem)? Associated with the philosophy of science are concepts such as epistemology (study of knowledge), empiricism (the importance of evidence), induction and deduction, parsimony, falsification, paradigm.... all of which will be discussed in this chapter.

Aristotle (384-322 BC) is often considered to be the founder of both science and the philosophy of science. His work covers many areas such as physics, astronomy, psychology, biology, and chemistry, mathematics, and epistemology. Attempting to not solely be Euro-centric, one should also mention the scientist and philosopher Ibn al-Haytham (Alhazen), who could easily be called the inventor of the peer-review system, on which this chapter too is created. In the modern era, Galileo Galilei and Francis Bacon take over from the Greek philosophy of thought (rationality) over evidence (empiricism). Rationalism was continued by Rene Descartes. David Hume introduced the problem of induction. A synthesis of rationalism and empiricism was provided by Emanuel Kant. Logical positivism (Wittgenstein, Bertrand Russel, Carl Hempel) ruled much of the early twentieth century. For example, Bertrand Russel attempted to reduce all of mathematics to logic (logicism). Any scientific theory then requires a method of verification using a logic calculus in conjunction with the evidence, to prove such theory true or false. Karl Popper appeared on the scene as a reaction to this type of reasoning, replacing verifiability with falsifiability, meaning that for a method to be called scientific, it should be possible to construct an experiment or acquire evidence that can falsify it. More recently Thomas Kuhn (and later Imre Lakatos) rejected the idea that one method dominates science. They see the evolution of science through structures, programs and paradigms. Some philosophers such as Feyerabend go even further (“Against method”, Feyerabend 1993) stating that no methodological rules really exist (or should exist).

The evolution of the philosophy of science has relevance to UQ. Simply replace the concept of “theory” with “model”, and observations/evidence with data. There is much to learn from how people’s viewpoints towards scientific discovery differs; how they have changed and how such change has affected our ways of quantifying uncertainty. One of the aims of this chapter therefore is to show that there is not really a single objective approach to uncertainty quantification based on some laws or rules provided by a passive, single entity (the truth-bearing clairvoyant God!). Uncertainty quantification just like science is dynamic, relies on interaction between data, models and predictions and evolving views on how these components interact. It is with high certainty that few methods covered in this chapter will not be used in 100 years; just consider the history of science as evidence.

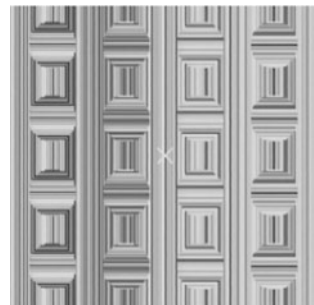
27.3 Science as Knowledge Derived from Facts, Data or Experience

Science has gained considerable credibility, including in everyday life, because it is sold as “being derived from facts”. It provides an air of authority, of truth to what are mainly uncertainties in daily life. This was basically the view with the birth of modern science in the seventeenth century. The philosophies that exalt this view are empiricism and positivism. Empiricism states that knowledge can only come from sensory experience. The common view was that (1) sensory experience produces facts to objective observers, (2) facts are prior to theories (3) facts are the only reliable basis for knowledge.

Empiricism is still very much alive in the daily practice of data collection, model building and uncertainty quantification. In fact, many scientists find UQ inherently “too subjective” and of lesser standing than “data”, physical theories or numerical modeling. Many claim that decisions should be based merely on observations, not models.

Seeing is believing. “Data is objective, models are subjective”. If facts are to be derived from sensory experience, mostly what we see, then consider Fig. 27.1. Most readers see a panel of squares, perhaps from a nice armoire. Others (very few) see circles and perhaps will interpret this as an abstract piece of art with interesting geometric patterns. Those who don’t see circles at first, need to simply look longer, with different focusing of their retinas. Hence, there seems to be more than meets the eyeball (Hanson 1958). Consider another example in Fig. 27.2. What do you see? Most will recognize this as a section of a geophysical image (whether seismic, radar etc....). A well-trained geophysicist will potentially observe a “bright spot” which may indicate the presence of a gas (methane, carbon dioxide) in the sub-surface formations. A sedimentologist may observe deltaic formations consisting of channel stacks. Hence, the experience in viewing an object is highly dependent on the interpretation of the viewer and not the pure sensory light perceptions hitting one’s retina. In fact, Fig. 27.2 is a modern abstract work of art by Mark Bardford (1963) on display in the San Francisco Museum of Modern Art (September 2016).

Fig. 27.1 How many circles do you see?



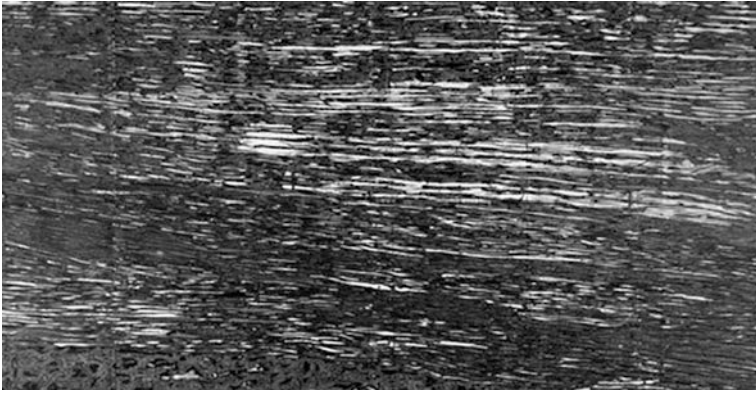


Fig. 27.2 What do you see?

Anyone can be trained to make interpretations, and this is usually how education proceeds. Even pigeons can be trained to spot cancers as well as humans, Levenson et al., PLOS ONE (18 November 2015) <http://www.sciencemag.org/news/2015/11/pigeons-spot-cancer-well-human-experts>. But this idea may also backfire. First off, the experts may not do better than random (Financial times, March 31, 2013: “Monkey beats man on stock market picks”, based on a study by the Cass Business School in London), or worse produce cognitive biases, as pointed out by a study of interpretation seismic images (Bond et al. 2007).

First facts, then theory. Translated to our UQ realm as “first data, then models”. Let’s consider another example in Fig. 27.3, now with actual geophysical data and not a painting. A statement of fact would then be “this is a bright spot”. Then, in the empiricist view, deduction, conclusions can be derived from it (“It contains gas”).

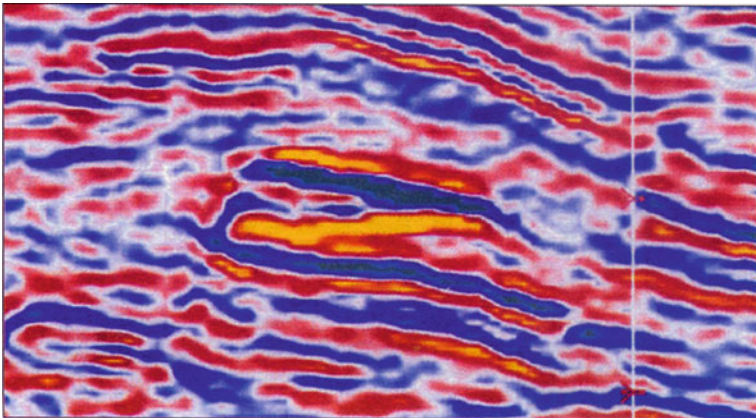


Fig. 27.3 Not art, just a geophysical image

However, what is relevant here is the person making this statement. A lay person will state as fact “There are squiggly lines”. This shows that any observable fact is influenced by knowledge (“the theory”) of the object of study. Statements of fact are therefore not simply recordings of visual perceptions. Additionally, quite an amount of knowledge is needed to consider taking the geophysical survey in the first place, hence facts do not proceed theory. This is the case for the example here but a reality for many scientific discoveries (we need to know where to look). A more nuanced view therefore is that data and models interact with each other.

Facts as the basis for knowledge. “Data precedes the model”. If facts depend on observers resulting in statements that depend on such observers, and if such statements are inherently subjective, then can we trust data as a prerequisite to models (data precede models)? It is now clear that data does not come without a model itself, and hence if the wrong “data model” is used, then the data will be used to build incorrect models. “If I jump in the air and observe that I land on the same spot, then ‘obviously’ the Earth is not moving under my feet”. Clearly the “data model” used here is lacking the concept (theory) of inertia. This again reinforces the idea that in modeling, and in particular UQ, data does not and should precede the model, or that one is subjective and the other somehow is not.

27.4 The Role of Experiments—Data

Progress in science is usually achieved by experimentation, the acquisition of information in a laboratory or field setting. Since “data” is central to uncertainty quantification, we spend some time on what “data” is, what “experiments” aim to achieve and what the pitfalls are in doing so.

First, the experiment is not without the “experimenter”. Perceptual judgements may be unreliable, and hence such reliance needs to be minimized as much as possible. For example, in Fig. 27.4, the uninformed observer may notice that the moon is larger when on the horizon, compared to higher up in the sky, which is merely an optical illusion (on which there still is no consensus as to why). Observations are therefore said to be both objective as well as fallible. Objective in the sense that they are shared (in public, presentations, papers, online) and subject to further tests (such measuring of the actual moon size by means of instruments, revealing the optical illusion). Often such progress happens when more advances in the ways of testing or gathering data occur.

Believing that a certain acquisition of data will resolve all uncertainty and lead to determinism on which “objective” decisions is an illusion because the real world involves many kinds of physical/chemical/biological processes that cannot be captured by one way of experimentation. For example, performing a conservative tracer test, to reveal better hydraulic conductivity, may in fact be influenced by the reactions in the subsurface taking place while doing such an experiment. Hence the

Fig. 27.4 The harvest moon appearing gigantic as compared to the moon in the high sky (https://commons.wikimedia.org/wiki/File:Harvest_Moon_over_looking_vineyards.jpg)



hydraulic conductivity measured and interpreted through some modeling without geochemical reactions may provide a false sense of certainty about the information deduced from such an experiment. In general, it is very difficult to isolate a specific target of investigation in the context of one type of experiment or data acquisition. A good example is in the interpretation of 4D geophysics (repeated geophysics). The idea of the repetition is to remove the influence of those properties that do not change in time, and therefore reveal only those that do change, for example, a change in pressure, a change in saturation, etc. ... However, many processes may be at work at the same time, a change in pressure, in saturation, rock compressibility, even porosity and permeability, geomechanical effects, etc. ... Hence someone interested in the movement of fluids (change in saturation) is left with a great deal of difficulty in unscrambling the time signature of geophysical sensing data. Furthermore, the inversion of data into a target of interest often ignores all these interacting effects. Therefore, it does not make sense to state that a pump test or a well test reveals permeability, it only reveals a pressure change under the conditions of the test and of the site in question, and many of these conditions may remain unknown or uncertain.

An issue that arises in experimentation is the possibility of a form of circular reasoning that may exist between an experimental set-up and a computer model aiming to reproduce the experimental set-up. If experiments are to be conducted to reveal something important about the subsurface (e.g. flow experiments in a lab), then often the results of such experiments are “validated” by a computer model. Is the physical/chemical/biological model implemented in the computer code derived from the experimental result, or, are the computer models used to judge the adequacy of the result? Do theories vindicate experiments and do experiments vindicate the stated theory? To study these issues better, we introduce the notion of induction and deduction.

27.5 Induction Versus Deduction

Bayesianism is based on inductive logic (Howson 1991; Howson et al. 1993; Chalmers 1999; Jaynes 2003; Gelman et al. 2004), although some argue that it is based both on induction and deduction (Gelman and Shalizi 2013). Given the above consideration (and limitations) of experiments (in a scientific context) and data (in a UQ context), the question now arises on how to derive theories from these observations. Scientific experimentation, modeling, studies often rely on a logic to make certain claims. Induction and deductions are such kinds of logic. What such logic offers, is a connection between premises and conclusions:

1. All deltaic systems contain clastic sands.
2. The subsurface system under study is deltaic.
3. The subsurface system contains clastic sands.

This logical deduction is obvious, but such logic only establishes a connection between premises 1 and 2 and the conclusion 3, it does not establish the truth of any of these statements. If that would be the case, then also:

1. All deltaic systems contain steel;
2. The subsurface system under study is deltaic;
3. The subsurface system contains steel.

is equally “logic”. The broader question therefore is if scientific theories can be derived from observations. The same question occurs in the context of UQ: can models be derived from data. Consider an experiment in a lab doing a set of n experiments.

Premises:

1. The reservoir rock is water-wet in sample 1.
2. The reservoir rock is water-wet in sample 2.
3. The reservoir rock is water-wet in sample 3.

...

20. The reservoir rock is water-wet in sample 20.

Conclusion: the reservoir is water-wet (and hence not oil-wet).

This simple idea is mimicked from Bertrand Russell’s Turkey argument (in his case it was a chicken). “I (the turkey) am fed at 9 am” day after day, hence “I am always fed at 9 am”, until the day before Thanksgiving (Chalmers 1999). Another form of induction occurred in 1907: “But in all my experience, I have never been in any accident ... of any sort worth speaking about. I have seen but one vessel in distress in all my years at sea. I never saw a wreck and never have been wrecked nor was I ever in any predicament that threatened to end in disaster of any sort. (E. J. Smith 1907, Captain, RMS Titanic)”.

Any model or theory derived from observations can never be proven in the sense as being derived from it (David Hume).

This does not mean that induction (deriving models from observations) is completely useless. Some inductions are more warranted than others. Specifically, in the case when the observations set is “large”, performed and under a “wide variety of conditions”, although these qualitative statements depend clearly on the specific case. “When I swim with hungry sharks, I get bitten”, needs really be asserted only once.

The second qualification (variety of conditions) requires some elaboration because we will return to it when discussing Bayesianism. Which conditions are being tested is important (the age of the driller for example is not), hence in doing so we rely on some prior knowledge of the particular model or theory being derived. Such prior knowledge will determine which factors will be studied, which are influencing the theory/model and which not. Hence the question is to how this “prior knowledge” itself is asserted by observations. One runs into the never-ending chain of what prior knowledge is used to derive prior knowledge. This point was made clear by David Hume, an eighteenth-century Scottish philosopher (Hume 2000, originally 1739). Often the principle of induction is argued because it has “worked” from experience. The reader needs simply to replace the example of the water-wet rocks with “Induction has worked in case *j*” etc.... to understand that induction is, in this way, “proven” by means of induction. The way out of this “mess” is to not make true/false statements, but to use induction in a probabilistic sense (probably true), a point to which we will return when addressing Bayesianism.

27.6 Falsificationism

A Reaction to Induction

Falsificationism, as championed by Karl Popper (1959) starting in the 1920s was born partly as a reaction to inductionism (and logical positivism). Popper claimed that science should not involve any induction (theories derived from observations). Instead, theories are seen as speculative or tentative, as created by the human intellect, usually to overcome limitations of previous theories. Once stated, such theories need to be tested rigorously with observations. Theories that are inconsistent with such observation should be rejected (falsified). The theories that survive are the best theories, currently. Hence, falsificationism has a time component and aims to describe progress in science, where new theories are born out of old ones by a process of falsification.

In terms of UQ, one can then see models not as true representations of actual reality but as hypotheses. One has as many hypotheses as models. Such a hypothesis can be constrained by previous knowledge, but real field data should be used not to confirm a model (it confirms this with data) but to falsify a model (reject, the model does not confirm with data). A simple example illustrates the difference:

Induction:

Premise: All rock samples are sandstones.

Conclusion: The subsurface system contains only sandstone.

Falsification:

Premise: A sample has been observed that is shale.

Conclusion: The subsurface system does not consist just of sandstone.

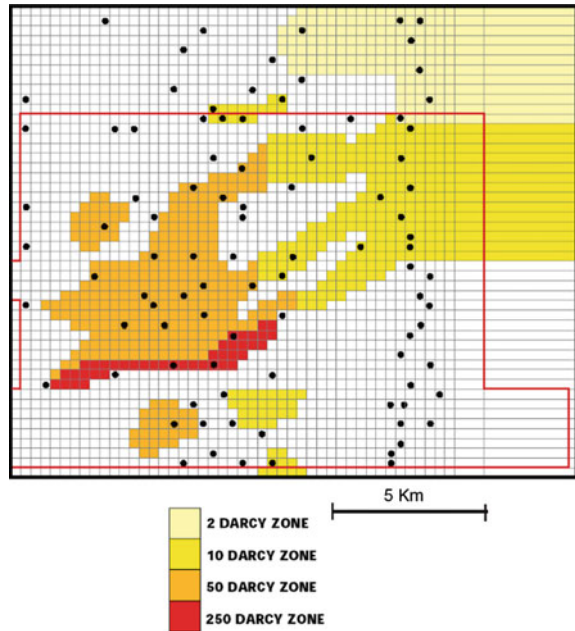
The latter is clearly a logically valid deduction (true). Falsification therefore can only proceed with hypotheses that are falsifiable (this does not mean that one has to falsify the observations, but that such observation could exist). Some hypotheses are not falsifiable; for example, “the subsurface system consists of rock that are sandstone or not sandstone”. This then raises the question of the degree of falsifiability of a hypothesis and the strength (precision) of the observation in falsifying. Not all hypotheses are equally falsifiable and not all observations should be treated on the same footing. A strong hypothesis is one that makes strong claims, there is a difference between:

1. Significant accumulation in the Mississippi delta requires the existence of a river system; and
2. Significant accumulation in all deltas require the existence of a river system.

Clearly 2 has more consequences than 1. Falsification therefore invites stating bold conjectures rather than safe conjectures. Science advances through a large number of bold conjectures that would be easily falsifiable. As a result, a hypothesis *B* that is offered after hypothesis *A* should also be more falsifiable.

The latter has considerable implications in UQ and model building. Inductionists tend to bet on one model, the best possible, best explaining most observations, within a static context, without the idea that the model they are building will evolve. Inductionists do evolve models, but that is not the outset of their viewpoint, there is always the hope that the best possible will remain the best possible. The problem with this inductionist attitude is that new observations that cannot be fitted into the current model are used to “fix” the model with ad hoc modifications. A great example of this can be found in the largest oil reservoir in the world, namely the Ghawar field (see *Twilight in the Desert: The Coming Saudi Oil Shock and the World Economy*, Matt Simmons). Before 2000, most modelers (geologists, geophysicist, engineers) did not consider fractures as being a driving heterogeneity for oil production. However, flow meter observations in wells indicated significant permeability. To account for this data, the existing models with already large permeabilities (1000–10,000mD) were modified to 200D, see Fig. 27.5. While this dramatic increase in permeability in certain zones did lead to explaining the flow meter data, the ad hoc modification cannot be properly tested with the current observations. It is just a fix to the model (the current “theory” of no fractures). Instead, a new test would be needed, such as new drilling to confirm or not the presence of a gigantic cave that can explain such ridiculous permeability values. Today, all models built of the Ghawar field contain fractures.

Fig. 27.5 A reservoir model developed to reflect super permeability channels; note the legend with permeability values (Valle et al. 1993)



Falsificationism does not use ad hoc modification, because the ad hoc modification cannot be falsified. In the Ghawar case, the very notion of fluid flow by means of large matrix permeability tells the falsificationist that bold alternative modifications to the theory are needed and not simple ad hoc fixes, in the same sense that science does not progress by means of fixes. An alternative therefore to the inductionist approach in Ghawar could be as follows: most fluid flow is caused by large permeability, except in some area where it is hypothesized that fractures are present despite the fact that we have not directly observed them. The falsificationist will now proceed by finding the most rigorous (new) test to test this hypothesis. This could consist of acquiring geomechanical studies of the system (something different than flow) or by means of geophysical data that aims to detect fractures (AVOZ data). New hypotheses also need to lead to new tests that can falsify them. This is how progress occurs. The problem is often “time”; a falsificationist takes the path of high risk, high gain, but time may run out on doing experiments that falsify certain hypothesis. “Failures” are often seen as that and not as lessons learned. In the modeling world one often shies away from bold hypothesis (certainly if one wants to obtain government research funding!) and that modelers, as a group tends to gravitate towards some consensus under the banner of being good at “team-work”. It is the view of the authors that such practice is however the death of any realistic UQ. UQ needs to include bold hypothesis, model conjectures that are not the norm, or based on any majority vote, or by playing it safe, being conservative. Uncertainty cannot be reduced by just great team-work,

it will require equally rigorous observations (data) that can falsify any (preferably bold) hypothesis.

This does not mean that inductionist type of modeling and falsification type of modeling cannot co-exist. If inductionism leads to cautious conjectures and falsification leads to bold conjectures. Cautious conjectures may carry little risk, and hence, if they are falsified, then insignificant advance is made. Similarly, if bold conjectures cannot be falsified with new observations, significant advance is made. The matter that is important in all this however is the nature of the background knowledge (recall, the prior knowledge), what is currently known about what is being studied. Any “bold” hypothesis is measured against such background knowledge. Likewise, the degree to which observations can falsify hypothesis needs to be measured against such knowledge. This background knowledge changes over time (what is bold in 2000 may no longer be bold in 2020), and such change, as we will discuss is explicitly modeled in Bayesianism.

Falsificationism in Statistics

Schools of statistical inference are sometimes linked to the falsificationist views of science, in particular the work of Fischer, Neyman and Pearson; all well-known scientists in the field of (frequentist) statistics (Fisher and Fisher 1915; Fisher 1925; Rao 1992; Pearson et al. 1994; Berger 2003; Fallis 2013 for overviews and original papers). Significance tests, confidence intervals p -values are associated with a hypothetico-deductive way of reasoning. Since these methods are pervasive in all areas of science, particularly in UQ, we present some discussion on its rationality as well as the opposing views of inductionism within this context.

Historically, Fisher can be seen as the founder of classical statistics. His work has a falsificationist foundation, steeped in statistical “objectivity” (lack of necessary subjective assumption, which is the norm in Bayesian methods). The now well-known procedure starts by stating a null-hypothesis (a coin is fair), then defines an experiment (flipping), a stopping rule (e.g. number of flips) and a test-statistic (e.g. number of heads). Next, the sampling distribution (each possible value of the test-statistic), assuming the null-hypothesis is true, is calculated. Then, we calculate a probability p that our experiment falls in an extreme group (e.g. 4 heads or less which hypothesis has only a probability of 1.2% for 20 flips). Then a convention is taken to reject (falsify) the hypothesis when the experiment falls in the extreme group, say $p \leq 0.05$.

Fisher’s test works only on isolated hypotheses, which is not how science progresses; often many competing hypotheses are proposed that require testing under some evidence. Neyman and Pearson developed statistical methods that involve rival hypotheses, but again reasoning from an “objective” perspective, without relying on priors or posteriors of Bayesian inductive reasoning. For example, in the case of two competing hypotheses H_1 and H_2 , Neyman-Pearson reasoned that either of the hypotheses are accepted or rejected, leading to two kinds of errors (stating that one is false, while the other is false and vice versa), better known as type I and II errors. Neyman and Pearson improved on Fischer in better defining “low probability”. In the coin example, a priori, any combination of 20

tosses has a probability of 2^{-20} , even under a fair coin, most tosses have small probability. Neyman-Pearson provide some more definition of this critical region (where hypotheses are rejected). If X is the random variable describing the outcome (e.g. a combination of tosses), then the outcome space is defined by the following inequality:

$$L(X) = \frac{P(X|H_1)}{P(X|H_2)} \leq \delta \quad P(L(X) \leq \delta|H_1) = \alpha \quad (27.1)$$

with δ depending on the significance level α and the nature of the hypothesis. This theorem known as the Fundamental Lemma (Neyman and Pearson 1933) defines the most powerful test to reject H_1 in favor of H_2 at significance level α for a threshold δ . The interpretation of likelihood ratio was provided by Bayesianists as the Bayes' factor (the evidential force of evidence). This was however not the interpretation of Neyman-Pearson, who rejected subjective models.

What then does a significance test tell us about the truth (or not) of a hypothesis? Since the reasoning here is in terms of falsification (and not induction), the Neyman-Pearson interpretation is that if a hypothesis is rejected, then "one's actions should be guided by the assumption that it is false" (Lindgren 1976). Neyman-Pearson gladly admit that significance tests tell nothing about whether a hypothesis is true or not. However, they do attach the notion of "in the long run", interpreting the significance level as, for example, the number of times in 1000 times that the same test is being done. The problem here is that no testing can be done and will be done in exactly the same fashion, under the exact same circumstances. This idea would also invoke the notion that under a significance level of 0.05, a *true* hypothesis would be rejected with a probability of 0.05. The latter violates the very reason on which significance tests were formed: events with probability p can never be proven to occur (that requires subjectivity!), let alone with the exact frequency of p .

The point here is to show that classical statistics should not be seen as purely falsificationist, a logical hypothetic-deductive way of reasoning. Reasoning in classical statistics comes with its own subjective notions of personal judgements (choosing which hypothesis, what significance level, stopping rules, critical regions, iid assumptions, Gaussian assumptions etc. ...). This was in fact later acknowledged by Pearson himself (Neyman and Pearson 1967, p. 277).

Limitations of Falsificationism

Falsificationism comes with its own limitations. Just as induction cannot be induced, falsificationism cannot be falsified, as a theory. This becomes clearer when considering real-world development of models or theories. The first problem is similar to the one discussed in using inductive and deductive logic. Logic only works if the premises are true, hence falsification, as a deductive logic cannot distinguish between a faulty observation and a faulty hypothesis. The hypothesis does not have to be false when inconsistent with observations, since observations can be false. This is an important problem in UQ that we will revisit later.

The real world involves considerably more complication than “the subsurface system is deltaic”. Let’s return to our example of monitoring heat storage using geophysics. A problem that is important in this context is to monitor whether the heat plume remains near the well and is compact, so that it does not start to disperse, since then recovery of that heat becomes less efficient. A hypothesis could then be “the heat plume is compact”, geophysical data can be used to falsify this by, for example, observing that the heat plume is indeed influenced by heterogeneity. Unfortunately, such data does not directly observe “temperature”, instead it measures resistivity, which is related to temperature and other factors. Additionally, because monitoring is done at a distance from the plume (at the surface), the issue of limited resolution occurs (any “remote sensing” suffers from this limited resolution). This is then manifested in the inversions of the ERT data into temperature, since many inversion techniques result in smooth versions of actual reality (due to this limited resolution issue), from which the modeler may deduce that homogeneity of the plume is not falsified. How do we find where the error lies? In the instrumentation? In the instrumentation set-up? In the initial and boundary conditions that are required to model the geophysics? In the assumptions about geological variability? In the smoothness of the inversion? Falsification does not provide a direct answer to this. In science, this problem is better known as the Duhem–Quine thesis after Pierre Duhem and Willard Quine (Ariew 1984). This thesis states that it is impossible to falsify a scientific hypothesis in isolation, because the observations required for such falsification themselves rely on additional assumptions (hypothesis) than cannot be falsified separately from the target hypothesis (or vice versa). Any particular statistical method that claims to do so, ignores the physical reality of the problem.

A practical way to deal with this situation is not consider just falsification, but sensitivity to falsification. What impacts the falsification process? Sensitivity, even with limited or approximate physical models provide more information that can lead to (1) changing the way data is acquired (the “value of information”) changing the way the physics of the problem (e.g. the observations) is modeled by focusing on what matters most towards testing the hypothesis.

More broadly, falsification does not really follow the history of the scientific method. Most science has not been developed by means of bold hypothesis that are then falsified. Instead, theories that are falsified are carried through history; most notably, because observations that appear to falsify the theory can be explained by means of causes other than the theory that was the aim of falsification. This is quite common in modeling too: observations are used as claims that a specific physical model does not apply, only to discover at a later time that the physical model was correct but that the data could be explained by some other factor (e.g. a biological reason, instead of a physical reason). Popper himself acknowledged this dogmatism (hanging onto models that have “falsified” to “some degree”). As we will see later, one of the problems in the application of probability (and Bayesianism) is that zero probability models are deemed “certain” not to occur. This may not reflect the actual reality that models falsified under such Popper-Bayes philosophy become

“unfalsified” later by new discoveries and new data. Probability and “Bayesianism” are not at fault here, but the all too common underestimation of uncertainties in many applications.

27.7 Paradigms

Thomas Kuhn

From the previous presentation, one may argue that both induction and falsification provide too much of a fragmented view of the development of scientific theory or methods that often do not agree with reality. Thomas Kuhn, in his chapter “The Structure of Scientific Revolution” (Kuhn 1996) emphasizes the revolutionary character of scientific methods. During such revolution one abandons one “theoretical” concept for another, which is incompatible with the previous one. In addition, the role of scientific communities is more clearly analyzed. Kuhn describes the following evolution of science:

paradigm → crises → revolution → new paradigm → new crisis.

Such a single paradigm consists of certain (theoretical) assumptions, laws, methodologies and applications adapted by members of a scientific community. Probabilistic methods, or Bayesian methods, can be seen as such paradigms: they rely on axioms of probability and the definition of a conditional probability, the use of prior information, subjective beliefs, maximum entropy, principle of indifference, algorithms of MCMC, etc. ... Researchers within this paradigm do not question the fundamentals of such paradigm, the fundamental laws or axioms. Activities within the paradigm are then puzzle-solving activities (e.g. studying convergence of a Markov chain) governed by the rules of the paradigm. Researchers within the paradigm do not criticize the paradigm. It is also typical that many researchers within that paradigm are unaware of the criticism on the paradigm or ignorant as to the exact nature of the paradigm, simply because it is a given: who is really critical of the axioms of probability when developing Markov chain samplers? Or, who questions the notion of conditional probability when performing stochastic inversions? Puzzles that cannot be solved are deemed to be anomalies, often attributed to the lack of understanding of the community about how to solve the puzzle within the paradigm, rather than a question about the paradigm itself. Kuhn considers such unsolved issues as anomalies rather than what Popper would see as potential falsifications of the paradigm. The need for greater awareness and articulation of the assumptions of a single paradigm becomes necessary when the paradigm requires defending against offered alternatives.

Within the context of UQ, a few such paradigms have emerged reflecting the concept of revolution as Kuhn describes. The most “traditional” of paradigms for quantifying uncertainty is by means of probability theory and its extension of Bayesian probability theory (the addition of a definition of conditioning).

We provide here a summary account of the evolution of this paradigm, the criticism leveled, the counter-arguments and the alternatives proposed, in particular possibility theory.

Is Probability Theory the Only Paradigm for Uncertainty Quantification?

The Axioms of Probability: Kolmogorov—Cox

The concept of numerical probability emerged in the mid-seventeenth century. A proper formalization was developed by (Kolmogoroff 1950) based on classical measure theory. A comprehensive study of its foundations is offered in Fine (1973). The treatment is vast and comprises many works of particular note (Gnedenko et al. 1962; Fine 1973; de Finetti 1974, 1995; de Finetti et al. 1975; Jaynes 2003; Feller 2008). Also of note is the work of (Shannon 1948) on uncertainty-based information in probability. In other words, the concept of probability has been around for three centuries. What is probability? It is now generally agreed (the fundamentals of the paradigm) that the axioms of Kolmogorov form the basis, as well as the Bayesian interpretation by Cox (1946). Since most readers are unfamiliar with the Cox theorem and the consequences for interpreting probability, we provide some high-level insight.

Cox works from a set of postulates for example (we focus on just two of three postulates)

- “A proposition p and its negation $\neg p$ is certain” or $plaus(p \cap \neg p) = 1$ which is also termed the logical principle of the excluded middle. *plaus* stands for plausibility.
- Consider now two propositions p and q and the conjunction between them $p \cap q$. This postulate states that the plausibility of the conjunction is the only function of the plausibility of p and the plausibility of q given that p is true. In other words

$$plaus(p \vee q) = f(plaus(p), plaus(q|p)).$$

The traditional laws are recovered when setting *plaus* to be a probability measure or P or stating as per the Cox theorem “any measure of belief is isomorphic to a probability measure”. This seems to suggest that probability is sufficient in dealing with uncertainty, nothing else is needed (due to this isomorphism). The consequence is that one can now perform calculations (a calculus) with “degrees of belief” (subjective probabilities) and even mix probabilities based on subjective belief with probabilities based on frequencies. The question is therefore whether these subjective probabilities are the only legitimate way of calculating uncertainty? For one, probability requires that either the fact is there, or it is not there, nothing is left in the “middle”. This then necessarily means that probability is ill-suited in cases where the excluded middle principle of logic does not apply. What are those cases?

Intuitionism

Probability theory is truth driven. An event occurs or does not occur. The truth will be revealed. From a hard scientific, perhaps engineering approach this seems perfectly fine, but it is not. A key figure in this criticism is the Dutch mathematician and philosopher Jan Brouwer. Brouwer founded the mathematical philosophy of intuitionism countering the then-prevailing formalism, in particular of David Hilbert as well as of Bertrand Russell, claiming that mathematics can be reduced to logic; the epistemological value of mathematical constructs lies in the fundamental nature of this logic.

In simplistic terms perhaps, intuitionists do not accept the law of excluded middle in logic. Intuitionism reasons from the point that science (in particular mathematics) is the result of the mental construction performed by humans rather than principles founded in the actual objective reality. Mathematics is not “truth”, rather it constitutes applications of internally consistent methods used to realize more complex mental constructs, regardless of their possible independent existence in an objective reality. Intuition should be seen in the context of logic as the ability to acquire knowledge without proof or without understanding how the knowledge was acquired.

Classic logic states that existence can be proven by refuting non-existence (the excluded middle principle). For the intuitionist, this is not valid; negation does not entail falseness (lack of existence), it entails that the statement is refuted (a counter example has been found). For an intuitionist a proposition p is stronger than a statement of not (not p). Existence is a mental construction, not proof of non-existence. One specific form and application of this kind of reasoning is fuzzy logic.

Fuzzy Logic

It is often argued that epistemic uncertainty (or knowledge) does not cover all uncertainty (or knowledge) relevant to science. One such particular form of uncertainty is “vagueness” which is borne out of the vagueness contained in language (note that other language dependent uncertainties exist such as “context-driven”). This may seem rather trivial to someone in the hard sciences, but it should be acknowledged that most language constructs (“this is air”, meaning 78% nitrogen, 21% oxygen, and less than 1% of argon, carbon dioxide, and other gases) are a purely theoretical construct, of which we still may not have incomplete understanding. The air that is outside is whatever that substance is, it does not need human constructs, unless humans use it for calculations, which are themselves constructs. Unfortunately (possibly flawed) human constructs is all that we can rely on.

The binary statements “this is air” and “this is not air” are again theoretical human constructs. Setting that aside, most of the concepts of vagueness are used in cases with unclear borders. Science typically works with classification systems (“this is a deltaic deposit”, “this is a fluvial deposit”), but such concepts are again man-made constructs. Nature does not decide to “be fluvial”, it expresses itself through laws of physics, which are still not fully understood.

A neat example presents itself in the September 2016 edition of EOS: “What is magma?” Most would think this is a problem which has already been solved, but it isn’t, mostly due to vagueness in language and the ensuing ambiguity and difference in interpretation by even experts. A new definition is offered by the authors: “*Magma*: naturally occurring, fully or partially molten rock material generated within a planetary body, consisting of melt with or without crystals and gas bubbles and containing a high enough proportion of melt to be capable of intrusion and extrusion.”

Vague statements (“this may be a deltaic deposit”) are difficult to capture with probabilities (it is not impossible, but quite tedious and construed). A problem occurs in setting demarcations. For example, in air pollution, one measures air quality using various indicators such as PM2.5, meaning particles which pass through a size-selective inlet with a 50% efficiency cut-off at 2.5 μm aerodynamic diameter. Then standards are set, using a cut-off to determine what is “healthy” (a green color) and what is “not so healthy” (orange color) and “unhealthy” (a red color) (the humorous reader may also think of terrorist alert levels). Hence, if the particular matter changes by one single particle, the air goes suddenly from “healthy” to “not so healthy”?

In several questions of UQ, both epistemic and vagueness-based uncertainty may occur. Often vagueness uncertainty exists at a higher-level description of the system, while epistemic uncertainty may then deal with questions of estimation because of limited data within the system. For example, policy makers in the environmental sciences may set goals that are vague, such as “should not exceed critical levels”. Such a vague statement then needs to be passed down to the scientist who is required to quantify risk of attaining such levels by means of data and numerical models, where epistemic uncertainty comes into play. In that sense there is no need to be rigorously accurate, for example according to a very specific threshold, given the above argument about such thresholds and classification systems.

Does probability easily apply to vagueness statements? Consider a proposition “the air is borderline unhealthy”. The rule of the excluded middle no longer applies because we cannot say that the air is either not unhealthy or unhealthy. Probabilities no longer sum to one. It has therefore been argued that the propositional logic of probability theory needs to be replaced with another logic: fuzzy logic (although other logics have been proposed such as intuitionistic, trivalent logic, we will limit the discussion to this one alternative).

Fuzzy logic relies on fuzzy set theory (Zadeh 1965, 1975, 2004). An example of fuzzy set A such as “deltaic” is said to be characterized by a membership function $\mu_{\text{deltaic}}(u)$ representing the degree of membership given some information u on the deposit under study, for example $\mu_{\text{deltaic}}(\text{deposit})=0.8$ for a deposit with info u under study. Probabilists often claim that such membership function is nothing more than a conditional probability $P(A|u)$ in disguise (Loginov 1966). The link is made using the following mental construction. Imagine 1000 geologists looking at the same limited info u and then voting whether the deposit is “deltaic” or “fluvial”. Let’s assume these are the two options available. $\mu_{\text{deltaic}}(\text{deposit})=0.832$ means that 832 geologists picked “deltaic” and hence a vote picked at random has 83.2% chance of being deltaic. However, the conditional probability comes with its

limitations as it attempts to cast a very precise answer into what is still a very vague concept. What really is “deltaic”? Deltaic is simply a classification made by humans to describe a certain type of depositional system subject to certain geological processes acting on it. The result is a subsurface configuration, termed architecture of clastic sediments. In modeling subsurface systems, geologists do not observe the processes (the deltaic system) but only the record of it. However, there is still no full agreement as to what is “deltaic” or when “deltaic” ends and “fluvial” starts as we go more upstream? (Recall our discussion on “magma”) What are the processes which are actually happening and how all this gets turned into a subsurface system? Additionally, geologist may not have a consensus on what “deltaic” is, where “fluvial” starts, or, may classify based on personal experiences, different education (schools of thought about “deltaic”), and different education levels. What then does 0.832 really mean? What is the meaning of the difference between 0.832 and 0.831? Is this due to education? Misunderstanding or disagreement on the classification? Lack of data provided? It clearly should be a mix of all this, but probability does not allow an easy discrimination. We find ourselves again with a Duhem–Quine problem.

Fuzzy logic does not take the binary route of voting up or down, but allows a grading in the vote of each member, meaning that it allows for more gradual transition between the two classes for each vote. Each person takes the evidence at his/her value and makes a judgement based on their confidence and education level: I don’t really know, hence 50/50; I am pretty certain, hence 90/10. (More advanced readers in probability theory may now see a mixture of the models of probability stated based on the evidence of what the u is. However, because of the overlapping nature of how evidence is regarded by each voter, these prior probabilities are no longer uniform).

The Dogma of Precision

Clearly probability theory (randomness) does not work well when the event itself is not clearly defined, subject to discussion. Probability theory does not support the concept of a fuzzy event, hence such information (however vague and incomplete) becomes difficult and non-intuitive to account for. Probability theory does not provide a system for computing with fuzzy probabilities expressed as likely, unlikely and not very likely. Subjective probability theory relies on the elicitation rather than the estimation of a fuzzy system. It cannot address questions of the nature “What is the probability that the depositional system *may* be deltaic”. One should question, under all this vagueness and ambiguity what is really the meaning of the digit “2” or “3” is in $P(A|u) = 0.832$. The typical reply of probabilists to possibilists is to “just be more precise” and the problem is solved. But this would ignore a particular form of lack of understanding, which goes to the very nature of UQ. Precision is required that does not agree with the realism of vagueness on concepts, which are as yet imprecise (such as in subsurface systems).

The advantage and the disadvantage of the application of probability to UQ are that, dogmatically, it requires, precision. It is an advantage in the sense that it attempts to render subjectivity into quantification, that the rules are very well understood, the methods deeply practiced, because of the nature of the rigor of the

theory, the community (of 300 years of practice) is vast. But, this rigor does not always jive with reality. Reality is more complex than “Navier Stokes” or “Deltaic”, so we apply rigor to concepts (or even models) that probably deviate considerably from the actual processes occurring in nature. Probabilists often call this “structural” error (yet another classification and often ambiguous concept, because it has many different interpretations) but provide no means of determining what exactly this is and how it should be precisely estimated, as is required by their theories. It is left as a “research question”, but can this question be truly answered within probability theory itself? For the same reasons, probabilistic method (in particular Bayesian, see the following sections are computationally very demanding, exactly because of this dogmatic quest for precision.

Possibility Theory: Alternative or Compliment?

Possibility theory has been popularized by Zadeh (1978), also by Dubois and Prade (1990). The original notion goes back further to the economist (Shackle 1962) studying uncertainty based on degrees of potential surprise of events. Shackle also introduces the notion of conditional possibility (as opposed to conditional probability). Just as probability theory, possibility theory has axioms. Consider Ω to be a finite set, with subsets A and B that are not necessarily disjoint:

axiom 1: $pos(\emptyset) = 0$ (Ω is exhaustive)

axiom 2: $pos(\Omega) = 1$ (no contradiction)

axiom 3: $pos(A \cup B) = \max(pos(A), pos(B))$ (“additivity”)

Noticeable difference with probability theory is that addition is replaced with “max” and the subsets for axiom 3 need not be disjoint. Additionally, probability theory uses a single measure, the probability, whereas possibility theory uses two concepts, the possibility and the necessity of the event. This necessity, another measure is defined as:

$$nec(A) = 1 - pos(\bar{A}) \quad (27.2)$$

If the complement of an event is impossible, then the event is necessary. $nec(A) = 0$ means that A is unnecessary. One should not be “surprised” if A does not occur, it says nothing about $pos(A)$. $nec(A) = 1$ means that A is certainly true, which implies $pos(A) = 1$. Hence nec carries a degree of surprise: $nec(A) = 0.1$ a little bit surprised, $nec(A) = 0.9$ very surprised if A is not true. Possibility also allows for indeterminacy (which is not allowed in epistemic uncertainty), this is captured by $nec(A) = 0, pos(A) = 1$.

Logically then

$$nec(A \cap B) = \min(nec(A), nec(B)) \quad (27.3)$$

Possibility does not follow the rule of the excluded middle because

$$pos(A) + pos(\bar{A}) \geq 1 \tag{27.4}$$

Take the following example. Consider a reservoir. It either contains oil (A) or contains no oil (\bar{A}) (something we like to know!). $pos(A) = 0.5$ means that I am willing to bet that the reservoir contains oil so long as the odds are even *or better*. I would not bet that it contains oil. Hence this describes a degree of belief very different from subjective probabilities.

Possibilities are sometime called “imprecise probabilities” (Hand and Walley 1993) or are interpreted that way. “Imprecise” need not be negative, as discussed above, it has its own advantages, in particular in terms of computation. In probability theory, information is used to update degrees of belief. This is based on Bayes’ rule whose philosophy will be studied more closely in the next section. A counterpart to Bayes’ rule exists in possibility theory, but because of the imprecision of possibilities over probabilities, no unique way exists to update possibilities into a new possibility, given new (vague) information. Recall that Bayes’ rule relies on the product (corresponding to a conjunction in classical logical)

$$P(A|B) = \frac{P(B|A)}{P(B)} P(A) \tag{27.5}$$

Consider first the counterpart of the probability density function $f_X(x)$ in possibility theory: namely the possibility distribution $\pi_X(x)$. Unlike probability densities which could be inferred from data, possibility distributions are always specified by users, and hence take simple form (constant, triangular) functions. Densities express likelihoods, a ratio of the densities assessed in two outcomes denotes how much more (or less) likely one outcome is over the other. A possibility distribution simply states how possible an outcome x is. Hence a possibility distribution is always equal or less than unity (not the case for a density). Also, note that $P(X=x) = 0$, always if X is a continuous variable, while $pos(X=x)$ is not zero everywhere. Similarly, in the case of a joint probability distribution, we can define a joint possibility distribution as $\pi_{X,Y}(x,y)$ and conditional possibility distributions as $\pi_{X|Y}(x|y)$. The objective now is to infer $\pi_{X|Y}(x|y)$ from $\pi_{Y|X}(y|x)$ and $\pi_X(x)$.

As mentioned above, probability theory relies on a logical conjunction, see Fig. 27.6. This conjunction has the following properties:

- $a \cap b = b \cap a$ (commutativity)
- if $a \leq a'$ and $b \leq b'$ then $a \cap b \leq a' \cap b'$ (monotonicity)
- $(a \cap b) \cap c = a \cap (b \cap c)$ (associativity)
- $a \cap 1 = a$ (neutrality)

Possibility theory, as it is based on fuzzy sets, rather than random sets, relies on an extension of the conjunction operation. This new conjunction is termed a

| a | b | $a \cap b$ | a | b | $T_1(a, b)$ | $T_2(a, b)$ | $T_3(a, b)$ |
|---|---|------------|-----|-----|-------------|-------------|-------------|
| 0 | 0 | 0 | 0.1 | 0.1 | 0.1 | 0.01 | 0.053 |
| 1 | 0 | 0 | 0.9 | 0.1 | 0.1 | 0.09 | 0.098 |
| 0 | 1 | 0 | 0.1 | 0.1 | 0.1 | 0.09 | 0.098 |
| 1 | 1 | 1 | 0.9 | 0.9 | 0.9 | 0.81 | 0.82 |

$$T_1(a, b) = \min(a, b); \quad T_2(a, b) = a \cdot b; \quad T_3(a, b) = \frac{ab}{a+b-ab}$$

Fig. 27.6 Example of t-norms for conjunction operations

triangular norm (T-norm) (Jenei and Fodor 1998; Höhle 2003; Klement et al. 2004) because it follows the following four properties:

- $T(a, b) = T(b, a)$ (commutativity)
- if $a \leq a'$ and $b \leq b'$ then $T(a, b) = T(a', b')$ (monotonicity)
- $T(a, T(b, c)) = T(T(a, b), c)$ (associativity)
- $T(a, 1) = a$ (neutrality)

Recall that Cox relied on the postulate that $plaus(p \cap q) = f(plaus(p), plaus(q|p))$. Similarly, possibility theory relies on:

$$\pi_{Y|X}(y|x) = T(\pi_X(x), \pi_{Y|X}(y|x)) = T(\pi_Y(x), \pi_{X|Y}(x|y)) \tag{27.6}$$

For example, for the minimum triangular norms we get

$$\pi_{X|Y}(x|y) = \begin{cases} 1 & \text{if } \pi(x) = \min\{\pi(x), \pi_{Y|X}(x|y)\} \\ \min\{\pi(x), \pi_{Y|X}(x|y)\} & \text{if } \pi(x) > \min\{\pi(x), \pi_{Y|X}(x|y)\} \end{cases} \tag{27.7}$$

and for the product triangular norm, we get something that looks Bayesian

$$\pi_{X|Y}(x|y) = \frac{\pi_{Y|X}(x|y)\pi(x)}{\pi(y)} \tag{27.8}$$

27.8 Bayesianism

Thomas Bayes

Uncertainty quantification, today often has a Bayesian flavor. What does this mean? Most researchers simply invoke Bayes' rule, as a theorem within probability theory.

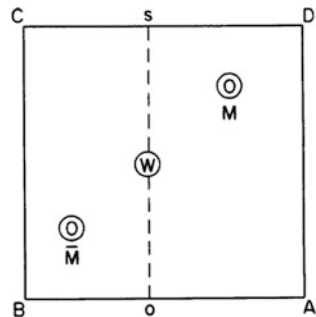
They work within the paradigm. But what is really the paradigm of Bayesianism? It can be seen as a simple set of methodologies, but it can also be regarded as a philosophical approach to doing science, in the same sense as empiricism, positivism, falsificationism or inductionism. The reverend Bayes' would perhaps be somewhat surprised by the scientific revolution and main stream acceptance of the philosophy based on his rule.

Thomas Bayes was a statistician, philosopher and Reverend. Bayes presented a solution to the problem of inverse probability in "An Essay towards Solving a Problem in the Doctrine of Chances". This essay was read after his death, by Richard Price for the Royal Society of London, a year after his death. Bayes' theorem remained in the background until reprinted in 1958, and even then it took a few more decades before an entirely new approach to scientific reasoning, Bayesianism was created (Howson et al. 1993; Earman 1992).

Prior to Bayes' most works on chance were focused on direct inference, such as the number of replications needed to calculate a desired level of probability (how many flips of the coin are needed to assure 50/50 chance?). Bayes' treated the problem of inverse probability: "given the number of times an unknown event has happened and failed: required the chance that the probability of its happening in a single chance lies between any two degrees of probability that can be named" (see the Biometrika publication of Bayes' essay). Bayes' essay has essentially four parts. Part 1 consists of a definition of probability and some basic calculation which are now known as the axioms of probability. The second part uses these calculations in a chance event related to a perfectly leveled billiard table, see Fig. 27.7. Part 3 consists of using the equations obtained from the analysis of the billiard problem to his problem of inverse probability. Part 4 consists of more numerical studies and applications.

Bayes, in his essay, was not concerned with induction and the role of probability in it. Price, however, in the preface to the essay did express a wish that the work would in fact lead to a more rational approach to induction than was then currently available. What is perhaps less known is that "Bayes' theorem" in the form that we now know it, was never written by Bayes'. However, it does occur in the solution to his particular problem. As mentioned above, Bayes' was interested in a chance event with unknown probability (such as in the billiard table problem), given a

Fig. 27.7 Bayes' billiard table: "to be so made and leveled that if either of the ball O and W thrown upon it, there shall be the same probability that it rests upon any one equal part of the plane as another" (Bayes and Price 1763)



number of trials. If M counts the number of times that an event occurs in n trials, then the solution is given through the binomial distribution

$$P(p_1 \leq p \leq p_2 | M = m) = \frac{\int_{p_1}^{p_2} \binom{n}{m} p^m (1-p)^{n-m} P(dp)}{\int_0^1 \binom{n}{m} p^m (1-p)^{n-m} P(dp)} \quad (27.9)$$

where $P(dp)$ is the prior distribution over p . Bayes' insight here is to "suppose the chance is the same that it (p) should lie between any two equi-different degrees". $P(dp) = dp$, in other words the prior is uniform, leading to

$$P(p_1 \leq p \leq p_2 | M = m) = \frac{(n+1)!}{m!(n-m)!} \int_{p_1}^{p_2} \binom{n}{m} p^m (1-p)^{n-m} dp \quad (27.10)$$

Why uniform? Bayes' does not reason from the current principle of indifference (which can be debated, see later), but rather from an operation characterization of an event concerning the probability which we know absolutely nothing about prior to the trials. The use of prior distributions however was one of the key insights of Bayes' that very much lives on.

Rationality for Bayesianism

Bayesians can be regarded more as relativists than absolutists (such as Popper). They believe in prediction based on imperfect theories. For example, they will take an umbrella on their weekend, if their ensemble Kalman filter prediction of the weather at their trip location puts a high (posterior) probability of rain in 3 days. Even if the laws involved are imperfect and probably can be falsified (many weather predictions are completely wrong!), they rely on continued learning from future information and adjustments. Instead of relying on Popper's zero probability (rejected or not), they rely more on an inductive inference yielding non-zero probabilities.

If we now take the general scientific perspective (and not the limited topic of UQ), then Bayesians see science progress by hypothesis, theories and evidence offered towards these hypotheses as all quantified using probabilities. In this general scientific context, we may therefore state hypothesis H , gather evidence E , with $P(H|E)$ the probability of the hypothesis in the light of the evidence, $P(E|H)$ the probability that the evidence occurs when the hypothesis is true, $P(H)$ the probability of the hypothesis without any evidence and $P(E)$ the probability of the evidence, without stating any hypothesis being true.

$$P(H|E) = \frac{P(E|H)}{P(E)} P(H) \quad (27.11)$$

$P(H)$ is also termed the prior probability and $P(H|E)$ the posterior probability. We provided some discussion on a logical way of explaining this theorem (Cox 1946) and the subsequent studies that showed this was not quite as logical as it seems (Halpern 1995, 2011). Few people today know that Bayesian probability has 6 axioms (Dupré and Tiplery 2009). Despite these perhaps rather technical difficulties, a simple logic underlies this rule. Bayes' theorem states that the extent to which some evidence supports a hypothesis is proportional to the degree to which the evidence is predicted by the hypothesis. If the evidence is very likely ("Sandstone has lower acoustic impedance than shale) then the hypothesis ("Acoustic impedance depends on mineral composition") is not supported significantly when indeed we measure that "Sandstone has lower acoustic impedance than shale". If, however, the evidence is deemed very unlikely, (e.g. "Shale has higher acoustic impedance than sandstone"), then the hypothesis of another theorem ("acoustic impedance depends not only on mineralization, but also fluid content") will be highly confirmed (have high posterior probability).

Another interesting concept is how Bayes deals with multiple evidences of the same impact on the hypothesis. Clearly, more evidence leads to an increase in the probability of a hypothesis supported by that evidence. But evidences of the same impact will have a diminishing effect. Consider that a hypothesis has as equal probability as some alternative hypothesis:

$$P(H) = 0.5$$

Now consider multiple evidence sources such that

$$P(H|E_1) = 0.8; P(H|E_2) = 0.8; P(H|E_3) = 0.8;$$

Then according to a model of conditional independence and Bayes' theorem (Bordley 1982; Journel 2002; Clemen and Winkler 2007):

$$P(H|E_2, E_1) = 0.94; P(H|E_3, E_2, E_1) = 0.98;$$

Compounding evidence leads to increasing probability of the hypothesis.

Objective Versus Subjective Probabilities

In the early days of the development of Bayesian approaches, several general principles were stated under which researchers "should" operate, resulting in an "objective" approach to the problem of inference, in the sense that everyone is following that same logic. One such principle is the principle of maximum entropy (Jaynes 1957), of which the principle of indifference (Laplace) is a special case. Subjectivists do not see probabilities as objective (leading to prescribing zero probabilities to well-confirmed ideas). Rather, subjectivists (Howson et al. 1993) see Bayes' theorem as an objective theory of inference. Objective is the sense that *given* prior probabilities and evidence, posterior probabilities are calculated. In that

sense, subjective Bayesian make no claim on the nature of the propositions on which inference is being made (in that sense, they are also deductive).

One interesting application of reasoning in this way results when disagreement occurs on the same model. Consider modeler A (the conformist) who assigns a high probability to some relatively well-accepted modeling hypothesis and low probability to some rare (unexpected) evidence. Consider modeler B (the skeptic) who assigns low probability to the norm and hence high probability to any unexpected evidence. Consequently, when the unexpected evidence occurs and hence is confirmed $P(E|H) = 1$, then the posterior of each is proportional to $1/P(E)$. Modeler A is forced to increase their prior more than the Modeler B. Some Bayesians therefore state that the prior is not that important as continued new evidence is offered. The prior will be “washed out” by cumulating new evidence. This is only true for certain highly idealized situations. It is more likely that two modelers will offer two hypotheses, hence evidence needs to be evaluated against each other. However, there is always a risk that neither model can be confirmed, regardless how much evidence is offered, hence the prior model space is incomplete, which is the exact problem of the objectivist Bayes. Neither objective nor subjective Bayes’ addresses this problem.

Bayes with Ad Hoc Modifications

Returning now to the example of Fig. 27.5. Bayesian theory, if properly applied allow for assessing these ad hoc model modifications. Consider that a certain modeling assumption H is prevailing in multi-phase flow: “oil flow occurs in rock with permeability of 10-10000 md” (H), now this modeling assumption is modified ad hoc to “oil flow occurs in rock with permeability of 10-10000md and 100-200D ($H \cap AdHoc$). However, this ad hoc modification, under H , has very low probability, $P(AdHoc) \simeq 0$ and hence $P(H \cap AdHoc) \simeq 0$. The problem, in reality is that those making the ad hoc modification often do not use Bayesianism, hence never assess or use the prior $P(AdHoc)$.

Criticism of Bayesianism

What is critical to Bayesianism is the concept of “background knowledge”. Probabilities are calculated based on some commonly assumed background knowledge. Recall that theories cannot be isolated and independently tested. This “background” consists of all the available assumptions tangent to the hypothesis at hand. The problem often resulting with using Eq. (27.11) is that such “background knowledge” BK is taken implicit:

$$P_{BK_0}(H|E) \simeq P_{BK_0}(E|H)P_{BK_0}(H) \rightarrow P_{BK_1}(H) \quad (27.12)$$

where 0 indicated at time $t=0$. The posterior then includes the “new knowledge” which is included in the new background knowledge at the next stage $t=1$. A problem occurs when applying this to the real world: what is this “background knowledge”? In reality, the prior and likelihood are not determined by the same person. For example, in our application, the prior may be given by a geologist, the likelihood by a data scientist. It is unlikely that they have the same “background

knowledge” (or even agree on it). A more “honest” way of conveying this issue is to make the background knowledge explicit. Suppose that $BK^{(1)}$ is the background knowledge of person 1, who deals with evidence (the data scientist) then

$$P(H|E \cap BK^{(1)}) \simeq P(E \cap BK^{(1)}|H)P(H|BK^{(1)}) \quad (27.13)$$

Suppose $BK^{(2)}$ is person 2 (geologist) who provides the “prior”, meaning provides background knowledge on his/her own, without evidence. Then, the new posterior can be written as

$$P(H|E \cap BK^{(1)} \cap BK^{(2)}) \simeq P(E \cap BK^{(2)}|H)P(H|BK^{(2)})P(H|BK^{(1)}) \quad (27.14)$$

assuming however, there is no overlap between background knowledge. In practice, the issue that different components of the “system” (model) are done by different modelers with different background knowledge is ignored. Even if one would be aware of this issue, it would be difficult to implement in practice. The ideal Bayesian approach rarely occurs. No single person understands all the detailed aspects of the scientific modeling study at hand. A problem then occurs with dogmatism. The study in Fig. 27.5 illustrates this. Hypotheses that are given very high probability (no fractures) will remain high, particularly in the absence of strong evidence (low to medium $P(E)$). Bayes’ rule will keep assigning very high probabilities to such hypotheses, particularly due to the dogmatic belief of the modeler or the prevailing leading idea of what is going on. This is not the problem of Bayes’, but its common (faulty) application. Bayes’ itself cannot address this.

More common is to select a prior hypothesis based on general principles or mathematical convenience, for example using a maximum entropy principle. Under such a principle, complete ignorance results in choosing for uniform distribution. In all other cases, one should pick the distribution that makes the least claims, from whatever information is currently available, on the hypothesis being studied. The problem here is not so much the ascribing of uniform probabilities but providing a statement of what all the possibilities are (on which then uniform probabilities are assigned). Who chooses these theories/models/hypotheses? Are those the only ones?

The limitation therefore of Bayesianism is that no judgment is leveled to the stated prior probabilities. Hence, any Bayesian analysis is as strong as the analysis of the prior. In subsurface modeling this prior is dominated by the geological understanding of the system. Such geological understanding and its background knowledge is vast, but qualitative. Later we will provide some ideas on how to make quantitative “geological priors”.

Deductive Testing of Inductive Bayesianism

The leading paradigm of Bayesianism is to subscribe to an induction from of reasoning: learning from data. Increasing evidence will lead to increasing probabilities of certain theories, models or hypothesis. As discussed in the previous section, one of the main issues lies in the statement of a prior distribution, the initial universe of

possibilities. Bayesianism assume that a truth exists, that such truth is generated by a probability model, and also than any data/evidence are generated from this model. The main issue occurs when the truth is not even with the support (the range/span) generated by this (prior) probability model. The truth is not part of this initial universe. What happens then? The same goes when the error distribution on the data is chosen at too optimistic a level, in which case the truth may be rejected. Can we verify this? Diagnose this? Figure out whether the problem lies with the data or the model? Given the complexity of models, priors, data in the real world this issue may in fact go undiagnosed if one stops the analysis with the generation of the posterior distribution. Gelman and Shalizi (2013) discuss how mis-specified prior models (the truth is not in the prior) may result in either no solution, multi-model solutions to problems that are unimodal or complete non-sense.

Recent work (Mayo 1996) started to look at these issues. They attempt to frame these tests within classical hypothesis testing. Recall that classical statistics rely on a deductive form of hypothesis testing, which is very similar in flavor to Popper's falsification. In a similar vein, some form of model testing can be performed posterior to the generation of the posterior. Note that Bayesian model averaging (Rings et al. 2012; Henriksen et al. 2012; Refsgaard et al. 2012; Tsai and Elshall 2013) or model selection are not tests of the posterior, rather, they are consequences of the posterior distribution, yet untested! Classical checks are whether posterior models match data, but these are checks based on likelihood (misfit) only.

Consider a more elaborate testing framework. These formal test rely on generating replicates of the data given some model hypothesis and parameters are the truth. Take a simple example of a model hypothesis with two faults ($H =$ two faults) and the parameters θ representing those faults (e.g. dip, azimuth, length etc.). The bootstrap allows for a determination of achieved significance level (*ASL*) as

$$ASL(\theta) = P(S(\mathbf{d}_{rep}) \geq S(\mathbf{d}_{obs}) | H, \theta) \quad (27.15)$$

here, we consider calculating some summary statistic of the data as represented by the function S . This summary statistic could be based on some dimension reduction method; for example, a first or second principal component score. The uncertainty on θ is provided by its posterior distribution, hence we can sample various θ from the posterior. Therefore we first sample \mathbf{d}_{rep} from the following distribution (averaging out over posterior in θ)

$$P(\mathbf{d}_{rep} | H, \mathbf{d}_{obs}) = \int P(\mathbf{d}_{rep} | H, \theta) P(\theta | H, \mathbf{d}_{obs}) d\theta \quad (27.16)$$

and calculate average *ASL* over the posterior distribution. Analytically this equals to

$$ASL = \int ASL(\theta) P(\theta | H, \mathbf{d}_{obs}) d\theta \quad (27.17)$$

or for given limited sample $\theta^{(\ell)}$, $\ell = 1, \dots, L \sim P(\theta | H, \mathbf{d}_{obs})$

$$ASL = \frac{1}{L} \sum_{\ell=1}^L ASL(\boldsymbol{\theta}^{(\ell)}) \quad (27.18)$$

These tests are not used to determine whether a model is true, or even should be falsified but whether discrepancies exist between model and data. The nature of the functions S defines the “severity” of the tests (Mayo 1996). Numerous complex functions will allow for a more severe testing of the prior modeling hypothesis. We can learn how the model fails by generating several of these summary statistics, each representing different elements of the data (a low, a middle and some extreme case etc....).

Within this framework of deductive tests, the prior is no longer treated as “absolute truth”, rather the prior becomes a modeling assumption that is “testable” given the data. Some may however disagree on this point: why should the data be any better than the prior? In the next section, we will try to get out of this trap, by basing priors on physical processes, with the hope that such priors are more realistic representations of the universe of variability, rather than simply relying on statistical methods that are devoid of physics.

27.9 Bayesianism for Subsurface Systems

What is the Nature of Geological Priors?

Constructing Priors from Geological Field Work

In a typical subsurface system, the model variables are parameterized in a certain way, for example with a grid, or a set of objects with certain lengths, widths dips, azimuths etc. What is the prior distribution of these model variables? Since we are dealing with a geological system, e.g. a delta, a fluvial or turbidite systems, a common approach is to do geological field work. This entails measuring and interpreting the observed geological structures, on outcrops, and creating a history of their genesis, with an emphasis on generating (an often qualitative) understanding of the processes that generated the system. The geological literature contains a vast amount of such studies.

To gather all this information and render it relevant for modeling UQ, geological databases based on classification systems have been compiled (mostly by the Oil industry). Analog databases, for example, on proportions, paleo-direction, morphologies and architecture of geological bodies or geological rules of association (Eschard and Doligez 2000; Gibling 2006) for various geological environments (FAKT: Colombera et al. 2012; CarbDB: Jung and Aigner 2012; WODAD: Kenter and Harris 2006; Paleoreefs: Kiessling and Flügel 2002; Pyrcz et al. 2008) have been constructed. Such relational databases employ a classification system based on geological reasoning. For example, the FAKTS database classifies existing studies, whether literature-derived or field-derived from modern or ancient river systems,

according to controlling factors, such as climate, and context-descriptive characteristics, such as river patterns. The database can therefore be queried on both architectural features and boundary conditions to provide the analogs for modeling subsurface systems. The nature of the classification is often hierarchical. The uncertain style or classification, often termed “geological scenario” (Martinius and Naess 2005) and variations within that style.

While such approach appears to gather information, it leaves the question of whether the collection of such information and the extraction of parameters values to state prior distribution produce realistic priors (enough variance, limited bias) for what is actually in the subsurface. Why?

- Objects and dimensions in the field are only apparent. An outcrop is only a 2D section of a 3D systems. This invokes stereological problems in the sense structural characteristics (e.g. shape, size, texture) of 2D outcrops are only apparent properties of the three-dimensional subsurface. These apparent properties can drastically change depending on the position/orientation of the survey (e.g. Beres et al. 1995). Furthermore, interpreted two-dimensional outcrops of the subsurface may be biased because large structures are more frequently observed than small structures (Lantuéjoul 2013). The same issue occurs for those doing 2D geophysical surveys to interpret 3D geometries (Sambrook Smith et al. 2006). For example, quantitative characterization of two-dimensional ground penetrating radar (GPR) imaging (e.g. Bristow and Jol 2003) ignore uncertainty on the three-dimensional subsurface characteristics resulting from the stereological issue.
- The database is purely geometric in nature. It records the end-result of deposition not the process of deposition. In that sense it does not include any physics underlying the processes that took place and therefore may not capture the complexity of geological processes fully to provide a “complete” prior. For that reason, the database may aggregate information that should not be aggregated, simply because each case represents different geological processes, accidentally creating similar geometry. For modeling, this may appear irrelevant (who cares about the process), yet it is highly relevant. Geologists reason based on geological processes, not just the final geometries, hence this “knowledge” should be part of a prior model construction. Clearly prior should not ignore important background knowledge, such as process understanding.

The main limitation is that this pure parameterization-based view (the geometries, dimensions) lacks physical reasoning, hence ignore important prior information. The next section provides some insight into this problem as well as suggests a solution.

Constructing Priors from Laboratory Experiments

Depositional systems are subject to large variability whose very nature is not fully understood. For example, channelized transport systems (fan, rivers, delta, etc.) reconfigure themselves more or less continually in time, and in a manner often difficult to predict. The configurations of natural deposits in the subsurface are thus

uncertain. The quest for quantifying prior uncertainty necessitates understanding the sedimentary systems by means of physical principles, not just information principles (such as the principle of indifference). Quantifying prior uncertainty thus requires stating all configurations of architectures of the system deemed *physically* possible and at what frequency (a probability density) they occur. This probability density need not be Gaussian or uniform. Hence, the question arises: what is this probability density for geological systems, and how does one represent it in a form that can be used for actual predictions using Bayesianism?

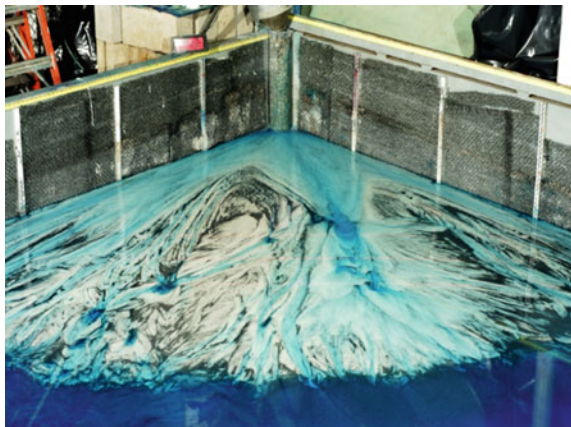
The problem in reality is that we observe geological processes over a very short time span (50 years of satellite data and ground observations), while the deposition of the relevant geological systems we work with in this chapter may span 100.000 years or more. For that reason, the only way to study such system is either by computer models or by laboratory experiment. These computer models solve a set of partial differential equations that describe sediment transport, compaction, diagenesis, erosion, dissolution, etc. (Koltermann and Gorelick 1992; Gabrovsek and Dreybrodt 2010; Nicholas et al. 2013). The main issue here is that PDEs are a limited representation of the actual physical process and require calibration with actual geological observations (such as erosion rules), require boundary conditions and source terms. Often their long computing times limit their usefulness for constructing complete priors.

For that reason, laboratory experiments are increasingly used to study geological deposition, simply because physics occurs naturally, and not as constructed with an artificial computer code. Next, we provide some insight into how laboratory experiments work and how they can be used to create realistic analogs of depositional systems.

Experimenting the Prior

We consider a delta constructed in an experimental sedimentary basin subject to constant external boundary conditions (i.e. sediment flux, water discharge, subsidence rates), see Fig. 27.8. The data set used is a subset of the data collected during

Fig. 27.8 Flume experiment of a delta with low Froude number performed by John Martin, Ben Sheets, Chris Paola and Michael Kelberer. Image source https://www.esci.umn.edu/orgs/seds/Sedi_Research.htm



an experiment in the Tulane Delta Basin, conducted in 2010 (Wang et al. 2011). Basin dimensions were 4.2 m long, 2.8 m wide and 0.65 m deep. The sediment consisted of a mix of 70% quartz sand and 30% anthracite coal sand. These experiments are used for a variety of reasons. One of them is to study the relationship between the surface processes and the subsurface deposition. An intriguing aspect of these experiments is that much of the natural variability is not due to forcing (e.g. uplift, changing sediment source), but due to the internal dynamics of the system itself, i.e. it is autogenic. In fact, it is not known if the autogenic behavior of natural channels is chaotic (Lanzoni and Seminara 2006), meaning one cannot predict with certainty the detailed configuration of even a single meandering channel very far into the future. This then has an immediate impact on uncertainty in the subsurface in the sense that configuration of deposits in the subsurface cannot be predicted with certainty away from wells. The experiment therefore investigates uncertainty related to the dynamics of the system, our lack of physical understanding (and not some parameter uncertainty or observational error). All this is a bit unnerving, since this very fundamental uncertainty is *never* included in any subsurface UQ. At best, one employs a Gaussian prior, or some geometric prior extracted from observation databases, as discussed above. The fundamental questions are:

1. Can we use these experiments to construct a realistic prior, capturing uncertainty related to the physical processes of the system?
2. Can a statistical prior model represent (mimic) such variability?

To address these questions and provide some insight (not an answer quite yet!), we run the experiment under constant forcing for long enough to provide many different realizations of the autogenic variability—a situation that would be practically impossible to find in the field. The autogenic variability in these systems is due to temporal and spatial variability in the feedback between flow and sediment transport, weaving the internal fabric of the final subsurface system.

Under fixed boundary conditions, the observed variability in deposition is therefore the result of only the autogenic (intrinsic) variability in the transport system. The data-set we use here is based on a set of 136 time-lapse overhead photographs that capture the dynamics of flow over the delta approximately every minute. Figure 27.9 shows representative images from this database. This set of images represents a little more than 2 h of experimental run time. Figure 27.9b shows the binary (wet-dry) images for the same set, which will be used in the investigation.

The availability of a large reference set of images of the sedimentary system enables testing any statistical prior by allowing a comparison of the variability of the resulting realizations, since all possible configurations of the system are known. In addition, the physics are naturally contained in the experiment (photographs are the result of the physical depositional processes). A final benefit is that a physical analysis of the prior model can be performed, which aids in understanding what depositional patterns should be in the prior for more sophisticated cases.

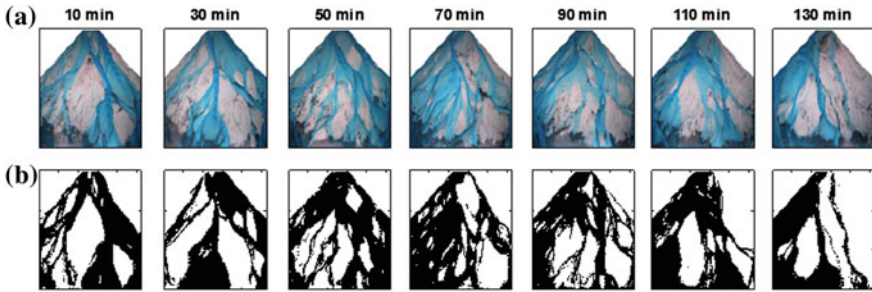


Fig. 27.9 Examples of a few photographic images of the flume experiment for different time periods. Flow is from top to bottom. **a** Photographs of the experiments. The blue pixels indicate locations where flow moves over the surface. The black sediment is coal which is the mobile fraction of the sediment mixture, and the tan sediment is sand. **b** Binary representation of the photographs. Black represents wet (flow) pixels, white represents dry (no flow) pixels

Reproducing Physical Variability with Statistical Models

In this study we employ a geostatistical method termed multiple-point geostatistics. MPS methods have grown popular in the last decade due to their ability to introduce geological realism in modeling via the training image (Mariethoz and Caers 2014). Similar to any geostatistics procedure, MPS allows for the construction of a set of stochastic realizations of the subsurface. Training images, along with trends (usually modeled using probability maps or auxiliary variables) constitute the prior model as defined in the traditional Bayesian framework. The choice of the initial set of training images has a large influence on the stated uncertainty, and hence a careful selection must be done to avoid artificially reducing uncertainty from the start.

It is unlikely that all possible naturally-occurring patterns can be contained in one single training image within the MPS framework (although this is still the norm; similarly, it is the norm to choose for a multi-Gaussian model by default). To represent realistic uncertainty realizations should be generated from multiple TIs. The set of all these realizations then constitutes a wide prior uncertainty model. The choice of the TIs brings a new set of questions: how many training images should one use, and which ones should be selected? Ideally, the TIs should be generated in such a way that natural variability of the system under study is represented (fluvial, deltaic, turbidite, etc.), hence all natural patterns are covered in the possibly infinite set of geostatistical realizations. Scheidt et al. (2016) use methods of computer vision to select a set of representative TIs. One such computer vision method evaluates a rate of change between images in time, and the training images are selected in periods of relative temporal pattern stability (see Fig. 27.10).

The training image set shown in Fig. 27.10 displays patterns consistent with previous physical interpretations of the fundamental modes of this type of delta system: a highly channelized, incisional mode; a poorly channelized, depositional mode; and an intermediate mode. This suggests that some clues to the selection of

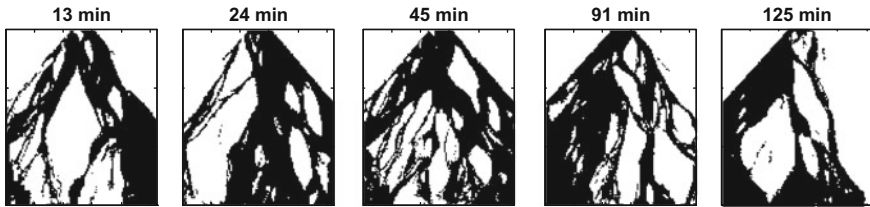


Fig. 27.10 Selected images by clustering based on the modified Hausdorff distance. The value at the top of the image represents the time in minutes of the experiment

appropriate training images lie in the physical properties of the images from the experiment.

With a set of training images available, multiple geostatistical realization per each training image can be generated (basically a hierarchical model of realizations). These realizations can now be compared with the natural variability generated in the laboratory experiments, to verify whether such set of realizations can in any way reproduce natural variability. Scheidt et al. (2016) calculate the Modified Hausdorff Distance (MHD, a distance used in image analysis), between any two geostatistical realization and also between any two overhead shots. A QQ-plot of the distribution of the MHD between all the binary snapshots of the experiment and the MPS models is shown in Fig. 27.11a, showing similarity in distribution.

The result is encouraging but also emphasizes a mostly ignored question of what a complete geological prior entails, that the default choices (one training image, one Boolean model, one multi-Gaussian distribution) make very little sense when dealing with realistic subsurface heterogeneity. The broader question remains as to how such a prior should be constructed from physical principles and how statistical models, such as geostatistics should be employed in Bayesianism when applied to

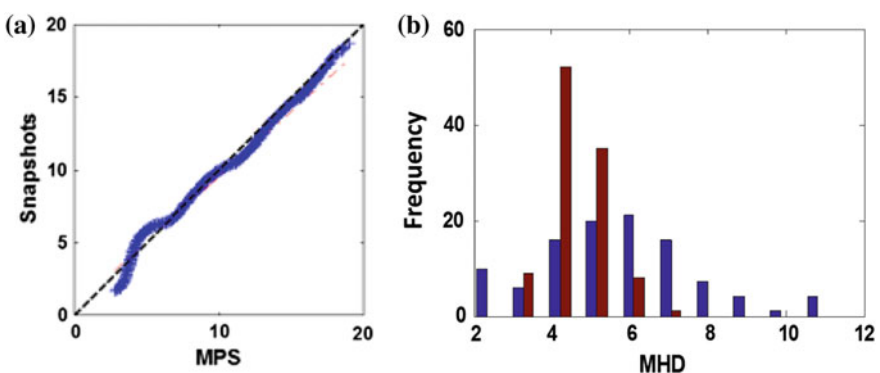


Fig. 27.11 **a** QQ-plot of the MHD distances between the 136 images from the experiment and 136 images generated using DS. **b** Comparison of the variability, as defined by MHD, between generated realizations per each training image (red) and the images from the experiment (blue) closest (in MHD) to the selected TI

geological systems. This fundamental question remains unresolved and certainly under-researched.

Field Application

The above flume experiments have helped in the understanding of the nature of a geological prior, at least for deltaic type deposits. Knowledge accumulated from these experiments will create scientific understanding on the fundamental processes involved in the genesis of these deposits and thereby understand better the range of variability of the generated stratigraphic sequences.

It is unlikely, however, that laboratory experiments will be of direct use in actual applications, since they take considerable time and effort to set them up. In addition, there is a question of how these scale to the real world. It is more likely in the near future that computer models, built from such understanding, will be used in actual practice. Various such computer models exist for depositional systems (process-based, process-mimicking, etc.).

We consider here one such computer model, FLUMY (Cojan et al. 2005), which is used to model meandering channels, see Fig. 27.12. FLUMY uses a combination of physical and stochastic process models to create realistic geometries. It is not an object-based model, which would focus on the end result, but it actually creates the depositional system. The input parameters are therefore a combination of physical parameters as well as geometrical parameters describing the evolution of the deposition.

Consider a simple application to an actual reservoir system (Courtesy of ENI). Based on geological understanding generated from well data and seismic, modelers are asked to input the following FLUMY parameters: channel width, depth and sinuosity (geometric), and two aggradation parameters: (1) decrease of the alluvium thickness away from the channel, and, (2) maximum thickness deposited on levees during an overbank flood. More parameters exist but these are kept fixed for this simple application.

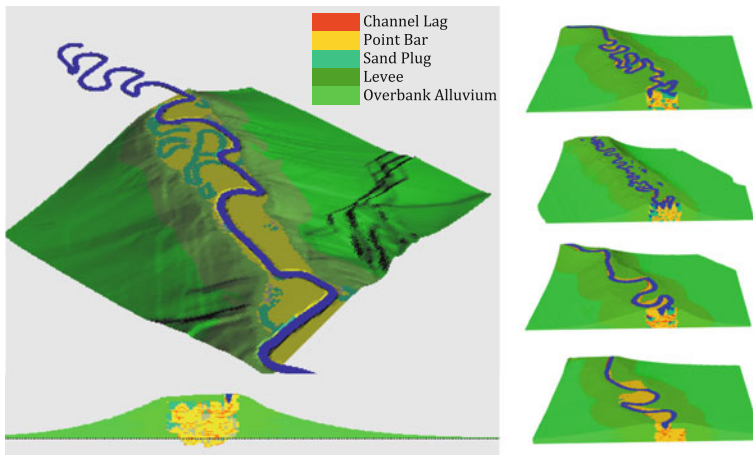


Fig. 27.12 Example of a FLUMY model with several realizations of the prior generated from FLUMY with uncertain input parameters

The prior belief now consists of (1) assuming the FLUMY model as a hypothesis that describes variability in the depositional system and (2) prior distributions of the five parameters. After generating 1000 s of FLUMY models (see Fig. 27.12), we can run the same analysis as done for the flume experiment to extract modes in the system that can be used as training images for further geostatistical modeling.

27.10 Summary

Eventually philosophical principles will need to be translated into workable practices, ultimately into data acquisition, computer codes, and actual decisions. A summary of some important observations and perhaps also personal opinion based on this chapter are:

- *Data acquisition, modeling and predictions “collaborate”*; going from data to models to prediction ignores the important interactions that take place between these components. Models can be used, prior to actual data acquisition to understand what role they will play in modeling and ultimately into the decision-making process. The often classical route of first gathering data, then creating models, may be completely inefficient if the data has no or little impact on any decision. This should be studied beforehand and hence requires building models of the data, not just of the subsurface.
- *Prior model generation is critical to Bayesian approaches* in the subsurface and statistical principles of indifference are very crude approximations of realistic geological priors. Uniform and multi-Gaussian distributions have been clearly falsified with many case studies (Gómez-Hernández and Wen 1998; Feyen and Caers 2006; Zinn and Harvey 2003). They may lead to completely erroneous predictions when used in subsurface applications. One can draw an analogy here with Newtonian physics: it has been falsified but it is still around, meaning it can be useful to make many predictions. The same goes with multi-Gaussian type assumptions. Such choices are logical for an “agent” that has limited knowledge and hence (rightfully) uses the principal of indifference. More informed agents will however use more realistic prior distribution. The point therefore is to use more informed agents (geologists) into the quantification of prior. The use of such agents would make use of the vast geological (physical) understanding that has been generated over many decades.
- *Falsification or prior*. It now seems logical to propose workflows of UQ that have both the induction and deduction flavors. Falsification should be part of any a priori application of Bayesianism, and also on the posterior results. Such approaches will rely on forms of sensitivity analysis as well as developing geological scenarios that are tested against data. The point here is not to state rigorous probabilities on scenarios but to eliminate scenarios from the pool of possibilities because they have been falsified. The most important aspect of geological priors are not the probabilities given to scenarios but the generation

of a suitable set of representative scenarios to represent the geological process taking place. This was illustrated in the flume experiment study.

- *Falsification of the posterior.* The posterior is the result of the prior model choice, the likelihood model choice and all of the auxiliary assumptions and choices made (dimension reduction method, sampler choices, convergence assessment etc. ...). Acceptance of the posterior “as is” would follow the pure inductionist approach. Just as the prior, it would be good practice to attempt to falsify the posterior. This can be done in several ways, usual using hypothetico-deductive analysis, such as the significance tests introduced in this chapter.

References

- Ariew R (1984) The duhem thesis. *Br J Philos Sci* 35(4):313–325
- Bayes M. & Price, M., 1763. An essay towards solving a problem in the Doctrine of chances. By the Late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philos Trans R Soc Lond* 53(0):370–418
- Berger JO (2003) Could Fisher, Jeffreys and Neyman have agreed on testing? *Stat Sci* 18(1):1–32
- Beres M, Green A, Huggenberger P, Horstmeyer H (1995) Mapping the architecture of glaciofluvial sediments with three-dimensional georadar. *Geology* 23(12):1087–1090
- Bond CE et al (2007) What do you think this is? “Conceptual uncertainty” in geoscience interpretation. *GSA Today* 17(11):4–10
- Bordley RF (1982) A multiplicative formula for aggregating probability assessments. *Manage Sci* 28(10):1137–1148
- Bristow CS, Jol HM (2003) An introduction to ground penetrating radar (GPR) in sediments. *Geol Soc* 211(1):1–7. London, Special Publications
- Chalmers AF (1999) What is this thing called science?, 3rd edn. Metascience, London
- Clemen RT, Winkler RL (2007) Aggregating probability distributions. *Adv Decis Anal* 120(919): 154–176
- Cojan I et al (2005) Process-based reservoir modelling in the example of meandering channel. In: *Geostatistics Banff 2004*. Springer, Netherlands, pp 611–619
- Colombera L et al (2012) A database approach for constraining stochastic simulations of the sedimentary heterogeneity of fluvial reservoirs. *AAPG Bull* 96(11):2143–2166
- Cox RT (1946) Probability, frequency and reasonable expectation. *Am J Phys* 14(1):1
- Dubois D, Prade H (1990) The logical view of conditioning and its application to possibility and evidence theories. *Int J Approx Reason* 4(1):23–46
- Dupré MJ, Tiplery FJ (2009) New axioms for rigorous Bayesian probability. *Bayesian Anal* 4(3): 599–606
- Earman J (1992) Bayes or bust: a critical examination of Bayesian confirmation theory
- Eschard R, Doligez B (2000) Using quantitative outcrop databases as a guide for geological reservoir modelling. In: *Geostatistics Rio 2000*. Springer, pp 7–17
- Fallis A (2013) Fisher, Neyman and the creation of classical statistics
- Feller W (2008) An introduction to probability theory and its applications. 2nd edn, vol 2, p xxiv-669
- Feyn L, Caers J (2006) Quantifying geological uncertainty for flow and transport modeling in multi-modal heterogeneous formations. *Adv Water Resour* 29(6):912–929
- Feyerabend, P., 1993. Against method
- Fine A (1973) Probability and the interpretation of quantum mechanics. *Br J Philos Sci* 24(1):1–37

- de Finetti B (1974) The value of studying subjective evaluations of probability. In: *The concept of probability in psychological experiments*, pp 1–14
- de Finetti B, Machí A, Smith A (1975) *Theory of probability: a critical introductory treatment*
- de Finetti B (1995) The logic of probability. *Philos Stud* 77(1):181–190
- Fisher R (1925) *Statistical methods for research workers*. <http://psychclassics.yorku.ca/Fisher/Methods>
- Fisher R, Fisher R (1915) Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 10(4):507–521
- Gabrovsek F, Dreybrodt W (2010) Karstification in unconfined limestone aquifers by mixing of phreatic water with surface water from a local input: a model. *J Hydrol* 386(1–4):130–141
- Gelman A et al (2004) *Bayesian data analysis*
- Gelman A (2008) Objections to Bayesian statistics Rejoinder. *Bayesian Anal* 3(3):467–477
- Gelman A, Shalizi CR (2013) Philosophy and the practice of Bayesian statistics. *Br J Math Stat Psychol* 66(1996):8–38
- Gibling MR (2006) Width and thickness of fluvial channel bodies and valley fills in the geological record: a literature compilation and classification. *J Sediment Res* 76(5):731–770
- Gnedenko BV, Aleksandr I, Khinchin A (1962) *An elementary introduction to the theory of probability*
- Gómez-Hernández JJ, Wen X-H (1998) To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology. *Adv Water Resour* 21(1):47–61
- Halpern J (2011) A counter example to theorems of Cox and Fine, vol 10, pp 67–85. [arXiv:1105.5450](https://arxiv.org/abs/1105.5450)
- Halpern JY (1995) A logical approach to reasoning about uncertainty: a tutorial. In: *Discourse, interaction, and communication*, pp 141–155
- Hand DJ, Walley P (1993) Statistical reasoning with imprecise probabilities. *Appl Stat* 42(1):237
- Hanson (1958) Patterns of discovery. *Philos Rev* 69(2):247–252
- Henriksen HJ et al (2012) Use of Bayesian belief networks for dealing with ambiguity in integrated groundwater management. *Integr Environ Assess Manag* 8(3):430–444
- Höhle U (2003) *Metamathematics of fuzzy logic*
- Howson C (1991) The “old evidence” problem. *Br J Philos Sci* 42(4):547–555
- Howson C, Urbach P, Gower B (1993) *Scientific reasoning: the Bayesian approach*
- Hume D (2000) *A treatise of human nature*. *Treatise Hum Nat* 26(1739):626
- Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev* 106(4):181–218
- Jaynes ET (2003) Probability theory: the logic of science. *Math Intell* 27(2):83
- Jenei S, Fodor JC (1998) On continuous triangular norms. *Fuzzy Sets Syst* 100(1–3):273–282
- Journal AG (2002) Combining knowledge from diverse sources: an alternative to traditional data independence hypotheses. *Math Geol* 34(5):573–596
- Jung A, Aigner T (2012) Carbonate geobodies: hierarchical classification and database—a new workflow for 3D reservoir modelling. *J Pet Geol* 35:49–65
- Kenter JAM, Harris PM (2006) Web-based Outcrop Digital Analog Database (WODAD): archiving carbonate platform margins. In: *AAPG international conference (November 2006)*. p 5–8
- Kiessling W, Flügel E (2002) Paleoreefs—a database on Phanerozoic reefs
- Klement EP, Mesiar R, Pap E (2004) Triangular norms. Position paper I: basic analytical and algebraic properties. In: *Fuzzy sets and systems*, pp 5–26
- Klir GJ (1994) On the alleged superiority of probabilistic representation of uncertainty. *IEEE Trans Fuzzy Syst* 2(1):27–31
- Kolmogoroff A (1950) *Foundations of the theory of probability*
- Koltermann C, Gorelick S (1992) Paleoclimatic signature in terrestrial flood deposits. *Science* 256(5065):1775–1782
- Kuhn TS (1996) *The structure of scientific revolution*

- Lantuéjoul C (2013) Geostatistical simulation: models and algorithms. Springer Science & Business Media
- Lanzoni S, Seminara G (2006) On the nature of meander instability. *J Geophys Res Earth Surf* 111(4)
- Lindgren B (1976) Statistical theory. MacMillan, New York
- Loginov VJ (1966) Probability treatment of Zadeh membership function and their use in pattern recognition. *Eng Cybern*, 68–69
- Mariethoz G, Caers J (2014) Multiple-point geostatistics: stochastic modeling with training images. Wiley Blackwell, Hoboken
- Martinius AW, Naess A (2005) Uncertainty analysis of fluvial outcrop data for stochastic reservoir modelling. *Pet Geosci* 11(3):203–214
- Mayo D (1996) Error and the growth of experimental knowledge. Chicago University Press, Chicago
- Neyman J, Pearson ES (1967) Joint statistical papers
- Neyman J, Pearson ES (1933) On the problem of the most efficient tests of statistical hypotheses. *Philos Trans R Soc A Math Phys Eng Sci* 231(694–706):289–337
- Nicholas AP et al (2013) Numerical simulation of bar and island morphodynamics in anabranching megarivers. *J Geophys Res Earth Surf* 118(4):2019–2044
- Pearson K, Fisher RA, Inman HF (1994) Karl Pearson and R. A. Fisher on statistical tests: a 1935 exchange from nature. *Am Stat* 48(1), 2–11
- Popper ER (1959) The logic of scientific discovery. Hutchinson, London
- Pyrce MJ, Boisvert JB, Deutsch CV (2008) A library of training images for fluvial and deepwater reservoirs and associated code. *Comput Geosci* 34:542–60
- Rao CR (1992) R. A. Fisher: the founder of modern statistics. *Stat Sci* 7(1):34–48
- Refsgaard JC et al (2012) Review of strategies for handling geological uncertainty in groundwater flow and transport modeling. *Adv Water Resour* 36:36–50
- Rings J et al (2012) Bayesian model averaging using particle filtering and Gaussian mixture modeling: theory, concepts, and simulation experiments. *Water Resour Res* 48(5)
- Sambrook Smith GH, Ashworth PJ, Best JL, Woodward J, Simpson CJ (2006) The sedimentology and alluvial architecture of the sandy braided South Saskatchewan River, Canada. *Sedimentology* 53(2):413–434
- Scheidt C et al (2016) Quantifying natural delta variability using a multiple-point geostatistics prior uncertainty model. *J Geophys Res Earth Surf*
- Shackle GLS (1962) The stages of economic growth. *Polit Stud* 10(1):65–67
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(July 1928), pp 379–423
- Tsai FTC, Elshall AS (2013) Hierarchical Bayesian model averaging for hydrostratigraphic modeling: uncertainty segregation and comparative evaluation. *Water Resour Res* 49(9):5520–5536
- Valle A, Pham A, Hsueh PT, Faulhaber J (1993) Development and use of a finely gridded window model for a reservoir containing super permeable channels. In: Middle east oil show. Society of Petroleum Engineers
- Wang P (2004) The limitation of Bayesianism. *Artif Intell* 158(1):97–106
- Wang Y, Straub KM, Hajek EA (2011) Scale-dependent compensational stacking: an estimate of autogenic time scales in channelized sedimentary deposits. *Geology* 39(9):811–814
- Zadeh L (1965) Fuzzy sets. *Inf Control* 8:338–353
- Zadeh LA (1975) Fuzzy logic and approximate reasoning. *D Reidel Publ Co* 30(i):407–428
- Zadeh LA (1978) Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst* 1(1):3–28

- Zadeh LA (2004) Fuzzy logic systems: origin, concepts, and trends. *Science*, pp 16–18
- Zinn B, Harvey CF (2003) When good statistical models of aquifer heterogeneity go bad: a comparison of flow, dispersion, and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields. *Water Resour Res* 39(3):1–19

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

