Applied Geochemistry 27 (2012) 1899–1919

Contents lists available at SciVerse ScienceDirect

Applied Geochemistry

journal homepage: www.elsevier.com/locate/apgeochem

Review

Models, validation, and applied geochemistry: Issues in science, communication, and philosophy

D. Kirk Nordstrom

U.S. Geological Survey, 3215 Marine St., Boulder, CO 80303, USA

ARTICLE INFO

Article history:
Available online 17 July 2012

ABSTRACT

Models have become so fashionable that many scientists and engineers cannot imagine working without them. The predominant use of computer codes to execute model calculations has blurred the distinction between code and model. The recent controversy regarding model validation has brought into question what we mean by a 'model' and by 'validation.' It has become apparent that the usual meaning of validation may be common in engineering practice and seems useful in legal practice but it is contrary to scientific practice and brings into question our understanding of science and how it can best be applied to such problems as hazardous waste characterization, remediation, and aqueous geochemistry in general. This review summarizes arguments against using the phrase model validation and examines efforts to validate models for high-level radioactive waste management and for permitting and monitoring open-pit mines. Part of the controversy comes from a misunderstanding of 'prediction' and the need to distinguish logical from temporal prediction. Another problem stems from the difference in the engineering approach contrasted with the scientific approach. The reductionist influence on the way we approach environmental investigations also limits our ability to model the interconnected nature of reality. Guidelines are proposed to improve our perceptions and proper utilization of models. Use of the word 'validation' is strongly discouraged when discussing model reliability.

Published by Elsevier Ltd.

Contents

E-mail address: dkn@usgs.gov

0883-2927/$ - see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.apgeochem.2012.07.007

1. Introduction

"Is there any knowledge in the world which is so certain that no reasonable man could doubt it?" These words were written by Bertrand Russell in the opening sentence of *The Problems of Philosophy* (Russell, 1912), a short book of 100 pages. He continues with "This question, which at first sight might not seem difficult, is really one of the most difficult that can be asked." Philosophers, scientists, and lawyers have struggled with this question since the beginnings of recorded history. The question continued to interest Russell, who did change some of his opinions over the years, and led him to publish *Human Knowledge* (Russell, 1948), some 36 years later and 438 pages longer. He arrived at 5 basic postulates (quasi-permanence, separable causal lines, spatio-temporal continuity, common causal origin, and analogy) which he concluded are necessary to validate scientific knowledge. Did that end the discussion? No, but it did provide interesting and useful insight. Russell's question lies at the core of the ongoing controversy regarding models, their validation in the geosciences and other field-based sciences, and what we really know and do not know.

Models are one of the principal tools of modern science and engineering, e.g. global climate models, groundwater models, ecological models, geochemical models, rainfall–runoff models, planetary models, radionuclide-transport models, reactive-transport models, etc.

"Scientists spend a great deal of time building, testing, comparing and revising models, and much journal space is dedicated to introducing, applying and interpreting these valuable tools" (Frigg and Hartmann, 2009). There seems to be a model for every need and purpose. Inevitably the question is asked: How good is your model? Most scientists and engineers prefer to rephrase the question to: How reliable is your model for the intended purpose? Both questions are vague enough that one could argue either way, positively or negatively and it begs the question, how do we judge correctness? The discussion is inextricably tied to the meaning of words and phrases. Hence, it involves semantics, linguistics, ontology, epistemology, in addition to science, the philosophy of science, and engineering. Indeed, most of philosophy is concerned with meaning and how best to express whatever subject is under investigation. Definitions become essential even though whatever definitions one chooses are likely to be criticized and found wanting in some respect (see examples in the Appendix A).

Attempts to answer the question of model reliability in both science and engineering practice have led to efforts to validate models and to numerous publications on model validation. Some of these were misguided attempts to claim certainty where little certainty existed and led to more papers for further clarification. This debate involved primarily ecologists, hydrogeologists, and engineers working in an environmental context, often related to hazardous waste disposal. Few geochemists seem to have contributed to this discussion. Having seen first-hand how easily scientists and engineers get drawn into believing the 'truth' of their models, including geochemical models, I was prompted to write this paper. Receiving the International Ingerson Lecture Award in 2010 provided me the opportunity to organize my thoughts on this subject for this review.

How well or how little we understand science and the limits of its application determines how we perceive our modeling efforts. Model reliability and validation is more than just semantic arguments. They continue to be an evolving discussion of how we see the place of science in our world and how science can contribute to important policies and decision making.

2. Science and the logic of induction and deduction

Science is the application of common sense to an uncommon degree. As Einstein (1936) once said, "The whole of science is nothing more than a refinement of everyday thinking." And more specifically, "Science is the attempt to make the chaotic diversity of our sense-experience correspond to a logically uniform system of thought" (Einstein, 1940). Science is the discovery of laws, theories, and hypotheses that interpret empirical observations through the use of logic. Conversely, hypotheses, theories, and laws are used, consciously or unconsciously, to guide and propose quantitative empirical observations. Observation and theory are inextricably intertwined in the scientific method and should not be thought of as separate endeavors.

Science searches for explanations of natural phenomena and tries to put these explanations into comprehensive generalizations. It is also about communication of these generalizations. Science benefits no one if it is not effectively communicated. The controversy surrounding 'model validation' arose because of misunderstandings about what science is, how to communicate science effectively, and how science can effectively help the regulatory process and the resolution of major environmental issues. The model validation controversy, unfortunately, received even more attention when it entered the courtrooms (Bair, 1994, 2001). It behooves every scientist to at least be familiar with the main points of debate about model validation.

This controversy is not new. It can be considered a continuation of philosophical discourses that began in ancient civilizations and developed further in western civilization, especially in the 17th and 18th centuries. Aristotle developed logic and the concepts of deduction and induction. The inductive–deductive approach became a backbone of the scientific method. But it also left us with a conundrum. Deduction was understandable – given two premises (a major and a minor premise), a conclusion is unequivocally reached. If we say, for example, "all men are mortal and Socrates is a man, then Socrates is mortal," the syllogism seems simple enough. But how do you know the premises are correct? The premises usually include a general inductive statement that is assumed to be true. Induction is an inference from individual observations to a general statement or conclusion. Of course, if one of the premises is false, the conclusion might be false. If we say all substances that dissolve in water exist as ions in aqueous solution and CH_4 dissolves in water, it is deduced that CH_4 exists as an ion in aqueous solution. However, this is recognized as being incorrect because the major premise is incorrect. Not all substances that dissolve in water exist as ions in solution. How is it that CH_4 can dissolve in water, a polar solvent, without ionizing? There are two other forces that can account for the solubility of CH_4 in water, induced polarization (polarization of CH_4 induced by water molecules) and Van der Waals forces (weak attractive forces between molecules, for example between the electrons in one molecule with the nuclei in another). Because these are weak forces, the solubility of CH_4 in water is low, unless high pressure is applied. High pressures often exist in deep groundwaters so that the solubility of CH_4 in a groundwater at a kilometer depth would be considerably larger than that near the Earth's surface. The problem here is that we can be led astray by an incorrect major (or minor) premise.

A premise is a general statement based on numerous observations that is inferred by induction. We might not have all the necessary information in a premise to make deductive conclusions. How do we know if a premise is correct? It is deductively 'correct' or 'valid,' if and only if, there are no exceptions to the statement. Because it is always possible for there to be an exception not yet known, general statements or inferences cannot be valid in a deductive sense. "Inductive inferences are never deductively valid" (Boyd, 1991; of course, those who love the self-reference game might ask if this statement is deductive or inductive).

The philosopher who argued most effectively that logic cannot prove (or even claim highly probable) inductive inference is Hume (1739). "So Hume noticed a very general, and very curious, point about any reasoning concerning the future; namely all such reasoning rests on an assumption that cannot be logically justified. And the more general implication is difficult to avoid: if all reasoning about the future rests on a logically unjustified assumption, then our conclusions based on such reasoning, that is, our conclusions about what is likely to happen in the future, are equally logically unjustified" (DeWitt, 2010). As Russell (1945) summarized "What these arguments [of Hume] prove – and I do not think the proof can be controverted – is that induction is an independent logical principle, incapable of being inferred from experience or from other logical principles, and that without this principle science is impossible." Another apt summary came from Gorham (2009) "Justifying induction inductively is as hopeless as relying on a hunch to vindicate the power of intuition."

Hume crowned the argument of the 'empiricists' (a line of philosophers who argued that all knowledge begins with observation via our senses) with his observation that knowledge can only be gained by experience; it is a posteriori. Traditionally, the opposing argument comes from the 'rationalists' or 'idealists' who say that all knowledge begins with reason; that knowledge is a priori. This unfortunate division is artificial, an outgrowth of our attempts to categorize everything and condense types of knowledge into a simple framework. Bronowski (1951) had a much more meaningful description when he said "In order to act in a scientific manner, in order to act in a human manner at all, two things are necessary: fact and thought. Science does not consist only of finding the facts; nor is it enough only to think, however rationally. The processes of science are characteristic of human action in that they move by the union of empirical fact and rational thought, in a way which cannot be disentangled."

Instead of stating that laws, theories, and principles of science are true – an opinion that is not logically justifiable – we often say that they are approximations which become more certain the more they are tested and their consequences shown to correspond with reality. Such testing has often been called validation when the results compare well. But I am jumping ahead of myself. This correspondence principle is one of the major tenets of 'truth' and is directly applicable to the scientific method. Unfortunately, the correspondence theory of truth has several flaws because of its ambiguities. What is the reality that thoughts or beliefs correspond to? Does this reality only refer to physically-based, sense-impression objects or can it include more general classes of objects and their relationships? Can we recognize reality? Can we measure or objectify reality sufficiently to make a comparison? What does 'correspond' mean? If it does not mean exact agreement then what criteria determine a meaningful comparison? Is the subjectivism implied by the last question unavoidable? The Ptolemaic theory of an earth-centered solar system corresponded to observations known at the time but ultimately found to be 'incorrect' a century later. The same might be true today for models of radionuclide migration from a leaking high-level radioactive waste repository. The difference being that instead of waiting for a century to get confirmation, we must wait for many millennia. In contrast, a geochemical model can be used to calculate the saturation state of a groundwater with respect to various minerals and compared to our assumption that some minerals should have reached solubility equilibrium after a few years residence time. For example, shallow groundwaters in a limestone aquifer were shown to have reached equilibrium with respect to calcite (Langmuir, 1971) demonstrating that for that type of aquifer the equilibrium calculation 'makes sense,' or 'is confirmed.' Some people would say "Isn't that an example of validating a geochemical model?" To that question I would reply: It is indeed a test of the model and the corroboration is helpful but why do you need to call it model validation which has a different meaning? Do you validate a model only once? If you have to re-validate it many times over then you do not really mean validation. The geochemical model that Langmuir (1971) used was actually inadequate in many ways but for quantitatively determining the solubility saturation state of calcite in this shallow, dilute Pennsylvanian aquifer, it worked quite well enough. For other situations it may not work at all.

These concerns about the correspondence theory led to alternative theories of truth such as the coherence theory. 'Coherence,' in this context, refers to the consistency of a belief or theory with the body of existing knowledge. The greater its consistency, the more likely it is to be true. "Even though parts of the edifice may be found to be rotten, the coherence of a body of scientific truths accounts for its stability over long periods of time. 'Scientific knowledge,' John Ziman rightly observes, 'eventually becomes a web or network of laws, models, theoretical principles, formulae, hypotheses, interpretations, etc., which are so closely woven together that the whole assembly is much stronger than any single element'" (Newton, 1997). Unfortunately, there are examples in the history of science where consistency has not been upheld as a trustworthy guide. New scientific discoveries have been known to be inconsistent with some of the established body of knowledge. Friedrich Wöhler synthesized an organic compound, urea, from inorganic compounds (Wöhler, 1828) and this result was contrary to the prevailing Vital Force Theory, or vitalism, at the time which held that organic compounds could only be made from organic compounds and living organisms.

Because both the 'correspondence' and the 'coherence' theories cannot, by themselves, be a criterion for the truth of scientific statements, other theories arose of which the most important would be pragmatism. Instead of thinking of a proposition as true or false, we should think of them as useful or useless. Both the ion-association model and Pitzer's specific-ion interaction model have been found to be useful for interpreting the behavior of aqueous solutions of electrolytes within different ranges of conditions, whereas the Debye-Hückel limiting law (DHLL) is fairly useless. The problem here is that the DHLL is part of the other two models and must be approached in the limit of infinite dilution. Hence, it is meaningful as a component within a larger context but not useful by itself.

You may have noticed that I have been making reference to what a word, phrase, theory, statement or proposition means. Often the meaning is more important than the truth or falsehood of a statement, although one would think they would go hand-in-hand. When Kekulé dreamt of a snake biting its own tail, that gave him the idea of the shape of the benzene ring (see Benfey, 1958), there was no literal truth to the snake but there was insightful meaning. Historically there have been two quantum mechanical approaches to interpreting the chemical bond, valence-bond theory and molecular-orbital theory. Both can explain many aspects of chemical structure and both are meaningful and useful. Each has certain advantages and disadvantages but it would not make much sense to say that one is closer to the truth than the other.

Another problem with 'truth' is that it comes in many different forms which can lead to misunderstandings. 'Truth' depends on the context and whether the truth refers to objective knowledge or subjective knowledge. There is the literal truth of the existence of physical reality. I can show someone a rock, explain that it has certain physical and mineralogical and chemical properties that can all be checked to demonstrate that it is, in truth, a rock of a certain type. I can also explain to someone what a unicorn is by describing its characteristics, but I cannot show it to anyone because it is a mythical creature that might have existed but there is no physical evidence for it. The objective knowledge for the unicorn does not exist. Once having described the unicorn, however, knowledge of a unicorn now exists. It can be found in paintings, novels, and emblems. It might be classified as subjective knowledge. It is a symbol for chastity and purity which gives it meaning. We know what these virtues are and, accordingly, we have knowledge of the unicorn. Kierkegaard (1941) expressed subjective knowledge as the relationship of a person to something. He has said (see Appendix A) that if the relationship is true then the individual is in the truth even if he is relating to something that is not objectively true. Much confusion and hostility has been caused by substituting literal truth for figurative truth and the lack of recognition for the many forms of truth.

We use the words 'true' and 'truth' in everyday language and the meaning is usually clear but in the application of science to such problems as hazardous waste disposal where models and their reliability are involved, these words no longer have a clear meaning in such a general form. At least the meaning is not clear without numerous caveats that are often site-specific. We are often guilty of making hasty generalizations about the truth or validity of our models.

This brief synopsis is meant to remind us of the philosophical underpinnings of science, especially the framework that relates to scientific modeling and the model validation controversy. 'Scientific truth' is a difficult and challenging concept to define. Every attempt (correspondence, coherence, pragmatism, meaningfulness) has captured an important element, but taken separately, each attempt has serious shortcomings. Logic is an essential element of good science, but it fails to justify induction, another essential element. Other well-established criteria for good science are reproducibility, simplicity (Occam's razor), and peer review, although none of them are guarantees of 'correctness.' They are simply helpful guidelines for good science that lead to better approximations of our understanding of natural phenomena.

3. Models and modeling

Much to my surprise, I have had some rather strange responses from colleagues when I have asked or suggested what might constitute a 'model.' I have heard "We know what a model is; let's move on," and "Qualitative models are better than any quantitative models." I have been told that there is no difference between a model and a computer code. I have heard that someone's model may not agree with my definition of a model but it is valid nonetheless. And I have had a statistician tell me that a statistical model is a scientific model. These comments come from some scientists and engineers who may not have given much thought to what science is and how it works.

'Model' is actually rather difficult to define, but perhaps the simplest and most concise definition is that it is a simplification of reality. This definition is found in many places and was felt appropriate for a National Academy's assessment of models for regulatory decision making (National Research Council, 2007). This assessment was consequently adopted by the U.S. Environmental Protection Agency (USEPA) in their guidebook on environmental model development, evaluation, and application (USEPA, 2009). Another common definition is that it is a representation of reality (e.g. Konikow and Bredehoeft, 1992). Other definitions are that it is an idealization of reality, a replica, a picture of how something works. I wish to be clear, whatever definition or description that I come up with is not 'mine' but rather a synthesis or repetition of what many others have described already. To make my point, I have compiled several definitions and descriptions in the Appendix A as examples. I do not agree with all of these definitions, but many of them have captured the essence of a scientific model.

Briefly I shall address my colleagues' other comments. Sometimes a qualitative model is better than a quantitative model, especially when the quantitative model is flawed. But science strives to be as quantitative as possible. When you can quantitate a qualitative model, it usually signifies an improvement in our understanding.

Equating a computer code with a model is simply misuse of language. A code is a kind of conversion algorithm that functions to convert one language into another. A computer code converts a written language that we can read (logic, sentences, or mathematics) into a language that a computer can read. A model can be converted into a computer code or not. There were plenty of scientific models before the computer was invented. When Linus Pauling discovered a model for the α-helix structure of protein in his room while visiting Oxford University by using paper cutouts and drawing base groups on it, there was no computer code involved (Eisenberg, 2003). When sea-floor spreading and continental drift were discovered these were conceptual models based on observational data which could not be formalized with mathematical equations or computer codes. Conceptual models, physical small-scale models, and experimental models are not computer codes. This difference between a model and a computer code is substantial and is necessary to avoid confusion and maintain clarity in communication. To call a model a computer code is to belittle the importance of the model in scientific research.

A computer code that incorporates a geochemical model is one of several possible tools for interpreting water–rock interactions in low-temperature geochemistry. It is unfortunate that one commonly finds, in the literature, reference to the MINTEQ or the PHREEQC model or the EQ3/6 model when these are not models but computer codes. Some of the models used by these codes are the same so that a different code name does not necessarily mean a different model is being used, but so it might be thought if no distinction is made between model and code; and vice versa, different models can be in the same code. If someone states only that they are using the PHREEQC model, then it is entirely ambiguous as to whether they used the ion-association model or the Pitzer model or whether they used the WATEQ database, the PHREEQC database, or the MINTEQ database. Of course they can (and should) specify these aspects but that does not make PHREEQC a model; it is a computer code.

Another misconception, common among engineers, is to think of models only in terms of mathematical equations. A typical example would be the American Society for Testing Materials (ASTM, 1984) "Standard practice for evaluating environmental fate models of chemicals" document in which a model is defined as "an assembly of concepts in the form of mathematical equations that portray understanding of a natural phenomenon." But, of course, the examples of Pauling and the protein structure (Eisenberg, 2003), plate tectonics, experimental models, physical analogues, and numerous other examples should make it clear that there are other types than just mathematical models. Greenwood

(1989) expressed it well when he said that a model "…is a well-constrained proposition, not necessarily mathematical, that has testable consequences." Indeed, physicists have so pursued the study of the most elementary particles that make up all larger particles that constitute all of matter that they discuss entities that can never be seen, or even explained, without the symbolism of higher mathematics. This state of affairs leads some scientists to question whether something that exists only as a mathematical assemblage has any real existence.

The strength of mathematics is in giving the appearance of certainty, its weakness comes from its inability to provide a consistent provable formal system of statements (Gödel, 1931). This argument of Gödel's became known as the incompleteness theorem. The problem with an inconsistent system is that you can prove anything, so it tends to be useless if you are trying to prove your argument with mathematics (Bronowski, 1978). Now most scientists would not say that mathematics is useless, as Bronowski (1978) points out, "It is the axiomatization, it is the formalization of the system which produces the trouble. Nature is not a gigantic formalizable system." Alfred Tarski (1983) carried this further by showing that there is no complete language of science in his theorem on the indefinability of truth (although it seems clear that Gödel was aware of this theorem also). He argues that truth in a language cannot be defined in itself. "I have told you that Tarski's proof that there is no complete closed scientific language depends on showing that as soon as you introduce the words 'is true' you get paradoxes" (Bronowski, 1978). These were formal arguments but Einstein (1921) foreshadowed these findings earlier with his often quoted statement, "…as far as the propositions of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality." "...there seems to be no unique way of combining a mathematical language and a part of the world" (Gregory, 1990).

By 1992 the ASTM changed their document title to "Standard practice for evaluating mathematical models for the environmental fate of chemicals" and a model (presumably just a mathematical model) was defined as "an assemblage of concepts in the form of equations that represent a mathematical interpretation of a natural phenomenon." Certainly this change reflected an improvement, but by 2001 a further change in thinking occurred at ASTM because the ASTM (1992) document was withdrawn with no replacement.

The most recent International Atomic Energy Agency (IAEA) definition for model is "An analytical representation or quantification of a real system and the ways in which phenomena occur within that system, used to predict or assess the behaviour of the real system under specified (often hypothetical) conditions" (IAEA, 2007). This is a somewhat restricted definition, but it is better than their original definition "In applied mathematics, an analytical or mathematical representation or quantification of a real system and the ways that phenomena occur within that system" (IAEA, 1982).

We often say that a model is supposed to represent reality. I know what is meant by this statement, but I think it is a bit presumptuous. It would be better to say that a model represents our thinking about reality rather than reality itself. This important distinction was driven home by Hughes (1997) when he stated "One major philosophical insight recovered by the romantic view of theories is that the statements of physical theory are not, strictly speaking, statements about the physical world. They are statements about theoretical constructs." "Science, like art, is not a copy of nature but a re-creation of her" (Bronowski, 1956).

It is also incorrect to say that we compare our models with reality. We compare the consequences of our models with independent observations. And those observations are not devoid of concepts and theories, they come with their own conceptual framework. Take the 'concept' of temperature. What is temperature? It is a quantitative measure of the 'hotness' or 'coldness' of objects which is a subjectively communicated sensation of heat transfer. Heat was a concept that was debated for centuries (e.g. the phlogiston theory, the caloric theory, the kinetic theory). It took about as long for experimenters to test different materials and designs that might be used to measure temperature. Standardization required further inventions and experiments. Middleton (1966) mentions 18 non-centesimal temperature scales that could be seen in the Copernicus Museum in Rome. Today there are three different temperature scales in use, the Celsius scale, the Fahrenheit scale, and the Kelvin or absolute scale. The common measurement of temperature is often taken for granted, but theories were required to establish the concept and how best to measure it. Einstein and Infeld (1938) made it clear that we cannot compare our theories with the real world; we can only compare the predictions from our theories with our theory-laden observations of the world.

There is no universally accepted theory for calculating activity coefficients for aqueous electrolyte solutions. There is the ion-association model which works well for a wide range of water compositions up to about seawater ionic strength and there is the Pitzer specific-ion interaction model that works well up to almost any concentration but for a limited range of compositions. And there are several hybrid models that utilize the best aspects of each. We know the strength and limitations of these models by comparing calculations with actual measurements of chemical potentials in single and mixed electrolyte solutions. The specific-ion interaction model is described as semi-empirical because it has some theoretical justification. It assumes electrolytes behave according to the virial formulation and the rest is curve-fitting. The ion-association model utilizes concepts of ion-pairing which has an intuitive appeal because it gives us an image of interacting ions in solution but it is also based on assumptions. Both approaches can be modified to give results that fit with observations, so the correspondence theory of truth does not lead to a definitive choice. Scientists tend to utilize the model that is most appropriate or useful for the particular problem at hand, hence, the choice is a pragmatic one.

Statistics is a powerful tool in the sciences and essential in the social sciences. However, in geochemistry we are usually looking at correlations and their probable importance. Correlative analysis suffers from the well-known fact that a correlation between two parameters does not equate to cause-and-effect. Many parameters correlate, but not always because one variable directly affects another. It is not possible to determine mechanisms or natural processes from correlations alone. Statistical evaluation of data without phenomenological context can suggest trends that may not have been noticed before, but it does not constitute a scientific model until it is interpreted with observational constraints. Once you provide an interpretation of a trend or correlation by applying knowledge of physico-chemical processes, you have a scientific model, not a statistical model. Before the interpretation is made, the statistical evaluation is an exercise in pure mathematics and not a scientific exercise. Statistics needs an empirical context to be a scientific model. "The pure mathematician who should forget the existence of the exterior world would be like a painter who knows how to harmoniously combine colors and forms, but who lacked models. His creative powers would soon be exhausted" (Poincare, as cited by Pappas (1995)).

So part of the problem is to understand what we mean by 'model.' A model is a theoretical construct that that begins with a concept (the conceptual model) and might be portrayed mathematically or diagrammatically or physically or by analogy. It is always a simplification, an idealization, a picture of how we think about some aspect of physical phenomena. In one sense, it is always incorrect because it is never complete or exact. In another sense, it may be correct because it might be our best understanding or approximation of something at a particular time in history.

A chemical model is a theoretical construct that permits the calculation of physicochemical properties and processes. A geochemical model is a chemical model applied to a geological context. Geochemical codes, such as PHREEQC and Geochemist's Workbench, use solution equilibrium models, such as the ion-association model and the Pitzer ion-interaction model, to calculate aqueous speciation. Algorithms similar to those used to solve speciation are used in these codes to solve equilibrium phase distribution, i.e. dissolution, precipitation, adsorption, desorption, degassing, and ingassing reactions. They can also be solved for non-equilibrium or kinetic phase distribution if the appropriate rate coefficients are known. Reactive-transport codes have incorporated geochemical models and fluid transport models using appropriate numerical approximation techniques that must solve both linear and nonlinear equations.

At the heart of every model is the conceptual model. "The conceptual model is the basic idea, or construct, of how the system or process operates; it forms the basic idea for the model (or theory)" (Bredehoft, 2005). Oreskes and Belitz (2001) further emphasized the importance of the conceptual model and the difficulty in recognizing mistaken concepts:

> "Conceptualization is probably the most thorny issue in modeling. It is the foundation of any model, and everyone knows that a faulty foundation will produce a faulty structure.... Yet what to do about it remains a problem. Much attention in model assessment has focused on quantification of error, but how does one quantify the error in a mistaken idea?... Almost by definition, conceptual error cannot be quantified."

Bredehoft (2003, 2005) has pointed out how important the conceptual model is to the overall success of modeling. It is the starting point and the end point of any modeling exercise. His comments on correcting misconceptions about the conceptual model are worth repeating:

"1. Modelers tend to regard their conceptual models as immutable.
2. Errors in prediction revolve around a poor choice of the conceptual model.
3. Data will fit more than one conceptual model equally well.
4. Good calibration of a model does not ensure a correct conceptual model.
5. Probabilistic sampling of the parameter sets does not compensate for uncertainties in what are the appropriate conceptual models, or for wrong or incomplete models."

4. Model validation

The phrase 'model validation' has been, and still is, used frequently as a part of scientific practice and communication, yet it is the subject of serious debate which continues to cause consternation inside and outside the scientific community. Try doing a Google search for validation and verification of simulation models. Doing this I got 24 million hits. The subject has become a whole new field of study.

My exposure to this issue began through my participation in radioactive waste disposal research. The nuclear waste management and waste regulatory agencies had a presumed need to prove to the public that they could safely dispose of high-level radioactive waste such that it would never harm the biosphere, especially any future human generations. Part of this proof consisted of modeling a leaking repository that had both engineered and natural barriers. Gradually critics asked a most pertinent question: "How do we know that the models are any good?" At that time, in the 1980s, few people understood that this question is general enough to be asked about anything we do, not only in science and engineering, but even in art, literature, politics, and religion.

"In 1980 the Swedish Nuclear Power Inspectorate (SKI) decided to investigate the possibility of initiating an international study on the comparison of model codes for radionuclide migration in geological formations" (Larsson, 1992). This was the beginning of a series of model validation studies that took on various acronymonus titles such as INTRAVAL, INTRACOIN, and HYDROCOIN. The topic was being addressed primarily by hydrogeologists and in 1992 a 2-part special issue of *Advances in Water Resources* of 10 papers was published with the theme "Validation of Geo-hydrological Models" (Celia et al., 1992). For validation of geochemical models there was the CHEMVAL project (Read and Broyd, 1992). When I first experienced these issues in the mid-1980s something did not seem right. I did not have a clear idea of what was bothering me, but I would hear presentations by professionals who would describe their field, theoretical, or laboratory work as having validated their model without having made clear what that meant. Gradually I began to see that some of this fuzziness stemmed from a rarely expressed misunderstanding of what science is and how it could play a supporting role in radioactive waste research. I delved into the philosophy of science (not an easy topic); I read arguments for and against 'model validation;' and I occasionally gave lectures on this subject to receive feedback from others. I was invited to put forth my concerns for the 5th natural analog meeting on the Alligator Rivers Analog Project (Nordstrom, 1994). The present paper is partly an outgrowth of observing and participating in this debate and a review of this controversy from the perspective of a hydrogeochemist who collected notes on the subject over several years.

It was and is commonly thought that good scientific practice should test model predictions with independent observations and, once having succeeded, one could say that the model has been validated. Comparing model predictions with independent observations is good scientific practice but what does it mean to say that the model has been validated? That point is where opinions begin to differ. To quote from Rykiel (1996): "Validation is a thorny issue for both ecological model builders and model users as exemplified by the confusing and often mutually exclusive statements in the literature. For example, model validation is sometimes considered essential (e.g. Gentil and Blake, 1981; Power, 1993), and sometimes considered impossible (e.g. Starfield and Bleloch, 1986; Oreskes et al., 1994). Some authors suggest that models can be validated (e.g. Law and Kelton, 1991), while others contend that models can only be invalidated (e.g. Popper, 1959; Holling, 1978; McCarl, 1984)."

The often-cited paper by Konikow and Bredehoeft (1992) entitled "Groundwater models cannot be validated," followed by the paper in *Science* on "Verification, validation, and confirmation of numerical models in the earth sciences" (Oreskes et al., 1994) should have clarified the issues, but many investigators objected to their theses or preferred their own interpretations. Enough concern was generated within the USDOE (US Department of Energy) that a 2½-day meeting was held in July, 1994 in Denver, Colorado for the Geochemical Integration Team to resolve issues on model validation, especially in regard to a proposed Yucca Mountain nuclear repository site. Another publication was an entire book on model validation (Anderson and Bates, 2001). Although the authors were primarily hydrologists, geologists, and engineers with two contributing philosophers (there were no ecologists, geochemists, microbiologists, or climatologists) a broad range of opinion is represented from those who think that validation is valid to those who do not.

Because there was no summary report from the 1994 DOE meeting that I am aware of, it seems relevant here to summarize some of my impressions made at that time. First, there was general agreement that Total System Performance Assessment (an assessment of the integrated effects of processes thought to affect transport of radionuclides from a nuclear waste repository to the biosphere) cannot have models that are validated. They can only present scenarios that are supported in that they do not contradict the more reliable models. Many felt that sufficiently small scale (laboratory) and process-level models could be partially validated. This argument changes the common meaning of 'validation.' Usually when something is validated, it is done once and thereafter the object is considered validated. Partial validation is an ambiguous phrase. Second, and most important, there was a consensus that the word 'validation' should not be used. The most popular substitute was 'confidence-building.' However, a specific methodology for confidence-building in models was discouraged and a better approach was to tailor confidence-building for a particular model and its application. Third, for some models convergence between prediction and observation is achievable and necessary, whereas for others bounding calculations are all that can be achieved. Non-uniqueness was agreed to be a characteristic of all models. Fourth, natural analogue studies were considered to be a valuable pursuit towards confidence-building in models. Fifth, the only way for assessment to work is for scientists, engineers, and performance assessors to have formal and informal meetings to clear up differences in perceptions, reach a clear understanding of goals, and to design integrated subsystem models. A presentation at the end of the meeting drove home the point that models are not unique, extrapolations far into the future carry enormous uncertainties, and it may not be possible to achieve regulatory standards as they now stand by using model calculations that are extrapolated beyond their time domain of calibration or history-matching.

So how do we define 'model validation?' I have compiled numerous documented definitions in the Appendix A and many of these have serious flaws. Regulatory agencies got off to a bad start with their quite fuzzy and unhelpful definitions. The IAEA defined it as follows: "A conceptual model and the computer code derived from it are 'validated' when it is confirmed that the conceptual model and the derived computer code provide a good representation of the actual processes occurring in the real system. Validation is thus carried out by comparison of calculations with field observations and experimental measurements" (IAEA, 1982). What is a 'good' representation? What are the 'actual' processes? Who decides when adequate confirmation has been achieved? If we knew the actual processes in sufficient detail then the model would not even be needed. It is precisely because the actual processes are not adequately known that we resort to modeling. I have also used the phrase 'real system' myself, but I added a footnote to warn readers that this is an oxymoron (Nordstrom and Munoz, 1994). A system is an arbitrarily defined anthropocentric concept and 'real' implies a physical reality apart from the human conceptualization of it. So there is an initial IAEA definition that defined little. The most recent IAEA update made some important modifications "The *process* of determining whether a *model* is an adequate representation of the real *system* being modelled, by comparing the predictions of the *model* with observations of the real *system*. Normally contrasted with *model verification*, although *verification* will often be a part of the broader *process* of *validation*. There is some controversy about the extent to which *model validation* can be achieved, particularly in relation to modelling the long term *migration* of radionuclides from *radioactive waste* in *repositories*" (IAEA, 2007). In this definition two important points are acknowledged: (i) that model validation is a process and (ii) that there is some controversy about the extent to which model validation can be achieved. Of course the meaning of validation implies a single event, not an ongoing process. The next aspect they need to define is 'adequate representation' and who determines it. These modifications show how convoluted the definitions have become. The more the IAEA tries to improve the definition, the more the definition drifts away from the original meaning. It would be far easier and more meaningful to abandon the word validation altogether.

The US Department of Energy (USDOE, 1986) defined validation even more loosely as the determination "that the code or model indeed reflects the behavior of the real world." Indeed! How that goal might be reached is left to the imagination of the reader. The U.S. Nuclear Regulatory Commission (USNRC, 1990) defined it as "assurance that a model, as embodied in a computer code, is a correct representation of the process or system for which it is intended." Again, how does one know what the correct representation is? For which intended purpose? If one knew the correct representation, then there would no longer be a need for the model or its validation. If the correct representation is not known, then it is not possible to validate the model. Of course, this appears to be a perfect lose–lose situation. Part of this problem has to do with trying to make a general statement for which only detailed context and specific information have meaning.

The Swiss Federal Nuclear Safety Inspectorate (HSK) did not fare much better with their definition for model validation, "Providing confidence that a computer code used in safety analysis is applicable for the specific repository system" (HSK, 1993). Again, a model is not a computer code. And who decides how much confidence is enough?

Nuclear waste management agencies were asked a simple question: How do we know that models are reliable for such long periods of time? The larger question for science and engineering is: How do we know that any model is reliable? To answer this question, nuclear waste agencies asked their consultants to 'validate' their models. Under such circumstances, would a consultant be likely to say that such a project was pointless? Would a consultant, having been given the project and a sizable chunk of funding, likely conclude that a model cannot be validated? I always wondered if a nuclear waste management agency asked the same consultant to invalidate a model, whether he/she could do that? The answer is likely to be affirmative. Consequently, anyone could have predicted the result – models can be validated, or invalidated. The agencies and environmentalists could always get whatever they asked for. One might say that models can be validated because nuclear waste agencies and their consultants knew what they wanted beforehand and could produce the appropriate results. This wish was compounded by a fallacy of logic known as 'affirming the consequent' (Shrader-Frechette, 1993). Affirming the consequent in this context means that, if a model calculation or prediction compares well with an independent observation, the model is assumed to be correct. Put another way, if the model has not been proved incorrect, it must be valid. The syllogism is false because the conclusion is assumed in the major premise.

All that can be said is that when a model prediction compares well with an independent observation is that the model is corroborated or confirmed. It lends some support to the model, but does not corroborate the model for all possible circumstances. The model can never be shown to be correct or validated because one would have to compare predictions for all possible situations to infer the overall correctness of it.

Does good agreement between a model result or prediction and observational measurements mean the model is correct? No, for three possible reasons, (i) if model parameters are not independent from the measurements they are being compared to, they should agree regardless of the correctness of the model, (ii) if the measurements are in error, both the measurements and the model could be in error, and (iii) the model results might agree with reliable measurements for the wrong reasons.

Does poor agreement between a model result and observations mean the model is incorrect? No, for similar possible reasons, (i) if the measurements are unreliable, the model may still be correct, (ii) model calculations could be in error whereas the conceptual model could be correct, and (iii) the criteria for good and poor agreement may be incompatible with the limitations and uncertainties of the model. When model predictions are not confirmed by independent observation, scientists often become very interested. The reason is because it means there is something going on that they had not already known. A new discovery could be made from this finding. That would be far more interesting than finding out things worked the way that a model predicted.

The radioactive waste agencies, in their haste to make safety guarantees to the public, promised results that are not deliverable. Modeling can improve our understanding of how to build a safe repository but it cannot provide a guarantee. Some of our modeling calculations may be the best estimates we have of possible future scenarios but we shall never know for sure. These calculations also provide a "starting point for policy decisions" (Drever, 2011), but they are fraught with large uncertainties. We are not even sure how uncertain are the uncertainties. One of the main problems in the application of science and engineering to major hazardous waste disposal issues is a lack of knowledge about the limitations of science and engineering.

Kirchner et al. (1996) argue for setting high standards for model testing and evaluation. "An important first step, in our view, is to ask modelers to use explicit performance criteria in evaluating their models, and to compare them against explicitly stated benchmarks. This would be a significant improvement over the subjective model evaluations that are common today. Explicitly testing models against other decision-making methods (such as expert opinion) would provide a particularly illuminating measure of the accuracy and reliability of model predictions" (Kirchner et al., 1996). This statement sounds reasonable until one considers the practical aspects of how to go about it. Different fields of science will have entirely different criteria. Who will set the criteria? How exact must the criteria be? Part of the problem here is that it is a moving target. Criteria and the standard-setters can change over time. Then there is the issue of who wants to do this type of investigation. Funding and prestige are major motivations for some scientists. How much funding and prestige can be associated with what most people would consider to be largely a non-innovative, tedious activity?

The word 'validation' is also used routinely as a synonym for quality assurance/quality control (QA/QC) in analytical chemistry (Taylor, 1983, 1986; Thompson et al., 2002; Fajgelj, 2007). Analytical chemists are usually referring to method validation not model validation but many of the same issues arise. If a method is used and validated there is still no guarantee that the analytical values are correct, it simply improves the probability that the results are accurate. Some would say there is a parallel here between 'method validation' and 'model validation,' but there is a big difference. Method validation, which I prefer to call QA/QC (quality control/quality assurance), has developed over a century and a half of testing over and over of chemical reactions and instrumental development and application. Consequently, we have a high degree of certainty, especially when compared to the highly complex, physical, chemical, biological, meteorological characteristics of the environment that are being modeled. Testing a high-level radioactive waste repository would take thousands of years.

Testing consequences of models is a process of making predictions. Again there are misunderstandings of what predictions are, what predictions can be made, and what predictions cannot be made with scientific knowledge.

5. Prediction

5.1. Two types of prediction: phenomenological and chronological

A distinction should be made between two types of predictions, (i) phenomenological or logical prediction and (ii) chronological or temporal prediction (Mayr, 1982; Strahler, 1992; Oreskes, 2000b; Iverson, 2003). Phenomenological prediction uses basic principles of science along with necessary assumptions to form a logical construct with testable consequences. The deduced consequences constitute the prediction. Chronological prediction is foretelling the future which "has traditionally been the province of mystics and clerics" (Iverson, 2003). Furthermore, phenomenological prediction falls into two general types, time-independent prediction and time-dependent prediction. Time-independent prediction would include any kind of chemical reaction outside of kinetics. I can predict that if you put a small piece of elemental Na into a beaker of water that it will react violently, catch on fire, and end with a small explosion. I know that because (i) my chemistry textbook tells me so and explains the reaction in terms of a stoichiometric equation, (ii) I have seen other people do this and it happens every time, and (iii) I have done it myself. The next time someone tries this experiment, it will happen again. I call this time-independent prediction because a phenomenon is predicted that has occurred before and not the time course of a future event. It does not depend on time as an explicit variable or as part of the chemical reaction equation that symbolizes the reaction. 'Prediction Central,' the nickname given to Professor Helgeson's laboratory of theoretical geochemistry at the University of California Berkeley, developed techniques of predicting fluid–rock interactions, given a sufficiently robust and comprehensive database. Similarly, the Marini and Accornero (2006) paper entitled "Prediction of the thermodynamic properties of metal–arsenate and metal–arsenite aqueous complexes to high temperatures and pressures and geological consequences" was predicting properties, not time-dependent processes. Classical equilibrium thermodynamics is time-independent and non-equilibrium or irreversible thermodynamics is time-dependent. A time-dependent prediction would be one in which the rate of the reaction of Na with water is known from the literature or from my own experiments and I have measured the mass of Na and I have a kinetic equation with time as an explicit variable. Then, I could predict how long the reaction would take. However, I still might not be able to predict exactly when I would actually do the experiment again because that would be a highly uncertain chronological prediction until I scheduled it (no current plans for this). Science has traditionally made phenomenological predictions not chronological predictions, but with major issues of global warming, climate change, water-supply limitations, water-quality degradation, air-quality degradation, natural-resource extraction limitations, and hazardous waste disposal challenges science and engineering experts are being asked to make chronological predictions.

Chronological prediction is foretelling the future. Frequently, it is based on human feelings, emotions, prejudices, and opinions with little use of logic. The prediction may be correct but it may not be science. It is more like betting on the horse races, water-witching, or the pronouncements of doomsday prophets. In contrast, groundwater models are developed to estimate present and future conditions given certain properties of the aquifer including permeability, porosity, recharge rates and changes in those rates, discharge rates, storage capacity, etc. These models are not guesses or mere opinions; they are based on a knowledge of average material properties of the aquifer and the fluid, on principles of fluid dynamics, and on assumptions or data about initial conditions, boundary conditions, and heterogeneities or the lack thereof. They also make assumptions about climate trends (which affect recharge rates, i.e. both precipitation and evapotranspiration rates) and withdrawal rates. It is these assumptions that limit the certainty of groundwater model calculations. In geochemistry, reaction rates between water and minerals and aqueous and gas phase reactions are also needed to predict the chemical evolution of an aquifer over time (as well as information from the groundwater model). Water–rock reaction rates are quite complicated functions of the degree to which the aquifer acts as an 'open' or 'closed' system, aquifer composition (mineralogy and chemistry), water composition, organic matter content, surface area, surface chemistry, temperature, pressure, and microbial ecology. To determine all these functionalities in the laboratory is not necessarily practical because it would take numerous lifetimes unless some practical generalizations could be demonstrated. Where possible, field rates should be measured and interpreted in terms of what is known from laboratory experiments. Long-term climate trends and long-term groundwater use by society cannot be predicted very well, if at all, and the longer the future extrapolation, the less certain the model results are and the less possibility there is of testing the predictions. The extent of reasonable future extrapolation for groundwater modeling is often stated to be of the same time frame as that for history-matching if any confidence in the extrapolation is desired. The state of the art for groundwater modeling and geochemistry including future needs and possibilities has been well summarized by Glynn and Plummer (2005) and Konikow and Glynn (2005).

Several scientific and engineering investigations attempt to connect phenomenological prediction with chronological prediction: planetary motion, comet recurrence, volcanic eruptions, earthquakes, groundwater supply and contaminant groundwater plume mobility, high-level nuclear waste mobility, mining effects on water quality, flooding, coastal erosion, and climate change. The deduced consequences from these studies reflect an enormous range of uncertainties. There are numerous measurements of planetary motions, the orbits are regular and cyclical, and the laws that govern them are few and not complicated. Hence, phenomenological and chronological predictions are virtually the same, predictions are accurate, and uncertainties in the calculations are quite small. The other extreme might be biological evolution. We understand that the DNA of an organism determines its physical characteristics and that natural selection and mutations determine evolution but, given an initial condition, one cannot predict the evolutionary path of an ecosystem in the same way that one can predict planetary motion. As Iverson (2003) mentions, the attributes of non-linearity, highly dissipative structures, and contingencies on past and future events break the connection between phenomenological and chronological predictions. Over the next 1 ka it would be very difficult to predict when and how many eruptions there might be from Etna and Pinatubo, but by measuring gas emissions, seismic activity, and heat flux one might be able to predict if an eruption is imminent or not. The closer it is to the time of an eruption, the more it is possible to predict the timing of the eruption to within weeks, days and sometimes hours. We can measure motion in the Earth's crust and the stress buildup between rock on opposite sides of a fault zone which allows us to constrain a major earthquake probability to a period of time measured in centuries or decades, but to predict the date to within weeks ahead is not possible at this time. What is possible is to be prepared for a large magnitude earthquake in high-risk regions, a much safer alternative to relying on earthquake prediction.

Regular cyclic events, governed by a few simple laws or principles, can be readily predicted. Indeed it appears that planetary motions were predicted by the builders of Stonehenge and the Egyptian pyramids before advanced mathematics and telescopes and without the benefit of the governing equations. Irregular and less frequent events like volcanic eruptions and earthquakes have very limited (and often not helpful) predictability. Although we understand the forces involved we do not have direct measurements of those forces in the deep subsurface near the sites where the greatest stress occurs. Groundwater conditions can be predicted if we choose appropriate initial conditions, boundary conditions, representative parameters, and future changes in input and output functions (climate variations and human activities). Because climate change and human activities cannot be forecast very far in the future, only groundwater conditions for the near future can be predicted with some confidence. An appreciation for the different types of predictions and their limitations can go a long ways towards improving the communication of science to policy makers and the public.

5.2. Predictions of rare or complex events are ambiguous because of model non-uniqueness

Some of the first geochemical calculations of water–rock interactions with a digital computer used a 'forward modeling' approach consisting of a given mineral assemblage and a given water composition (Helgeson et al., 1969, 1970) that are allowed to react. If any mineral solubility equilibrium was reached the mineral was allowed to precipitate. Constraints on these systems were generally minimal in that equilibrium was assumed, the primary mineral assemblage was reasonable for an initial rock type, and secondary minerals were those thought to form under certain conditions of pressure and temperature. Assumptions had to be made about whether the system was open or closed during the evolution of geochemical reactions, the correct activities of solid phases and solutions species, the correct speciation, whether solid-solution minerals form and how to formulate their activities and reactive processes, the effects of temperature and pressure gradients, and, if it is a hydrothermal fluid, at what temperature boiling takes place. These were major assumptions that were not adequately known or evaluated at the time. Although these calculations marked a tremendous advance in our ability to model geochemical processes, the uncertainties in the thermodynamic database and these assumptions did not provide confidence in the results. The calculated results were subject to numerous unquantifiable uncertainties. Hence, these models could give very different results if some of the assumptions or input data were changed. Confidence improved with advances in the underlying thermodynamic database and by increased testing against field data, but an enormous amount of data is needed to substantially reduce uncertainties. Reed and Palandri (2010) has shown that even with improvements in thermodynamic data and using the same data and the same initial conditions, it is possible to calculate the same mineral assemblage (an ore deposit) with very different pathways (or models), i.e. an isobaric pathway vs. an isentropic or isenthalpic pathway. Different models can produce the same result. By incorporating more data such as isotopic compositions and fluid inclusion data from a specific field site the needed constraints for the possible pathways might be found. To reduce possible model interpretations, limitations can only come from the field data, it cannot come from the modeling itself.

Even without these uncertainties, model calculations are not unique because there can be multiple roots to the governing equations (Bethke, 1992; Oreskes et al., 1994). Although this non-uniqueness is mathematical in nature, it is usually possible to resolve it through constraints derived from appropriate field conditions.

The problem of non-uniqueness is not just relegated to mathematical uncertainty or insufficient knowledge of thermodynamic and field data. Pierre Duhem (1954) recognized the non-uniqueness of scientific explanation in general. His writings on this subject, supported by those of W.V.O. Quine, became known as the

Duhem–Quine thesis (Oreskes et al., 1994). They argued that scientific theories are not uniquely determined by observational data (Derry, 1999). One of Duhem's examples (Duhem (1954) illustrates this point quite well. Léon Foucault modified an experiment originally devised by François Arago and improved by Hippolyte Fizeau to demonstrate that light consisted of waves, not particles (the corpuscular theory), by comparing the speed of light in water to that in air. Foucault and Fizeau demonstrated that the speed of light in water was slower than in air, contrary to the predictions of Newton and others who supported the corpuscular theory. This conclusion was thought to put an end to the corpuscular theory but Duhem pointed out that there might be another corpuscular theory based on different postulates that could still account for the difference in the speed of light. He made this statement just before Einstein had reported on his theory of light quanta, or photons, which, indeed, was such a theory. Duhem provided one of the clearest descriptions of his time of what science is really about. He pointed out that scientific laws are neither true nor false but approximate, that laws are provisional partly because they are symbolic, i.e. they are connections among symbols not the actual realities.

Another important step forward in geochemical modeling was taken with the introduction of mass balances into water–rock interactions, known as "inverse modeling" (Plummer, 1984; Glynn and Plummer, 2005). The basic concept is that if you have mineralogical data on a groundwater aquifer and you have access to evolving groundwater compositions along a flow path, you can solve a matrix for the relative proportions of minerals dissolving and precipitating, gases exsolving or dissolving, organic matter reacting, etc. The water analyses can also be used to calculate mineral saturation states that provide further constraints on the possible reactions. Instead of making a lot of assumptions as in forward modeling, you have fewer assumptions because your modeling is constrained by actual field data. If water residence times can be obtained, weathering rates can be calculated as well (Burns et al., 2003). Although there will always be some aspects of non-uniqueness, inverse modeling with sufficient field data, can go a long way toward narrowing the most likely model interpretations for water–rock interactions.

6. Prediction and validation for models of nuclear waste disposal

Scientists and engineers were asked to devise a high-level radioactive waste disposal strategy that would never allow radionuclides from a leaky repository to travel to the biosphere and harm human and environmental health for at least 10 ka, preferably 100 ka to 1 Ma (after which radioactive decay would have reduced the most harmful levels of radioactivity to safe levels). It is interesting to note that a repository would be assumed to leak in less than 10^4 years, a reasonable assumption, and that radionuclide transport over $\geqslant 10^4$ years could be reliably modeled, an unreasonable assumption. The arrogance in thinking that such models, based on huge extrapolations over time, were not only possible, but were considered probable and provided safety guarantees for the future led to a strange journey. We can summarize the substantial obstacles confronting those in the radioactive waste research community who attempt to validate models of radionuclide migration from a leaking repository: (i) the original definitions of model validation offered by federal nuclear waste agencies and regulatory agencies were far too vague to be useful, (ii) the 'problem of induction' or the problem of extrapolating from the specific to the general cannot establish the validity of a model, (iii) short-term tests cannot establish long-term responses to migrating radionuclides, (iv) overemphasis on validation rather than understanding the processes operative in a repository environment, (v) publicity campaigns focused on making guarantees to the public when such guarantees are not possible, (vi) affirming the consequent, and (vii) confusing modeling with reality. One summary statement by Leiss and Chociolko (1994) seems appropriate, "In general there is now widespread agreement that until recently the capacity of scientific risk assessment to render definitive verdicts for the non-expert public on the scope and magnitude of most hazards was 'oversold' by experts, and that lower expectations would be more fitting."

The main problem is that there is nothing scientific about modeling a high-level radioactive waste disposal site for 10 ka or more because the calculations cannot be tested and the technical basis makes no sense (Ramspott, 1993). The modeling efforts are not testable within the time frame that policy decisions must be made. 'Validating' a model that has no testable consequences is meaningless. This situation is reminiscent of the debate between the rationalists and the empiricists. Empiricists demand testability, or falsifiability, as Popper (1959) a well-known advocate of logical positivism, described. Testability connects ideas with reality. Without even the possibility of testing model calculations within a reasonable time period, the calculations are all just exercises in idealistic or rationalistic scenarios which may or may not have any relationship to reality because we cannot know the outcomes with any comfortable degree of certainty. They are chronological predictions without knowledge of long-term processes to make them phenomenological predictions.

Those who have worked on potential repositories such as Yucca Mountain, Nevada, have responded with the argument that component models of the overall assessment picture can be separately modeled and evaluated. Of course they can, but you are still limited to a much smaller time scale than $\geqslant 10^4$ years and, when combining component models into the larger picture of radionuclide transport to the biosphere, there are important feedbacks and synergisms that do not appear in the individual models (the total is greater than just the sum of its parts).

Whoever wrote the section on verification and validation for Wikipedia may have been an engineer and a QA/QC employee of the medical field (see Appendix A). The telltale signs include not only the numerous references to engineering, medical, and pharmaceutical sources but the categories of verification and validation are broken down into 'prospective validation,' 'retrospective validation,' 'full scale validation,' 'partial validation,' 'cross-validation,' 're-validation/locational or periodical validation,' and 'concurrent validation.' Who knew that there were so many ways to validate something? One must assume there are an equal number of ways to invalidate the same things. With these new definitions on hand, I'm sure that some folks would like to see retrospective validation of members of Congress and I would like my next parking coupon to be a full-scale validation (i.e. valid anytime, anywhere). I'm sure that the authors had good intentions and had strong support from certain quarters of business and engineering, but there is an inherent contradiction with the normal use of validation, which refers to something being correct or incorrect. If something is valid, it is founded on truth or fact, and it is considered stronger or more believable than something not based on truth or fact. If we now include 'partial validation' and a host of other types of validation then we have changed the original meaning of validation. A counter argument might be that when we validate something, we do so only for a limited set of conditions. If that is in fact the meaning, then, to this author, we should not be using the word validation in the first place.

Another factor at play here is the desire on the part of the safety assessment and nuclear waste management to provide convincing evidence to the public that it is possible to guarantee that this most hazardous waste will not affect future generations. By using the word validation instead of 'provisional acceptance' (or partial validation), a stronger and more confident meaning is implied than is actually warranted. What would your reaction be if someone told you that model A was provisionally accepted (or partially validated) and model B was validated? If both models served the same purpose, would not you feel more comfortable using model B? Yet we are told that validation really means provisional acceptance (Rykiel, 1994; Younker and Boak, 1994). I do not think I am speculating very much, if I suggest that most people do not think that validation is the same as provisional acceptance. And it does not help to claim that 'validation' is being used in a strictly technical sense (Rykiel, 1994) because this language is being used to explain to the non-technical public why a high-level nuclear waste repository will keep nuclear contamination safe from the biosphere. It does not help the credibility of the engineering community to tell the public that they must learn the technical language of the engineer. Indeed, it is the responsibility of the engineer to translate their technical jargon to everyday language that the public understands if they wish to obtain public acceptance of a repository design and location.

Tsang (1991), in his short review of modeling and validation within the nuclear waste context, took a broader view of the issue to include evaluation of how conceptual models were constructed and methodologies of parameter correlation. He recognized that it was illogical to validate a model in a generic sense and that models had to be validated with respect to specified processes and with respect to a given site. However, he also advocated the position of systems engineers, i.e. that validation was a necessary and defensible procedure in contrast to the scientific argument that questions the underlying meaning.

In spite of the numerous criticisms and logical flaws with the nuclear management agencies that propose model validation, the validation 'process' continued to be justified (Hassan, 2003).

7. Validation, science, and the relativity of wrong

At least three other phenomena have influenced the use of the phrase 'model validation.' One is the motivation on the part of many scientists to make a greater public impression of their science and to boost their chances of obtaining funding by using phrases such as model validation; another is a general misunderstanding of what science is and what are its strengths and limitations; and another is the tendency for scientists and the public to simplify descriptions of the natural world into black and white terms, i.e. models are validated or they are not. The importance of this controversy cannot be overemphasized because it bears directly on the role of science in society. It is not merely a 'matter of semantics.' Semantics is meaning, and meaning is at the core of communication. It is a matter of perception and understanding. Debates about models and their validation reflect our collective understanding of what science is and how science fulfills its contribution to society. Discussions about models and model validation are ultimately discussions about the meaning of and perceptions about science and engineering.

It is worthwhile asking why 'validation' is still used frequently in discussions on model reliability or even model hypothesis testing. It should be obvious that this word suffers from the misguided attempt to connote truthfulness and to connote reliability without qualification (Konikow and Bredehoeft, 1992; Oreskes et al., 1994). These problems are partly driven by regulatory needs, but even outside of regulations, scientists tend to use this word when they have found some corroboratory evidence for their hypotheses. It is human nature to want to neatly categorize everything into what is valid or 'right' and what is not valid or 'wrong.' We know this as the 'black or white syndrome.' It is much more difficult to recognize the grey in the universe even though grey is much more common than black or white. Isaac Asimov summed it up rather well when he said "The trouble, you see, is that people think that 'right' and 'wrong' are absolute; that everything that isn't perfectly and completely right is totally and equally wrong" (Asimov, 1988). He continued to develop this theme as a result of a letter he received from an English Literature major (who suffered from the idea that everything in science would eventually be proven wrong) and pointed out that instead of scientific theories being wrong there are really degrees of wrongness. The idea that the Earth was flat was much more wrong than thinking of the Earth as a sphere. But the earth is not a sphere, it is more of an oblate spheroid. But the difference between a sphere and an oblate spheroid is the difference between 8 in. (203.2 mm) to the mile (1.609 km) compared to 8.03 in. (204 mm) to the mile. Precise satellite measurements later showed that the Earth is asymmetric and slightly pear-shaped rather than an oblate spheroid. Each improvement in measurements of the Earth's shape is a closer approximation and less wrong than the previous one but it is a much bigger jump from flatness to sphereness than for the later improvements. As Asimov said so well "... when people thought the earth was flat, they were wrong. When people thought the earth was spherical, they were wrong. But if *you* think that thinking the earth is spherical is *just as wrong* as thinking the earth is flat, then your view is wronger than both of them put together."

Asimov (1988) also explained why we have this tendency to classify everything as right or wrong. He said that if we can categorize our experiences and our understanding into easily remembered categories such as 'black and white' or 'right and wrong,' life is easier – we do not have to think so much, but it is unrealistic.

Again, many of those involved in radioactive waste disposal and other areas of hazardous waste related research agree with these points and some even say that this is their own position. What I fail to understand is why they have not figured out that validation is just not the right word to use? The USEPA (2009), following the NRC (2007) lead, has clearly recognized that model validation is an inappropriate phrase and has dropped it from their lexicon and give the following two reasons (i) models contain simplifications and can never correspond exactly to reality and predictions can never be completely accurate (black/white syndrome) and (ii) those that are confirmed for one application cannot predict accurately for multiple applications. Of course, these reasons seem obvious and are stated in an impossibly exaggerated manner (who would ever think that model predictions would be completely accurate?). Instead of model validation, the USEPA and the NRC use the phrase 'model evaluation' and following Beck (2002) they define model evaluation as the attempt to answer four main questions:

i. How have the principles of sound science been addressed during model development?
ii. How is the choice of model supported by the quantity and quality of available data?
iii. How closely does the model approximate the real system of interest?
iv. How does the model perform the specified task while meeting the objectives set by QA project planning?

These four points are further elaborated upon by the USEPA, but it still begs a few questions. Does question ii refer to a necessary database such as thermodynamic or kinetic data or does it refer to field data or both? Thermodynamic and kinetic data evaluation is part of an ongoing and iterative process with its own QA/QC challenges. It is a whole separate field of study. How reliable does the thermodynamic or kinetic database have to be? Field data also has a large number of QA/QC challenges. How is question iii applied? We have already discussed the fact that we do not compare a model to a real system, so the question should be: How close to observational measurements do model consequences or predictions have to be? Who decides and how? Most of the answers to these questions can only come from the stated objectives of a particular study. They cannot be answered in the general form in which the questions are stated. An important conclusion that comes from this discussion is that there is a responsibility on the part of the project supervisors and project oversight committee to make clear what constitutes a necessary and sufficient effort to answer these questions because, if left open-ended, they could be never-ending.

There are two additional overview concerns that need to be covered in this review. One is the complexity paradox described by Oreskes (2000a). Another is the question of why there seems to be a driving need to model and predict future untestable scenarios (Oreskes, 2000b).

8. The complexity paradox

Numerous authors have made note of the fact that our ability to develop sophisticated codes and advanced mathematical models has outstripped our ability to acquire the complex data upon which the codes depend. It seems clear how this situation developed. Our early models and computer programs were quite simple, but as our understanding increased, as computers became faster with more capacity and with software that was more elegant, flexible, and advanced, and as our ability to couple physical and chemical processes progressed, our models and codes became more and more sophisticated. A model and its code would naturally become more complex with further development if its developers wanted them to be more realistic. The paradox is that as computer codes become more complex, they become more difficult to test and confirm that they operate according to basic scientific and mathematical principles and that they are correctly coded. Oreskes (2000a) describes this paradox in terms of representation *versus* refutability, "The closer a model comes to a full representation of a complex earth system, the harder it is to evaluate it. Put another way, the better the model is from the point of view of the modeler, the harder it is for others to evaluate the model. There is a trade-off between representation and refutability." The situation has finally reached the point where Silberstein (2006) asks, tongue-in-cheek, "Hydrological models are so good, do we still need data?" Both Silberstein and Oreskes make the point that for geoscientists the answer to the complexity paradox is relatively simple: collect more field data because that is where the rubber hits the road. Unfortunately, less money and effort has gone into collecting necessary field data compared to modeling exercises. Another benefit from collecting more and better field data is that it decreases litigation time and costs. Many lawsuits and other legal proceedings involving hazardous waste can be simply avoided by collecting more and better field data. Some parties complain that collecting field data takes time and is expensive, but it is trivial compared to the costs and the length of legal proceedings.

The complexity paradox has a more dangerous analogue in complex technical engineering. Not only can the models and the codes be increasingly difficult to test and ensure that quality assurance needs are met but the hardware, the elaborate interconnected array of technical equipment, become more difficult to test, evaluate, understand, repair and maintain. Dumas (1999) has described this situation well, "As technical systems become more complex, they become more opaque. Those who operate ever more complex systems usually cannot directly see what is going on. They must depend on readings taken from gauges and instruments, and this can be very misleading." He goes on to describe several examples of costly and frightening failures in the USA that were caused by this problem. Oreskes (1997) argues that many times models cannot be tested because of additional reasons such as the inaccessibility of some of the input data and inaccessibility of model predictions in practice and in principle. Petroski (1994) recounts several incidents that involved unacceptable belief in the reliability of computer software that led to either dangerous situations or unnecessary loss of life.

9. Modeling mine water chemistry

Pilkey and Pilkey-Jarvis (2007) in a most thought-provoking book, *Useless Arithmetic: Why Environmental Scientists Can't Predict the Future*, have heavily criticized the use of quantitative mathematical models as a useful means of predicting future trends for making policy decisions (or most anything). They cover a broad range of examples that include modeling to prevent the demise of the cod population along the Grand Banks of Newfoundland and Nova Scotia (it did not), Lord Kelvin's modeling for the age of the earth based on the cooling rate of molten rock (wrong), modeling in support of the proposed Yucca Mountain nuclear waste repository (fraught with questionable assumptions), modeling climate change and sea level rise, modeling coastal erosion (senior author's expertise), modeling post-mining water quality in pit lakes, and modeling invasive species. They are not against modeling, per se, but they argue in favor of qualitative models over quantitative models. Unfortunately some of their examples contradict this point. The example of Lord Kelvin's estimate of the age of the Earth by using physical processes, which Pilkey and Pilkey-Jarvis (2007) call quantitative, is a case in point. Lord Kelvin's method for estimating the Earth's age, which came to about 20 Ma on last modification (and supported by Clarence King, first director of the U.S. Geological Survey, who made separate calculations with the same conclusion; Dalrymple, 1991), was quantitative, and it was said to have trumped earlier qualitative methods by sedimentologists and paleontologists who were thinking in terms of an age nearer billions of years. The difference was that the sedimentologists/paleontologists made their estimates from much less faulty assumptions and concepts than Lord Kelvin's estimate. Sedimentologists could make qualitative estimates of how long it takes for a stratum to form and compare that with the thickness of sediments and the correlation of strata to arrive at an overall age of the Earth. There was a considerable difference in the conceptual model, but Lord Kelvin's was thought to be more reliable because of its quantitative nature. More than 50 years later the age of the Earth was settled at 4.55 Ga through the application of radioactive isotope dating (Patterson, 1956). The application of radiogenic isotope dating is certainly a mathematical and quantitative method of modeling through the use of isochron diagrams (isotope evaluation diagrams) and concordia–discordia diagrams (Dalrymple, 1991) which can all be represented mathematically and computer coded. I would call this a case of one quantitative model trumping another one because the conceptual model had improved considerably. It is the conceptual model that matters, not the mathematics.

In chapter seven ("Giant cups of poison"), Pilkey and Pilkey-Jarvis (2007) discuss the modeling of water quality for open-pit mine lakes. This modeling has been done primarily by consultants for obtaining permits for open-pit mines which have a finite lifetime and for which it must be demonstrated that when the mine closes, water quality will meet state and federal standards in perpetuity. Somehow mining consultants and regulators and others were influenced into believing that it really was possible to 'accurately' predict water quality of the lake infills to these open pits for 50–200 a into the future. A clear parallel exists here between safely disposing of high-level waste for a long period of time and promising good water quality after mining for a long period of time. There is also a parallel in that this modeling was also another example of affirming the consequent (they knew what they wanted beforehand). Pilkey and Pilkey-Jarvis (2007) are hardly subtle in their opinion, "Just as in other modeling arenas we have discussed, accurate prediction of future water quality is a fantasy supported by a hyperreligious faith in the predictive power of numbers." Then in their concluding paragraph, "Tempting as it will be to government bureaucrats to continue to use models, the predictive models for the long-term quality of water in abandoned open-pit mines should themselves be abandoned."

In contrast, the recent book on mine pit lakes (Castendyk and Eary, 2009) clearly states that accurate predictions are possible and a laudable goal. "An accurate conceptual model is an essential precursor to the development of an accurate numerical prediction of pit lake water quality," (Castendyk, 2009). Of course the quality of a conceptual model has enormous bearing on the quality of the numerical model, but what does 'accurate' mean? Predicted concentrations agree within 50% of observed concentrations? An order of magnitude? The definition of 'accurate numerical prediction' distinguishes a quantitative model result from a qualitative one, but who draws the line and where? Shafer and Eary (2009) correctly point out that the main problem with developing and applying geochemical pit lake models is the lack of data. There are so few adequate data for predicted and observed pit lake water chemistry data post-mining.

Moran (2000) describes a mine in Nevada for which the pit lake chemistry was predicted in the draft Environmental Impact Statement (EIS) to have a very favorable water quality. It received considerable criticism and was revised. Incredibly, there were no estimates of key metals such as Cu, Pb, Zn and Cd in the draft. In the final EIS, these metals were included and the calculations indicated there would be a water of pH 8.5–9.5 with high dissolved solids concentration and some constituents unable to meet water quality standards for drinking. Actual water composition was still different from that predicted. It is no wonder that Bob Moran, Orin Pilkey, and many others have no reason to believe pit lake model predictions. Unfortunately, these models can be often driven by politics and the need to obtain permits rather than by good science and engineering.

Other examples of comparing predicted with observed water quality for mined sites based on modeling were documented by Kuipers et al. (2006) and the majority of examples did not compare well. Again, these results should not surprise us because the predictions were mostly done to demonstrate future regulatory compliance and suffered from (i) affirming the consequent, (ii) lack of enough examples from similar sites, (iii) faulty assumptions and poor conceptual models, (iv) inexperience with the complicated nature of pit lake hydrology and geochemistry, and (v) inadequate peer review. It seems we have to learn the hard way that any environmental modeling that involves hydrology, geology, geochemistry, microbiology, meteorology, and climatology is extraordinarily complicated, suffers from a lack of data, and the best that we can hope for is to improve our understanding through experience. To expect anything approaching 'accurate' predictions (however that is defined) is generally unobtainable. Weather patterns cannot be predicted years in advance for any given catchment or basin. If we cannot accurately predict weather patterns and climate change for a small region of interest (such as a mine site), how do we expect to predict groundwater flow patterns and water compositions?

Another aspect of predicting water compositions is related to knowing the natural variation in groundwater or surface water composition at any one location over time. Only relatively recently have we learned that Fe photoreduction occurs (independent of pH and water composition), that diel cycles occur with metals and metalloids in streams receiving mine drainage, that storm events can cause huge variations in surface water chemistry (both increasing and decreasing concentrations), and that organic matter can have a major effect on the transport of metals and their sorption properties. Scientists are still discovering what causes changes in water chemistry through seasonal variations and storm events, let alone how to predict them. The answers can often be site-specific and not easily generalizable.

From 2001 to 2006, the present author led an USGS project near Questa, New Mexico, to determine premining groundwater quality at an active mine site in a mountainous terrain (Nordstrom, 2008). The area was highly altered locally with substantial quartz–sericite–pyrite mineralization. Consequently, both the mineral assemblages and the groundwaters were heterogeneous and could not be characterized by a single representative composition. During the first presentation of the final report to the oversight committee, it was mentioned that there would not be a single 'background' concentration for the 14 constituents of concern but rather a range of concentrations that represented the uncertainty in each value for each part of each catchment in the mine area. There were several catchments and usually two sometimes three different groundwater compositions in each catchment. Of course the regulators did not want to hear this result. This method of presenting the results was unacceptable to them. They expected a single concentration number for Cu, another for Zn, another for F^-, etc. for the whole mine area. They were then asked simply if they wanted me to lie. They said "no" and we proceeded with the final report. It ended up with about 12 tables of different ranges of concentrations for the different groundwaters which reflected the heterogeneous nature of both the geology and the groundwater chemistry (groundwater pH values ranged from 3 to 7.5). This was a good example of how the real world does not fit neatly into simple categories that regulatory agencies or industry consultants would prefer. We have to modify our thinking to allow for the large variations in natural phenomena. We have to be sure our conceptual models have considered appropriately the variability as well as the variety of processes and coupling of processes that can occur in nature.

The Questa project was also one in which mathematical models did not play any decisive role in reaching the conclusions. The final results were obtained by using a natural analogue site proximal to the mine site. The quality of the results largely depended on how well we could apply the data collected from the analogue site and how analogous it was. No amount of mathematical modeling could possibly have superseded a detailed study of an analogue site. Likewise with respect to nuclear waste disposal, some of the most useful research information pertinent to building confidence in a final repository design has come from natural analogue studies (Miller et al., 1994). Overall, these are qualitative to semi-quantitative models in terms of applicability to nuclear waste repositories although quantitative models were often used to better understand specific hydrogeochemical processes operating at the analogue sites.

From 1981 to 1989, the current author worked on the first underground research laboratory for testing the properties of granite as a potential containment rock for high-level radioactive waste disposal, the Stripa Project (Nordstrom et al., 1989a). Field work was done at depths of 250–1000 m in a granitic intrusion located at the Stripa mine. The author was a member of the hydrogeochemical advisory group which had the charge of investigating granitic groundwater chemistry and its implications for radioactive waste disposal. Other investigators were studying rock mechanics, hydrogeology, and developing methods of discerning rock fractures, porosity, and permeability in a very low permeability terrain. The groundwater chemistry was unusual in that the Cl^- concentration was elevated (up to 700 mg/L), the pH was elevated (up to 10), the alkalinity almost non-existent (a few mg/L), and the P_{CO_2} was also very low (10^{-5}–10^{-6}). While attempting to disprove the hypothesis that fluid inclusions were leaking into the groundwater from the granite as a source of salt, evidence was discovered that they were. The Br/Cl ratio of fluid inclusions in the granite was identical to that of the groundwater (Nordstrom et al., 1989b) and distinctly different from seawater. This discovery could not possibly have been made with mathematical modeling. It was only possible by formulating hypotheses about possible sources of the salt, testing the hypotheses, and changing the conceptual model when the evidence warranted it. Any effort to mathematically model the Stripa granite–water interactions without acknowledging the potential influence of fluid inclusions would be remiss and more likely to fail.

The International Stripa Project went through three phases and Phase III was being formulated as the hydrogeochemistry group was completing a major part of its work. We were invited to submit a proposal to the third and final phase which was to be an exercise in model validation. The validation exercise was proposed by a geological engineer and made no particular sense to the geochemical research that had been accomplished at that time. We proposed further hydrogeochemical research that would have tested the fluid-inclusion hypothesis and other hypotheses we had for the evolution of the groundwater chemistry. Our proposal was rejected because it did not fit within the 'validation' program that was being designed. When the validation program was finished and published, one of the researchers from Lawrence Berkeley National Laboratory, Jane Long (1994), had this insightful message:

> "Some of what was gained from the Stripa Project was not planned; as in most earth science research, the results may not exactly match the original goals. Nevertheless, the actual results are significant. The stated aims of the project included "validation" of fluid flow and transport codes. I would argue that this is not a possible achievement in a strict sense. Simply changing the definition of validation such that validation somehow becomes achievable trivializes and obfuscates an accurate assessment of the modeling effort. What we have learned is that the codes are a mathematical formalization of the exceedingly more important effort of 'conceptual modeling.'"

10. Science and engineering: a tale of two cultures

In the past, scientists did not routinely refer to corroboration as validation or verification when finding observational evidence for their theories or hypotheses. Today, I am seeing increasing use of these words in scientific papers, which I consider an unfortunate trend. These words are derived from the engineering literature, particularly concepts in systems engineering (operations research, systems analysis, software engineering). Systems engineers are most often defending the use of validation and verification because it is part of their lexicon and protocols. Few in their ranks have questioned the validity of these words in an environmental context. Because engineers are often directing large research operations such as radioactive waste research, groundwater contamination characterization and remediation and mine waste remediation, they are quite comfortable using verification and validation. But there is more to this story. I often wondered why mine site remediation would fail for reasons that seemed predictable. There is an important difference of cultures inherent in science compared to engineering. In scientific study, students and researchers want to find out how natural phenomena work and be able to explain natural processes. In engineering, by contrast, students and professionals want to build, operate, or fix things. The emphasis is not on understanding or explaining problem, it is on action to solve a problem. Of course, you would think that to properly fix a hazardous waste problem, you should first have to understand it. That is why Superfund studies go through a phase of RI/FS (remedial investigation/feasibility study) with an initial characterization phase followed by identifying and utilizing the best available technology for remediation. Unfortunately, characterization of a mine site or a contaminated aquifer does not always involve the expertise of those who specialize in environmental sciences – geologists, hydrologists and geochemists and it is in such cases when remediation most often fails. Another cause of failure is simply the lack of independent oversight and peer review by appropriate experts who do not have a conflict of interest. When scientists and engineers have worked closely together on a complex environmental problem and knowledgeable oversight was utilized, the project invariably progressed more favorably. In my experience, two good examples of this cooperation would be the remediation of the Iron Mountain Superfund site (Alpers et al., 2003) and the Questa baseline and premining groundwater quality investigation (Nordstrom, 2008). I am sure there are many others as well, but rather than being common practice it is still uncommon practice.

11. Reality and interconnectedness

Another underlying problem with our limited ability to understand, explain, and predict natural phenomena is the reductionist approach, which contrasts with the interconnectedness of reality. Occasionally, we admit to the complexity and interconnectedness of the natural world. The hydrologic cycle is a typical example of how moisture in the atmosphere is related to moisture in the soil is related to infiltration of meteoric water to the groundwater table is related to discharge to a lake or river is related to fluvial inflow to the ocean which returns to the atmosphere through evaporation. Now we understand that the variability in climate and weather conditions resulting from global warming affects the hydrologic cycle. Changing patterns of precipitation affect water management, weathering rates, ecosystem distributions, flora and fauna, etc. However, to study our world we tend to take single discipline approaches and then (i) draw boundaries for the system under investigation and (ii) make numerous simplifying assumptions for factors that should be negligible. We slice up the complex and interconnected world so that we can have a meaningful and manageable subsystem on which to apply our disciplinary knowledge. We have to do this because of our limited data and limited capabilities and we hope that there are no surprises through an unknown or unpredictable 'butterfly effect' (Lorenz, 1993). We need to always recognize this disconnect between our modeling and the reality of our physical world. Nature is integrated in ways that are not often known to us, we are not integrated in ways that are often known to us. "The problems of the world are a result of the difference between the way nature works and the way man thinks" (Bateson, 1972).

In addition to our limited subsystem approach to studying the environment, we do further injustice by communicating our findings. Not only is our thinking about reality limited but our ability to tell others about it is limited by language itself. "The minute we begin to talk about this world, however, it somehow becomes transformed into another world, an interpreted world, a world delimited by language" (Gregory, 1990). Science is an abstracting process that goes from sense impressions and percepts to concepts to communicated language. We lose the immediate experience by journeying through this abstracting process. Science is thus a process of transforming experiential knowledge into processed, communicable knowledge.

12. Recommended guidelines on the use of models

From my own experience, and from what I have learned from others, I offer these guidelines:

- *Modeling cannot give an exact answer, only an approximation*. Answers are not right or wrong, but can be useful or not useful. Modeling is not accurate. If we knew the correct answer we would not have any need of models.
- *Modeling can never substitute for reliable and relevant field data*. In earth sciences the best constraints on modeling come from field data. The most robust models will be those which have the widest available data such as chemical, isotopic, and hydrologic constraints.
- *The greatest weaknesses of any model computation are the quality of the input data and the adequacy of the assumptions (implicit and explicit)*; remember GIGO ('garbage in, garbage out').
- *It is the quality of the conceptual model that determines the usefulness and relevance of any mathematical modeling.*
- *Model computations are not unique*. They are not unique (i) for purely mathematical reasons, (ii) because there are insufficient data to constrain the computations to a unique model, and (iii) because there can be multiple plausible conceptual models.
- *Worst-case scenarios are needed to balance out any unrealistically optimistic scenarios*. If modeling results were done to produce a best-case scenario, then insist on a worst-case scenario to show the range of possibilities; if only a worst-case scenario was obtained, get the best-case scenario. Think of the worst-case and best-case scenarios as error bands on the modeling conclusions.
- *Model and code reliability can be tested in some limited ways to see if they will be useful for the particular problem*. Test cases should be standard practice for codes to see how well they perform before letting them loose on an important hazardous waste problem. Worked examples, provided in the user's manual for several geochemical codes, can fulfill a role as test cases.
- *The main conclusion or argument based on a complex computation should be reproducible in a simpler manner by hand calculation*. Very often a hand calculation can catch the main points of a modeling interpretation.
- *Model computations must be explicable to non-modelers*. Explaining geochemical model computations to non-modelers and especially to non-technical audiences is a test of how well the modeler understands the computations and should be mandatory if modeling is used as a basis for hazardous waste decisions.
- *No matter how much data is acquired, no matter how sophisticated the modeling, there are always factors one cannot know that prevent our ability to predict the future.*
- *The more sophisticated the modeling, the less we know about how well the model performs or how it works (the "complexity paradox")*. A balance has to be found between model sophistication and model transparency.
- *Any computer code that is used for regulatory or legal purposes must be transparent*. If important policy decisions are based partly on modeling with codes then the codes must be open to review and evaluation. Proprietary codes have no business in technical studies when results affect public policy.
- *If a model is said to have been validated, insist on having it invalidated also*. Then determine whether modeling has any meaning or usefulness for the specific problem.
- *Is it necessary to predict? Is this the best approach? Must we model to reach a pragmatic policy decision*? We cannot know well in advance exactly when an earthquake will hit, a volcano will erupt, or when and where a major flood will occur but we can prepare for these natural disasters by emergency planning, especially if field-based data are used rather than idealized mathematical models (Baker, 1994). If models are not linked to fundamental principles and actual field data, they will be especially prone to error (Iverson, 2003). Prediction is not necessarily the main point of science (Oreskes, 2000b).
- *A model represents our understanding of a particular system*. Models are useful because (i) they are heuristic, (ii) they can be incorporated into computer codes to manage large databases and test hypotheses, (iii) they integrate different types of data and processes for a given system, and (iv) they should provide new insight into complex problems. They are an important tool for guiding, but not replacing, our thinking in making decisions on hazardous waste remediation. They are approximate because our understanding is always approximate.

13. Conclusions

A consensus as to an acceptable definition for 'model validation' is lacking and there are major flaws with those that have been proposed. Scientists have argued against using this phrase, and philosophers have argued against it. By examining examples where validation was the main objective, the goal was not achieved; validation could either be always achieved or never achieved depending on one's point of view. Hence, the phrase became meaningless. The emphasis on validation was a carryover from engineering practice into environmental investigations in which complexity and the nature of science prevents useful application of 'model validation.' We would do well to drop this word from our vocabulary and strive for more humility in the face of challenging environmental problems.

Models and computer codes should be thought of as dynamic; they are continually updated and modified, sometimes discarded, sometimes merged. In spite of these changes, the modeler should be focused on interpreting the hydrogeochemical processes, not finessing the code. Someone who spends most of his time improving a code is a programmer, not a modeler. The modeler's main responsibility is to the conceptual model. Towards that goal, the modeler should be asking (i) Has the problem been well defined? (ii) Have I applied the Chamberlin (1897) method of multiple working hypotheses? (iii) Has the appropriate science been applied to the problem? (iv) Has a balanced, informative synthesis been developed?

Another concern that has surfaced in the 'model validation' discussions is the different perspectives of the scientist and the engineer. When these two perspectives are joined and work cooperatively towards a common goal, better results are achieved.

Thanks to the considerable efforts of Naomi Oreskes, Ken Belitz, John Bredehoft, Lenny Konikow, Kristin Shrader-Frechette, Mary Anderson, and many others, there seems to be increasing acceptance that model validation is a phrase that should not be used, especially in a regulatory context (NRC, 2007; USEPA, 2009). I am hopeful that this trend will continue.

Acknowledgments

I thank the International Ingerson Award Committee for honoring me with this award and for giving me the opportunity to put my thoughts on modeling down on paper. I thank the National Research Program of the USGS for support of this and my other, less philosophical, research. Most especially, I am deeply in debt to Russell Harmon, special editor of these few papers, and the other authors, whose patience I wore threadbare by taking much too long to write this paper. I am most grateful for the review comments provided by Ann Maest, Russ Harmon, Mike Edmunds, Lenny Konikow and Katie Walton-Day. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A. Selected definitions from the literature

1. Interconnectedness of reality

"Conventional thought is, in brief, the confusion of the concrete universe of nature with the conceptual things, events, and values of linguistic and cultural symbolism. For in Taoism and Zen the world is seen as an inseparably interrelated field or continuum, no part of which can actually be separated from the rest or valued above or below the rest. It was in this sense that Hui-neng, the Sixth Patriarch, meant that 'fundamentally one thing exists,' for he realized that things are not *terms*, not entities. They exist in the abstract world of thought but not in the concrete world of nature." Watts (1960)

"I believe that every event in the world is connected to every other event. But you cannot carry on science on the supposition that you are going to be able to connect every event with every other event. … It is, therefore, an essential part of the methodology of science to divide the world for any experiment into what we regard as relevant and what we regard, for purposes of that experiment, as irrelevant. We make a cut. We put the experiment into a box. Now the moment we do that, we do violence to the connections in the world." Bronowski (1978)

"The world is totally connected. Whatever explanation we invent at any moment is a partial connection, and its richness derives from the richness of such connections as we are able to make." Bronowski (1978)

"When we try to pick out anything by itself, we find it hitched to everything else in the universe." Muir (1911)

"Nature is one, and to me the greatest delight of observation and study is to discover new unities in the all-embracing and eternal harmony." Muir (1909)

"The farther and more deeply we penetrate into matter, by means of increasingly powerful methods, the more we are confounded by the interdependence of its parts….It is impossible to cut into this network, to isolate a portion without it becoming frayed and unraveled at all its edges." Teilhard de Chardin (1959)

2. Model

"An assembly of concepts in the form of mathematical equations or statistical terms that portrays a behavior of an object, process or natural phenomenon." Drinking Water Source Protection [www.sourcewaterinfo.on.ca/content/spProject/glossary.php *accessed 11-12-11*]

"…a well-constrained logical proposition, not necessarily mathematical, that has necessary and testable consequences." Greenwood (1989)

"model – an assembly of concepts in the form of mathematical equations that portray understanding of a natural phenomenon." ASTM (1984)

"There is a variety of things that are commonly referred to as models: physical objects, fictional objects, set-theoretic structures, descriptions, equations, or combinations of some of these. However, these categories are neither mutually exclusive nor jointly exhaustive. Where one draws the line between, say, fictional objects and set-theoretical structures may well depend on one's metaphysical convictions, and some models may fall into yet another class of things." Frigg and Hartmann (2009)

"Models are vehicles for learning about the world. Significant parts of scientific investigation are carried out on models rather than on reality itself because by studying a model we can discover features of and ascertain facts about the system the model stands for; in brief, models allow for surrogative reasoning." Frigg and Hartmann (2009)

"Models, in the sense in which I am using the word here, are imaginary simulations of the real natural systems we are trying to understand. The models include only properties and relationships that we need in order to understand those aspects of the real system we are presently interested in." Derry (1999)

"But whenever physicists want to emphasize their lack of commitment to the reality of what is described by a theory, or to express their consciousness of its limitations, they simply call it a model." Newton (1997)

"It is clear that models, metaphors, and analogies lack the attribute of truth." Newton (1997)

"Like other metascientific concepts, the notion of a model defies formal definition. One might say, perhaps, that a theoretical model is an abstract system used to represent a real system, both descriptively and dynamically." Ziman (2000), Giere (1988)

"For this discussion, we define a model as a representation of a real system or process." Konikow and Bredehoeft (1992)

"To call a model an idealization is to suggest that it is a simplification of what occurs in reality, usually a simplification that omits some relevant features, …" Cartwright (1983)

"But models are almost never realistic in the first sense; and I have been arguing, that is crucial to how physics works." Cartwright (1983)

"A simplification of reality that is constructed to gain insights into select attributes of a physical, biological, economic, or social system. A formal representation of the behavior of system processes, often in mathematical or statistical terms. The basis can also be physical or conceptual." NRC (2007); USEPA (2009)

"Fundamentally, all models are simplifications. Complex relationships are reduced, some relationships are unknown, and ones perceived to be unimportant are eliminated from consideration to reduce computational difficulties and to increase transparency. Thus, all models face inherent uncertainties because human and natural systems are always more complex and heterogeneous than can be captured in a model." NRC (2007)

"Every area of science uses models as intellectual devices for making natural processes easier to understand. The model that reliably predicts the outcome of real events, or that continues to fit new data, is essentially a kind of theory, a broad statement of how nature works." Lehr (1990)

"The word 'model' is used in everyday speech with three distinct meanings: 'a replica,' 'an ideal' and 'to display.' The concept of the model as adopted here combines aspects of all three meanings. In order to simplify environmental systems, models or replicas of them can be constructed. To be useful, these models must display or make clear its structure or how it works." White et al. (1992)

"A model is a substitute for a real system. Models are used when it is easier to work with a substitute than with the actual system. An architect's blueprint, and engineer's wind tunnel, and an economist's graphs are all models. They represent some aspect of a real system – a building, an aircraft, or the nation's economy. They are useful when they help us learn something new about the systems they represent." Ford (1999)

"By design, models are simplifications of the system under study." Ford (1999)

"Science and scientific models begin as ideas and opinions that are formalized into a language, often, but not necessarily, mathematical language." Nordstrom (2003)

"A model takes on the quality of theory when it abstracts from raw data the facts that its inventor perceives to be fundamental and controlling, and puts these into relation to each other in ways that were not understood before—thereby generating predictions of surprising new facts." Judson (1980)

"A model is a formulation that mimics a real-world phenomenon, and by means of which predictions can be made. …In summary, models are not intended to be exact copies of the real world but simplifications that reveal the key processes necessary for prediction." Odum (1971)

"A model of something is a simplified imitation of it that we hope can help us understand it better. A model may be a device, a plan, a drawing, an equation, a computer program, or even just a mental image. Whether models are physical, mathematical, or conceptual, their value lies in suggesting how things work or might work." AAAS (1990)

"Every theory of the course of events in nature is necessarily based on some process of simplification and is to some extent, therefore, a fairy tale." Sir Napier Shaw as cited by C.J. Walters in Odum (1971)

"Though we often think of 'models' in terms of equations and computers, they can be defined more generally as any physical or abstract representations of the structure and function of real systems." C.J. Walters in Odum (1971)

"The failure of a model to predict change is in itself useful, because it points out flaws in the conceptual framework from which the model was developed." C.J. Walters in Odum (1971)

"That will never be, for a bond does not really exist at all: it is a most convenient fiction…both to experimental and theoretical chemists." C.A. Coulson as cited by M.J. Nye (1994)

"Careful thought leads us to the following disturbing conclusion: *Every model is definitely false.* Although we may be able to say that one model is *better* than another, in the sense that it produces more accurate predictions, we cannot say that it is more probable. Therefore, it is inappropriate to try to assign probabilities to models." Morgan and Henrion (1990)

"Remember that all models are wrong; the practical question is how do they have to be to not be useful." Box and Draper (1987)

"Science models are rather a simulation of human consciousness than the reality of the universe." Nalimov (1981)

"Model. A system designed to possess some of the properties of another, generally more complicated, system." Lorenz (1993)

"The main import of mathematics is that it provides a universal system of symbols, rather than merely a means for quantitative judgment… The use of symbolism allows us to widen the horizon of our knowledge beyond immediate experience. Science is the "abstract" representation of reality. We build up science by constructing more and more abstract theories." Hutten (1967)

"The problem is that in mathematics a proposition is capable of being exclusively and immutably true or false. In science any proposition has to be interpreted in the light of a theory, and as such its truth value cannot only change but may not be strictly definable. Thus, there is a dichotomy between inductive and deductive logic." Sanitt (1996)

"A model is simply an abstraction or a simple representation of a real system or process." Hassan (2003)

"A model is, by definition, a simpler representation of the real thing. It is essentially a toy, albeit a useful one, as a mathematical mimic of the real, more complicated, system. It is not a unique opinion that modeling is fine as long as it is not confused with the real thing." Silberstein (2006)

"The characteristic – perhaps the only characteristic – that all theoretical models have in common is that they provide representations of parts of the world, or of the world as we describe it. But the concept of representation is as slippery as that of a model." Hughes (1997)

"A model is a copy or imitation or representation of a real thing." Rothman (1992)

"A model can be broadly defined as a concept or set of linked concepts that aim to explain some part of reality. A mode may be expressed in words, a picture or graph, a mechanical apparatus, or a set of equations that may or may not be solved by analytical or numerical means." Wilcock and Iverson (2003)

"One might say, perhaps, that a theoretical model is an abstract system used to represent a real system, both descriptively and dynamically." Ziman (2000)

"Mathematics and thermodynamics deal with models of reality, not with reality itself." Anderson and Crerar (1993)

"To communicate knowledge, we must use a simplified and abstract symbolism (words, mathematics, pictures, diagrams, analogues, allegories, three-dimensional physical constructs, etc.) to describe a material object or phenomenon, i.e., a model. Models take on many forms but they are all characterized by being a simplification or idealization of reality." Nordstrom (2003)

"A model of something is a simplified imitation of it that we hope can help us to understand it better. A model might be a device, a plan, a drawing, an equation, a computer program, or even just a mental image. Whether models are physical, mathematical or conceptual their value lies in suggesting how things either do work or might work." (AAAS, 1990)

"All models seek to simplify the complexity of the real world by selectively exaggerating the fundamental aspects of a system at the expense of incidental detail. In presenting an approximate view of reality, a model must remain simple enough to understand and use, yet complex enough to be representative of the system being studied." Anderson and Burt (1985)

"In order to simulate the workings of this machine, we usually describe a model made of simple units and obeying simple laws whose motions are then shown to take it to just those points in time and space where experiment can check against the physical world. It does not matter whether this model is made with pulleys and springs and cathode tubes whose behavior has become familiar to us, or it is simply an array of equations to be solved. Either is a model. The true essence of a model is that it is an axiomatic construction like that of Euclid." Bronowski (1951)

"A *mathematical model* is a description of a process or a prediction about the end result of a process, expressed as an equation or equations. A model is a numerical analogue – a set of equations that describes the relationships between parameters that control a process." Pilkey and Pilkey-Jarvis (2007)

"model – In applied mathematics, an analytical or mathematical representation or quantification of a real system and the ways that phenomena occur within that system." IAEA (1982)

"model. An analytical representation or quantification of a real system and the ways in which phenomena occur within that system, used to predict or assess the behaviour of the real system under specified (often hypothetical) conditions. A representation of a system and the ways in which phenomena occur within that system, used to simulate or assess the behaviour of the system for a defined purpose." IAEA (2007)

"The conceptual model is the basic idea, or construct, of how the system or process operates; it forms the basic idea for the model (or theory)." Bredehoeft (2005)

"A naturalistic 'picture' of a dynamic system is a model. Although this word means no more than a simplified representation of a complex entity, and is often used very loosely to mean any abstract theory, it conveys intuitive notions of internal structures and mechanisms." Ziman (2000)

"A model may be defined as a selected simplified version of a real system and phenomena that take place within it, which approximately simulates the system's excitation-response relationships that are of interest." Bear and Cheng (2010)

"So we must ask, what is a model? For one thing, a model is an abstraction of reality. Nature is simply too complex to understand in toto, and so we must abstract from nature's reality those elements that are important to any given circumstance." Hall (2000)

3. Model accreditation

"Accreditation is 'the official certification that a model or simulation is acceptable for use for a specific purpose." (DoD Directive 5000.59 http://triton.dmso.mil/docslib/mspolicy/directive.html). Balci (1997)

4. Model calibration

"The process for generating information over the life cycle of the project that helps to determine whether a model and its analytical results are of a quality sufficient to serve as the basis of a decision." Drinking Water Source Protection [www.sourcewaterinfo.on.ca/content/spProject/glossary.php *accessed 11-12-11*]

5. Model evaluation

"A comparison of model results with numerical data independently derived from experiments or observations of the environment." Drinking Water Source Protection [www.sourcewaterinfo.on.ca/content/spProject/glossary.php *accessed 11-12-11*]

6. Model testing

"Model testing is ascertaining whether inaccuracies or errors exist in the model. In model testing, the model is subjected to test data or test cases to determine if it functions properly. 'Test failed' implies the failure of the model, not the test. A test is devised and testing is conducted to perform either validation or verification or both." Balci (1997)

7. Model validation

"A test of a model with known input and output information that is used to adjust or estimate factors for which data are not available." Drinking Water Source Protection [www.sourcewaterinfo.on.ca/content/spProject/glossary.php *accessed 11-12-11*]

"Validation is a quality assurance process of establishing evidence that provides a high degree of assurance that a product, service, or system accomplishes its intended requirements. This often involves acceptance of fitness for purpose with end users and other product stakeholders. This is often an external process." Wikipedia [*accessed 11-12-11*]

"Model validation is substantiating that the model, within its domain of applicability, behaves with satisfactory accuracy consistent with the M&S [models and simulation] objectives. Model validation deals with building the model right." Balci (1997)

"There is no uniform procedure for validation. No model has ever been or ever will be thoroughly validated. Since, by design, models are all simplifications of the reference system, they are never entirely valid in the sense of being fully supported by objective truth. 'Useful,' 'illuminating,' 'convincing,' or 'inspiring confidence' are more apt descriptions applying to models than 'valid.'" Greenberger et al. (1976)

"... all models leave out a lot and are in that sense false, incomplete, inadequate. The validation of a model is not that it is 'true' but that it generates good testable hypotheses relevant to important problems. A model may be discarded in favor of a more powerful one, but it usually is simply outgrown when the live issues are not any longer those for which it was designed." Levins (1966)

"The terms 'validation' and 'assurance' prejudice expectations of the outcome of the procedure toward only the positive – the model *is* valid or its quality *is* assured – whereas evaluation is neutral in what might be expected of the outcome." NRC (2007)

"Ideally, comparing model results with a real-world situation, a process known as *calibration* or *validation*, tests a model." Pilkey and Pilkey-Jarvis (2007)

"validation: A conceptual model and the computer code derived from it are 'validated' when it is confirmed that the conceptual model and the derived computer code provide a good representation of the actual processes occurring in the real system. Validation is thus carried out by comparison of calculations with field observations and experimental measurements." IAEA (1982)

"Validation, a procedure that provides, by reference to independent sources, evidence that an inquiry is free from bias or otherwise conforms to its declared purpose." Pescatore (1995)

"In the establishment of scientific laws experience plays a twofold part. There is the obvious confirming or confuting of a hypothesis by observing whether its calculated consequences take place, and there is the previous experience which determines what hypotheses we shall think antecedently probable. But behind these influences of experience there are certain vague general expectations, and unless these confer a finite *a priori* probability on certain kinds of hypotheses, scientific inferences are not valid." Russell (1948)

"Each computer code to be used in safety analysis has to be verified. It also has to be shown that the models used are applicable for the specific repository system (validation), taken both individually and as an overall model chain." HSK (1993)

"Difference in validation criteria across the disciplines are accordingly vast." Wilson (1998).

"The essence of science is validation by observation. But it is not enough for scientific theories to fit only the observations that are already known. Theories should also fit additional observations that were not used in formulating the theories in the first place; that is, theories should have predictive power." (AAAS, 1990)

"Hence, every conclusion of compliance with government regulations, or every conclusion of repository safety, on the basis of "verified" or "validated" test or simulation results, is an example of affirming the consequent. Program *verification,* in other words, "is not even a theoretical possibility." One cannot *prove* safety. One can only demonstrate that one has attempted to falsify one's results and either has failed to do so or has done so." Shrader-Frechette (1996)

"In all but the most trivial cases, science does not produce logically indisputable proofs about the natural world. At best it produces a robust consensus based on a process of inquiry that allows for continued scrutiny, re-examination, and revision." Oreskes (2004)

8. Model verification

"The examination (normally performed by the model developers) of the numerical technique in *the* computer code to ascertain that it truly represents the conceptual model and that there are no inherent numerical problems with obtaining a solution." Drinking Water Source Protection [www.sourcewaterinfo.on.ca/content/spProject/glossary.php *accessed 11-12-11*]

"Verification and validation is the process of checking that a product, service, or system meets specifications and that it fulfills this intended purpose. Verification is a quality control process that is used to evaluate whether a product, service, or system complies with regulations, specifications, or conditions imposed at the start of a development phase. Verification can be in development, scale-up, or production. This is often an external process."

"It is sometimes said that validation can be expressed by the query 'Are you building the right thing?' and verification by 'Are you building it right?' 'Building the right thing' refers back to the user's needs, while 'building it right' checks that the specifications are correctly implemented by the system. In some contexts, it is required to have written requirements for both as well as formal procedures for determining compliance." Wikipedia [*accessed 11-12-11*]

"Verification is the confirmation of truth or authority. [It is] the evidence for such a confirmation or a formal assertion of validity. [It is] the establishment of the correctness of a theory, fact, etc. or evidence that provides proof of an assertion, theory, etc. Verification [is] additional proof that something that was believed (some fact or hypothesis or theory) is correct; [it] is an affidavit attached to a statement confirming the truth of that statement." The Free Dictionary [www.thefreedictionary.com *accessed 11-12-11*]

'The end result of verification is technically not a verified model, but rather a model that has passed all the verification tests." [jtac.uchicago.edu/conferences/05/resources/V&V_macal_pres.pdf *accessed 11-12-11*]

"Model verification is substantiating that the model is transformed from one form into another, as intended, with sufficient accuracy. Model verification deals with building the model right. The accuracy of transforming a problem formulation into a model specification or the accuracy of converting a model representation from a micro-flowchart form into an executable computer program is evaluated in model verification." Balci (1997)

"The problem of verification in empirical science resolves itself therefore, into four aspects: (a) the logical structure of the hypothesis and of the research design, (b) the precision and appropriateness of the methods, (c) the criteria of reliability and/or validity, and (d) the level of credibility of the investigator. Strength in one of more of these aspects does not compensate for weakness in the others; for like the proverbial chain, verification can be no stronger than the weakest link in the total research effort." Lastrucci (1963)

"verification: A computer code is 'verified' when it is confirmed that the conceptual model of the real system is adequately represented by the mathematical solution. Verification can thus be carried out, for example, by intercomparison of codes and by comparison of numerical codes with analytical solutions." IAEA (1982)

"The *process* of determining whether a *computational model* correctly implements the intended *conceptual model* or *mathematical model*." IAEA (2007)

"To put it in a nutshell: we can never rationally justify a theory – that is, a claim to know its truth – but we can, if we are lucky, rationally justify a preference for one theory out of a set of competing theories, for the time being; that is, with respect to the present state of the discussion. And our justification, though not a claim that the theory is true, can be the claim that there is every indication at this stage of the discussion that the theory is *a better approximation to the truth* than any competing theory so far proposed." Popper (1972)

9. On knowledge

"It isn't what we don't know that causes the trouble, it's what we think we know that just ain't so." Will Rogers (1879–1935)

"When you know a thing, to hold that you know it; and when you do not know a thing, to allow that you do not know it – this is knowledge." Confucius (551-479 BCE) The Analects

"To know that you do not know is the best. To pretend to know when you do not know is a disease." Lao-tse, #71 from the Tao Te Ching

"'You can know the name of a bird in all the languages of the world, but when you are finished, you'll know absolutely nothing about the bird.... So let's look at the bird and see what it's *doing* – that's what counts.' (I learned very early the difference between knowing the name of something and knowing something.)" Feynman (1988)

"What then is time? If no one asks of me, I know: if I wish to explain to him who asks, I know not." St. Augustine as cited by Russell (1948)

"'Knowledge,' as we have seen, is incapable of precision. All knowledge is in some degree doubtful, and we cannot say what degree of doubtfulness makes it cease to be knowledge, any more than we can say how much loss of hair makes a man bald." Russell (1948)

"'Knowledge' is a sub-class of true beliefs. We have just seen that 'belief' is not easy to define, and true is a very difficult term." Russell (1948)

"Indeed, such inadequacies as we have seemed to find in empiricism have been discovered by strict adherence to a doctrine by which empiricist philosophy has been inspired: that all human knowledge is uncertain, inexact, and partial. To this doctrine we have not found any limitation whatever." Russell (1948)

"We have found it of paramount importance that in order to progress we must recognize the ignorance and leave room for doubt. Scientific knowledge is a body of statements of varying degrees of certainty – some most unsure, some nearly sure, none *absolutely* certain." Feynman (1999)

"So what we end up having to contend with is the false idea that science is objective, value-free and is able to give the whole picture, and the equally false idea that science and literature are fundamentally different pursuits. Behind this is the problematic idea of truth in science and literature. As soon as the idea of an absolute truth is removed from the centre-stage position and replaced by a concept of truth relative to some theory, then literary and scientific truth may converge towards one another. The effect of literature – and at the heart it is this effect which is the fundamental measure of literature – is to mould the way we 'perceive and speak of the world.' What literature works on is human perception, through the medium of a fictional world or framework. This fictional world is not just an alternative to the real world, but is created in the mind of the subject through the text. Truth is thus not propositional but conceptual." Sanitt (1996)

"There may be an eternal objective truth beyond all of our words, but the minute that truth is spoken by a human being it ceases to be either eternal or objective. It becomes then truth compromised by time, concept, vocabulary, history, and prejudice." Spong (1991)

"Can we devise a universal litmus test for scientific statements and with it eventually attain the grail of objective truth? Current opinion holds that we cannot and never will. Scientists and philosophers have largely abandoned the search for absolute objectivity and are content to ply their trade elsewhere." Wilson (1998)

"For it is of the essence of scientific honesty that you do not pretend to know what you do not know, and of the essence of the scientific method that you do not employ hypotheses which cannot be tested." Watts (1951)

"Our knowledge of how the world works is limited by at least five kinds of uncertainty: (1) inadequate knowledge of all the factors that may influence something, (2) inadequate number of observations of these factors, (3) lack of precision in the observations, (4) lack of appropriate models to combine all the information meaningfully, and (5) inadequate ability to compute from the models." AAAS (1990)

"We must learn to understand that the content of all knowledge is empirical; that its test is whether it works; and we must learn to act on that understanding in the world as well as in the laboratory." Bronowski (1951)

References

AAAS, 1990. Science for All Americans. Am. Assoc. Adv. Sci., Project 2061. Oxford University Press, New York.

Alpers, C.N., Nordstrom, D.K., Spitzley, J., 2003. Extreme acid drainage from a pyritic massive sulfide deposit: the Iron Mountain end-member. In: Jambor, J.L., Blowes, D.W., Ritchie, A.I.M. (Eds.), Environmental Aspects of Mine Wastes, Mineral. Assoc. Canada, vol. 31, pp. 407–430.

Anderson, M.G., Bates, P.D. (Eds.), 2001. Model Validation: Perspectives in Hydrological Science. John Wiley & Sons, New York.

Anderson, M.G., Burt, T.P., 1985. Modelling strategies. In: Anderson, M.G., Burt, T.P. (Eds.), Hydrological Forecasting. Wiley & Sons, New York, pp. 1–13.

Anderson, G.M., Crerar, D.A., 1993. Thermodynamics and Geochemistry. Oxford University Press, New York.

Asimov, I., 1988. The Relativity of Wrong. Kensington Books, New York.

ASTM, 1984. Standard Practice for Evaluating Environmental Fate Models of Chemicals. Am. Soc. Test. Mater. Doc. E978-84.

ASTM, 1992. Standard Practice for Evaluating Mathematical Models for the Environmental Fate of Chemicals. Amer. Soc. Test. Mat. Doc. E978-92.

Bair, E.S., 1994. Model (in)validation – a view from the courtroom. Ground Water 32, 530–531.

Bair, E.S., 2001. Models in the courtroom. In: Anderson, M.G., Bates, P.D. (Eds.), Model Validation: Perspectives in Hydrological Science. John Wiley & Sons, New York, pp. 58–76.

Baker, V.R., 1994. Geomorphological understanding of floods. Geomorphology 10, 139–156.

Balci, O., 1997. Verification, validation and accreditation of simulation models. In: Proc. 29th Conf. Winter, Simulation, pp. 47–52.

Bateson, G., 1972. Steps to an Ecology of Mind. University of Chicago Press, Chicago.

Bear, J., Cheng, D.-D., 2010. Modeling Groundwater Flow and Contaminant Transport. Springer, New York.

Beck, M.B., 2002. Model evaluation and performance. In: Al-Shaarawi, A.H., Piegorsch, W.W. (Eds.), Encyclopedia of Environmentrics. Wiley, Chichester, UK.

Benfey, O.T., 1958. August Kekulé and the birth of the structural theory of organic chemistry in 1858. J. Chem. Ed. 35, 21–23.

Bethke, C.M., 1992. The question of uniqueness in geochemical modeling. Geochim. Cosmochim. Acta 56, 4315–4320.

Box, G.E.P., Draper, N.R., 1987. Empirical Model-Building and Response Surfaces. John Wiley & Sons, New York.

Boyd, R., 1991. Confirmation, semantics, and the interpretation of scientific theories. In: Boyd, R., Gasper, P., Trout, J.D. (Eds.), The Philosophy of Science. MIT Press, Cambridge, Massachusetts, pp. 3–36.

Bredehoeft, J., 2003. From models to performance assessment – the conceptualization problem. Ground Water 41, 571–577.

Bredehoeft, J., 2005. The conceptualization model problem – surprise. Hydrogeol. J. 13, 37–46.

Bronowski, J., 1951. The Common Sense of Science. Heinemann, New York.

Bronowski, J., 1956. Science and Human Values. Revised and republished in 1975 by Harper and Row, New York.

Bronowski, J., 1978. The Origins of Knowledge and Imagination. Yale University Press, New Haven, Connecticutt.

Burns, D.A., Plummer, L.N., McDonnell, J.J., Busenberg, E., Casile, G., Kendall, C., Hooper, R.P., Freer, J.E., Peters, N.E., Beven, K., Schlosser, P., 2003. Geochemical evolution of riparian ground water in a forested Piedmont catchment. Ground Water 41, 913–925.

Cartwright, N., 1983. How the Laws of Physics Lie. Clarendon Paperbacks. Oxford University Press, Oxford.

Castendyk, D.N., 2009. Conceptual models of pit lakes. In: Castendyk, D.N., Eary, L.E. (Eds.), Mine Pit Lakes: Characteristics, Predictive Modeling, and Sustainability. Soc. Min. Metall. Explor., Littleton, Colorado, pp. 61–76.

Castendyk, D.N., Eary, L.E. (Eds.), 2009. Mine Pit Lakes: Characteristics, Predictive Modeling, and Sustainability. Soc. Min. Metall. Explor., Littleton, Colorado.

Celia, M.A., Gray, W.G., Hassanizadeh, S.M., Carrera, J. (Eds.), 1992. Validation of Geo-hydrological models. Adv. Water Resour. 15, 1–214.

Chamberlin, T.C., 1897. The method of multiple working hypotheses. J. Geol. 5, 837–848.

Dalrymple, G.B., 1991. The Age of the Earth. Stanford University Press, Stanford, California.

Derry, G.N., 1999. What Science Is and How It Works. Princeton University Press, Princeton, New Jersey.

DeWitt, R., 2010. Philosophy of science. In: Fritz, A. (Ed.), Philosophies of the Sciences. Wiley-Blackwell, West Sussex, UK, pp. 9–37.

Drever, J.I., 2011. "Prediction is hard – particularly about the future" (editorial). Elements 7, 363.

Duhem, P., 1954. The Aim and Structure of Physical Science. Princeton University Press, Princeton, New Jersey, (translated from La Théorie Physique: Son Objet, Sa Structure, 2nd ed., 1914, by P.P. Wiener).

Dumas, L.J., 1999. Lethal Arrogance: Human Fallibility and Dangerous Technologies. St. Martin's Press, New York.

Einstein, A., 1921. Geometry and experience. Lecture before the Prussian Academy of Sciences, January 27, 1921. In: Seelig, C. (Ed.), 1954, Ideas and Opinions by Albert Einstein. Wings Books, New York.

Einstein, A., 1936. Physics and reality. J. Franklin Inst. 221, 349–382.

Einstein, A., 1940. Considerations concerning the fundaments of theoretical physics. Science 24, 487–492.

Einstein, A., Infeld, L., 1938. The Evolution of Physics. Simon and Schuster, New York.

Eisenberg, D., 2003. The discovery of the α-helix and the β-sheet, the principle structural features of proteins. Proc. Natl. Acad. Sci. USA 100, 11207–11210.

Fajgelj, A., 2007. Assuring Quality of Analytical Measurement Results: The IUPAC role. Chem. Inter. 29, May–June.

Feynman, R.P., 1988. What Do You Care What Other People Think? W.W. Norton, New York.

Feynman, R.P., 1999. The Pleasure of Finding Things Out. Perseus Publishing, Cambridge, Massachusetts.

Ford, A., 1999. Modeling the Environment. Island Press, Washington, DC.

Frigg, R., Hartmann, S., 2009. Models in Science. In: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy (Summer 2009 ed.). Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, Stanford, California. <http://plato.stanford.edu/archives/sum2009/entries/models-science/> (accessed 13.11.11).

Gentil, S., Blake, G., 1981. Validation of complex ecosystems models. Ecol. Model. 14, 21–38.

Giere, R., 1988. Explaining Science: A Cognitive Approach. University of Chicago Press, Chicago.

Glynn, P.D., Plummer, L.N., 2005. Geochemistry and the understanding of ground-water systems. Hydrogeol. J. 13, 263–287.

Gödel, K., 1931. Über formal unentscheidbare Sätze oder Principia Mathematica und verwandter Systeme. Teil I. Monat. Math. Physik 38, 173–189.

Gorham, G., 2009. Philosophy of Science: A Beginner's Guide. Oneworld, Oxford.

Greenberger, M., Crenson, M., Crissey, B., 1976. Models in the Policy Process. Russell Sage Foundation, New York.

Greenwood, H., 1989. On models and modeling. Can. Mineral. 27, 1–14.

Gregory, B., 1990. Inventing Reality, Physics as Language. John Wiley & Sons, New York.

Hall, C.W., 2000. Laws and Models: Science, Engineering, and Technology. CRC Press, Boca Raton, Florida.

Hassan, A., 2003. A Validation Process for the Groundwater Flow and Transport Model of the Faultless Nuclear Test at Central Nevada Test Area, Publication No. 45197, DOE/NV/136009-27.

Helgeson, H.C., Garrels, R.M., Mackenzie, F.T., 1969. Evaluation of irreversible reactions in geochemical processes involving aqueous solutions. II. Applications. Geochim. Cosmochim. Acta 33, 455–481.

Helgeson, H.C., Brown, T.H., Nigrini, A., Jones, T.A., 1970. Calculation of mass transfer in geochemical processes involving aqueous solutions. Geochim. Cosmochim. Acta 34, 569–592.

Holling, C.S., 1978. Adaptive Environmental Assessment and Management. John Wiley & Sons, New York.

HSK, 1993. Protection Objectives for the Disposal of Radioactive Waste. HSK-R-21/e, Villigen-HSK, Switzerland.

Hughes, R.I.G., 1997. Models and representation. Philos. Sci. 64, S325–S336.

Hume, D., 1739. A Treatise of Human Nature. John Noon, London.

Hutten, E.H., 1967. The Ideas of Physics. Oliver and Boyd, London.

IAEA, 1982. Radioactive Waste Management Glossary. Report IAEA-TECDOC-264.

IAEA, 2007. IAEA Safety Glossary 2007 Edition.

Iverson, R.M., 2003. How should mathematical models of geomorphic processes be judged? In: Wilcock, P.R., Iverson, R.M. (Eds.), Prediction in Geomorphology, Amer. Geophys. Union Monogr. 135, Washington, DC, pp. 83–94.

Judson, H.F., 1980. The Search for Solutions. Holt, Rinehart, and Winston, New York.

Kierkegaard, S., 1941. Concluding Unscientific Postscript to the "Philosophical Fragments", (originally published in Danish, 1846) transl. D.F. Swenson. Princeton University Press, New Jersey.

Kirchner, J.W., Hooper, R.P., Kendall, C., Neal, C., Leavesley, G., 1996. Testing and validating environmental models. Sci. Total Environ. 183, 31–47.

Konikow, L.F., Bredehoeft, J.D., 1992. Ground-water models cannot be validated. Adv. Water Resour. 15, 75–84.

Konikow, L.F., Glynn, P.D., 2005. Modeling groundwater flow and quality. In: Selinus, O. (Ed.-in-Chief), Essentials of Medical Geology: Impacts of the Natural Environment on Public Health. Elsevier Academic Press, Amsterdam, pp. 737–765.

Kuipers, J.R., Maest, A.S., MacHardy, K.A., Lawson, G., 2006. Comparison of Predicted and Actual Water Quality at Hardrock Mines: The Reliability of Predictions in Environmental Impact Statements. Earthworks, Washington, DC.

Langmuir, D., 1971. The geochemistry of some carbonate ground waters in central Pennsylvania. Geochim. Cosmochim. Acta 35, 1023–1045.

Larsson, A., 1992. The international projects INTRACOIN, HYDROCOIN, and INTRAVAL. Adv. Water Resour. 15, 85–87.

Lastrucci, C.L., 1963. The Scientific Approach: Basic Principles of the Scientific Method. Schenkman Publishing Company, Cambridge, Massachusetts.

Law, A.M., Kelton, W.D., 1991. Simulation Modeling and Analysis. McGraw-Hill, New York.

Lehr, J., 1990. The scientific process, Part II: Can we learn it? Editorial. Ground Water 28, 850–855.

Leiss, W., Chociolko, C., 1994 Risk and Responsibility. McGill-Queen's University Press, Montreal & Kingston, Canada.

Levins, R., 1966. The strategy of modeling building in population biology. Am. Sci. 54, 421–431.

Long, J.C.S., 1994. Validation of fracture-flow studies in the Stripa granite. In: In Situ Experiments in Granite, Proc. Inter. Stripa Project, Phase III, OECD/NEA.

Lorenz, E.N., 1993. The Essence of Chaos. University of Washington Press, Seattle.

Marini, L., Accornero, M., 2006. Prediction of the thermodynamic properties of metal–arsenate and metal–arsenite aqueous complexes to high temperatures and pressures and some geological consequences. Environ. Geol. 52, 1343–1363.

Mayr, E., 1982. The Growth of Biological Thought. Belknap Press of Harvard University Press, Cambridge, Massachusetts.

McCarl, B.A., 1984. Model validation: an overnew with some emphasis on risk models. Rev. Market. Agric. Econ. 52, 152–173.

Middleton, W.E.K., 1966. A History of the Thermometer and Its Uses in Meterology. The Johns Hopkins Press, Baltimore, Maryland.

Miller, W., Alexander, R., Chapman, N., McKinley, I., Smellie, J., 1994. Natural Analog Studies in the Geological Disposal of Radioactive Wastes. Elsevier, Amsterdam.

Moran, R.E., 2000. Is this number to your liking? Water quality predictions in mining impact studies. In: Sarewitz, D., Pielke, R.A., Jr., Byerly, R., Jr. (Eds.), Prediction, Science, Decision Making, and the Future of Nature. Island Press, Washington, DC, pp. 185–198.

Morgan, M.G., Henrion, M., 1990. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, Cambridge, UK.

Muir, J., 1909. Three Days with John Muir. The World's Work, vol. 17, Doubleday, pp. 11355–11356.

Muir, J., 1911. My First Summer in the Sierra. Houghton Mifflin Company, Boston, Massachusetts.

Nalimov, V.V., 1981. Faces of Science. ISI Press, Philadelphia, Pennsylvania.

National Research Council, 2007. Models in Environmental Regulation Decision Making. The National Academies Press, Washington DC.

Newton, R.G., 1997. The Truth of Science: Physical Theories and Reality. Harvard University Press, Cambridge, Massachusetts.

Nordstrom, D.K., 1994. On the evaluation and application of geochemical models. Appendix 2. In: Proc. 5th CEC Natural Analogue Working Group and Alligator Rivers Analogue Project, EUR 15176 EN, pp. 375–385.

Nordstrom, D.K., 2003. Modeling low-temperature geochemical processes. In: Drever, J.I. (Ed.), Weathering, Soils, Surface Waters and Groundwaters. In: Holland, H.D., Turekian, K.K. (Ex. Eds.), vol. 5, Treatise on Geochemistry. Elsevier, Amsterdam, pp. 37–72.

Nordstrom, D.K., 2008, Questa Baseline and Pre-Mining Ground-Water Quality Investigation. 25. Summary of Results and Baseline and Pre-mining Ground-water Geochemistry, Red River Valley, Taos County, New Mexico, 2001–2005. U.S. Geol. Surv. Prof. Paper 1728.

Nordstrom, D.K., Lindblom, S., Donahoe, R.J., Barton, C.C., 1989a. Fluid inclusions in the Stripa Granite and their possible influence on the groundwater chemistry. Geochim. Cosmochim. Acta 53, 1741–1756.

Nordstrom, D.K., Olsson, T., Carlsson, L., Fritz, P., 1989b. Introduction to the hydrogeochemical investigations within the International Stripa Project. Geochim. Cosmochim. Acta 53, 1717–1726.

Nye, M.J., 1994. From Chemical Philosophy to Theoretical Chemistry. University of California Press, Berkeley, California.

Odum, E.P., 1971. Fundamentals of Ecology, third ed. W.B. Saunders, Philadelphia.

Oreskes, N., 1997. Testing models of natural systems: Can it be done? In: Dalla Chiara, M.L., Doets, K., Mundici, D. (Eds.), Structures and Norms in Science, pp. 207–217.

Oreskes, N., 2000a. Why believe a computer? Models, measures, and meaning in the natural world. In: Schneiderman, J.S. (Ed.), The Earth around Us: Maintaining a Livable Planet. W.H. Freeman and Company, New York, pp. 70–82.

Oreskes, N., 2000b. Why predict? Historical perspectives on prediction in earth science. In: Sarewitz, D., Pielke, R.A., Jr., Byerly, R., Jr. (Eds.), Prediction, Science, Decision Making, and the Future of Nature. Island Press, Washington, DC, pp. 185–198.

Oreskes, N., 2004. Science and public policy: what's proof got to do with it? Environ. Sci. Policy 7, 369–383.

Oreskes, N., Belitz, K., 2001. Philosophical issues in model assessment. In: Anderson, M.G., Bates, P.D. (Eds.), Model Validation: Perspectives in Hydrological Science. John Wiley & Sons, New York, pp. 23–41.

Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, validation, and confirmation of numerical models in earth sciences. Science 263, 641–646.

Pappas, T., 1995. The Music of Reason. World Wide Publishing/Tetra.

Patterson, C.C., 1956. Age of meteorites and the earth. Geochim. Cosmochim. Acta 10, 230–237.

Pescatore, C., 1995. Validation: the eluding definition. Radio. Waste Manag. Environ. Restor. 20, 13–22.

Petroski, H., 1994. Failed promises. Am. Sci. 82, 6–9.

Pilkey, O.H., Pilkey-Jarvis, L., 2007. Useless Arithmetic: Why Environmental Scientists Can't Predict the Future. Columbia University Press, New York.

Plummer, L.N., 1984. Geochemical modeling: a comparison of forward and inverse methods. In: Hitchon, B., Wallick, E.I. (Eds.), First Canadian/American Conference on Hydrogeology. National Well Water Association, Worthington, Ohio, pp. 149–177.

Popper, K.R., 1959. The Logic of Scientific Discovery. Harper & Row, New York (originally published 1934 as Logik der Forschung).

Popper, K.R., 1972. Objective Knowledge: An Evolutionary Approach. Oxford University Press, Oxford.

Power, M., 1993. The predictive validation of ecological and environmental models. Ecol. Model. 68, 33–50.

Ramspott, L.D., 1993. The Lack of Technical Basis for Requiring a Ten Thousand Year Prediction for Nuclear Waste Management, UCRL-JC-115488 Preprint, Lawrence Livermore National Laboratory.

Read, D., Broyd, T.W., 1992. The international CHEMVAL Project: Verification and validation of geochemical model. In: Proc. 3rd Int. Conf. Las Vegas, Nevada, April 12–16, 1992, pp. 1421–1428.

Reed, M., Palandri, J., 2010. Ascent and cooling of magmatic fluids: Precipitation of vein and alteration minerals. In: Birkle, P., Torres-Alvarado, I.S. (Eds.), Water–Rock Interaction. Proc. 13th Int. Conf. Water–Rock Interaction, Guanajuato, Mexico, 16–20 August, pp. 37–40.

Rothman, M.A., 1992. The Science Gap: Dispelling the Myths and Understanding the Reality of Science. Prometheus Books, Buffalo, NY.

Russell, B., 1912. The Problems of Philosophy. Home University Library and (1959). Oxford University Press, UK, Oxford.

Russell, B., 1945. A History of Western Philosophy. Simon and Schuster, New York.

Russell, B., 1948. Human Knowledge: Its Scope and Limits. Routledge, London.

Rykiel Jr., E.J., 1994. The meaning of models. Letter to the editor. Science 264, 330–331.

Rykiel Jr., E.J., 1996. Testing ecological models: the meaning of validation. Ecol. Model. 90, 229–244.

Sanitt, N., 1996. Science as a Questioning Process. Institute of Physics Publishing, Bristol, UK.

Shafer, W.M., Eary, L.E., 2009. Approaches for developing predictive models of pit lake geochemistry and water quality. In: Castendyk, D.N., Eary, L.E. (Eds.), Mine Pit Lakes: Characteristics, Predictive Modeling, and Sustainability. Soc. Min. Metall. Expl., Littleton, Colorado, pp. 115–126.

Shrader-Frechette, K.S., 1993. Burying Uncertainty: Risk and the Case against Geological Disposal of Nuclear Waste. University of California Press, Berkeley, California.

Shrader-Frechette, K.S., 1996. Methodological rules for four classes of scientific uncertainty. In: Lemons, J. (Ed.), Scientific Uncertainty and Environmental Problem Solving. Blackwell Science, Cambridge, Massachusetts, pp. 12–39.

Silberstein, R.P., 2006. Hydrological models are so good, do we still need data? Environ. Model. Softw. 21, 1340–1352.

Spong, J.S., 1991. Rescuing the Bible from Fundamentalism. Harper, San Francisco.

Starfield, A.M., Bleloch, A.L., 1986. Building Models for Conservation and Wildlife Management. Macmillan, New York.

Strahler, A.N., 1992. Understanding Science: An Introduction to Concepts and Issues. Prometheus Books, Buffalo, New York.

Tarski, A., 1983. Logic, Semantics, Meta-Mathematics, second ed. Hackett Publishing Company, Indianapolis, Indiana (first published in 1936 as Der Wahrheitsbegriff in den formalisierten Sprachen. Studia, Philosophica, vol. 1, pp. 261–405).

Taylor, J.K., 1983. Validation of analytical methods. Anal. Chem. 55, 600A.

Taylor, J.K., 1986. On validation of analytical data. Mar. Chem. 18, 115–119.

Teilhard de Chardin, P., 1959. The Phenomenon of Man. Harper and Brothers, New York.

Thompson, M., Ellison, S.R.L., Wood, R., 2002. Harmonized guidelines for single-laboratory validation of methods of analysis. Pure Appl. Chem. 74, 835–855.

Tsang, C.-F., 1991. The modeling process and model validation. Ground Water 29, 825–831.

USDOE, 1986. Environmental Assessment: Yucca Mountain Site, Nevada Research and Development Area, Nevada, vol. 2 of DOE/RW-0073.

USEPA, 2009. Guidance on the Development, Evaluation, and Application of Environmental Models, EPA/100/K-09/003.

USNRC, 1990. U.S. Nuclear Regulatory Commission Radioactive Waste Management Program Guidelines. NUREG-0865, U.S. Nuclear Regulatory Commission, Washington, DC.

Watts, A., 1951. The Wisdom of Insecurity. Pantheon Books.

Watts, A., 1960. Beat Zen, Square Zen, and Zen. In: This is It and Other Essays on Zen and Spiritual Experience, and introductory essay in the 3-in-1 1994 volume: The Wisdom of Insecurity, The Way of Zen, and Tao: The Watercourse Way. Quality Paperback Book Club, New York.

White, I.D., Mottershead, D.N., Harrison, S.J., 1992. Environmental Systems: An Introductory Text. Chapman & Hall, London.

Wilcock, P.R., Iverson, R.M., 2003. Prediction in Geomorphology. Geophysical Monograph 135, Am. Geophys. Union, Washington, DC.

Wilson, E.O., 1998. Consilience. The Unity of Knowledge, Knopf, New York.

Wöhler, F., 1828. Über künstliche Bildung des Harnstoffs (On the artificial synthesis of urea). Ann. Phys. Chem. 88, 253–256.

Younker, J.L., Boak, J.M., 1994. Geological models. Letter to the editor. Science 264, 1065.

Ziman, J., 2000. Real Science: What it is and What it Means. Cambridge University Press, Cambridge, UK.