Ants, Bees and Spiders:
Hydrologic Models in Experiment and Theory

Paul A. Hofmann

Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Science in Hydrology

New Mexico Institute of Mining and Technology
Socorro, New Mexico

April, 1995

The men of experiment are like the ant; they only collect and use; the reasoners resemble spiders, who make cobwebs of their own substance. But the bee takes a middle course; it gathers material from the flowers of the garden and field, but transforms and digests it by a power of its own...

— *Francis Bacon (1620), Novum Organum*

Acknowledgments

When we arrived in Socorro in 1989, my wife Kathleen continued her bakery business on a casual scale. Her customers soon included the president of NMIMT, two vice-presidents, folks at the Registrar's Office and Financial Aid, and Dr. John L. and Betty Wilson. My success seemed assured. Tagging along on a delivery to John's house one evening, I was invited to work in the hydrology blob lab, an event that led to almost 6 years of observation and increasingly independent participation.

It was maybe my good fortune to be in school while hydrologists and regulators struggled over the validation of groundwater models at places like Yucca Mtn. The justification for attempting this fledgling philosophy of hydrology was easily come by under the circumstances. I was even more fortunate to have financial support from the U.S. Dept. of Energy in the form of an Environmental Restoration and Waste Management Employment Program Scholarship (1992-3) and Fellowship (1993-5). This outside support allowed me research independence of a sort few, I suppose, have exercised.

Thanks are due my undergraduate advisor in the NMIMT Mathematics Dept., Dr. Bill Stone. Dr. Laurence Lattman, former president of the Institute and bread connoisseur, provided the travel funds that took me to the presentation of a paper on Darcy's law, written with my brother Jim, at the 1992 Philosophy of Science Association meeting in Minneapolis; an expanded and reoriented version of that paper serves as the cornerstone of this thesis. Bhaskar Rao of the New Mexico State Engineer Office, Glenn Hearne of the United States Geological Survey, Denver, and Doug McAda of the USGS, Albuquerque, helped me understand the issues in modeling the Tesuque aquifer near Pojoaque, NM. The interpretation of events presented here is, of course, my responsibility alone. Li Chunhong, my friend and sometime research partner on solute mixing at fracture junctions, displayed unfailing good grace and cheer; he was also the only one I ever heard refer to the old version of Advanced Hydrology as "easy".

As my graduate advisor, John Wilson gave me the benefit of his very broad perspective on the earth sciences as both technical and adminstrative challenges. I appreciate his patience with the endless evolution of the work and also his willingness to leave me alone much of the time. The occasional 3-hour bull session always advanced my thinking, though to whatever degree the thesis remains incomprehensible, it is entirely my fault. Thanks go to committee members Drs. Rob Bowman of NMIMT and C. Kent Keller of Washington State University for their thoughtful criticisms. Special thanks are due Dr. Bruce H. Weber of California State University at Fullerton, biochemist and historian of science, for encouraging me to view this project in the broader context of the other empirical sciences. My invisible committee was composed of the dozens of scientists and philosophers, both the quick and the dead, from whose work and thought I have benefitted. This crowd did not hesitate to make many suggestions; some I accepted. In the end I have absorbed all of what I was capable, and made a beginning at a hydrologic tale told from a philosophical perspective.

My brother Jim, of the Philosophy Dept. of California State University at Fullerton, came along for the ride. Throughout my time at NMIMT, he acted as friend, financial lifeline, co-author, sounding board, and fellow hater of the New York Knicks. My wife Kathleen, son Walker and daughter Greta gave up a great deal in Wisconsin to follow me to New Mexico. I appreciate the good-natured way they supported my chance to return to school. Owen was born here and doesn't know what he's missing there; he deserves thanks just for being who he is.

Analytical Table of Contents

Introduction

Part One: Experiments of Fruit

models to generate approximate derivations of Darcy's law. Similar strategies with respect to applied models have been seen to impossibly complicate the validation process; in fact, modeling exercises that support hydrologic structural explanations on any scale do not converge to a final model. This fact, and the idealized conditions incorporated into models, made a realist interpretation or application of predictive models difficult to defend. Structural explanations of Darcy's law, however, do not serve as strict validation exercises, but instead serve two purposes: 1) to assimilate Darcy's law into the larger body of scientific understanding; since such exercises serve only to confirm and categorize empirical results, the absence of strict validation protocols is not fatal; and 2) to illuminate the nature and interdependence of selected fluid and medium properties. Several mutually inconsistent models may be used sequentially to explore the contributions of a wide variety of flow parameters. An analysis of the discovery and later treatment of Darcy's law provides a useful introduction to the nature and limits of a progressive hydrologic philosophy in which models serve a pivotal and intermediary role between experiment and theory.

The methodology of experiment, model and theory introduced in Chapter 5 can be updated and extended by looking at a contemporary example of process research. Mixing behavior at fracture junctions is an important process in field-scale contaminant transport problems. Until recently, detailed experimental work and convincing structural explanations were unavailable to regulate standard practice in numerical simulations of the phenomenon. One of two endmember mixing algorithms was generally favored: advectively controlled streamline distribution (relatively high flow rates; no mixing), or diffusion-controlled redistribution (relatively low flow rates; complete mixing). Computational convenience especially favors the complete mixing assumption. At one time, available understanding and data did not justify further model embellishments, a situation reminiscent of the simplifying assumptions of applied modelers considered in Chapters 2 and 3. When mixing assumptions are incorporated into process models, they lead to similarly indefensible conclusions in the absence of experimental corroboration.

An historical overview examines the main thread of mutually supportive experimental, theoretical and numerical investigations, beginning in 1986. These include research into the significance of flow rate, aperture width and angle of intersection - all in an effort to determine the governing mixing rule and any bounds on its validity. The final sections present the results of numerical and experimental work performed at New Mexico Institute of Mining and Technology. We adopted a mixed causal/Baconian approach, in which it was not thought unreasonable 1) to expect that a third mechanism might prevail at the intermediate flow rates; nor 2) to hope that this phenomenon might be robustly characterized so as to include the two existing mechanisms as endmembers. The causal portion of the exercise consisted of computer simulations resting on certain simplifying assumptions; the numerical results display the implications of the fundamental laws and assumptions invoked. The laboratory experience, on the other hand, proceeded in something of a micro-scale parallel to Darcy's experiment; it proves out Kuhn's comment that experimental data is hardly the *given* some theorists would have it, but rather *the collected with difficulty*. The net effect, however, is to constrain the field of plausible models with both theoretical and experimental benchmarks.

7. Progress: "The Believing Spirit" 229

We have surveyed both the practical efforts of hydrologists to contribute to public affairs and also the amended methodology typical of process research. The development of other science fields strongly suggests that within immature sciences the balance between empirical correlations and theory inclines toward the experimental. The immaturity of hydrology is manifest in the lack of dependable and robust theories needed for confident prediction in complicated systems. Present hydrologic theory constitutes at best a good set of underwear for the *Emperor's new clothes*, says Mary Anderson; at worst, even the underwear is full of holes. The interplay of experiment and theory figures prominently in our account of how hydrologists construct compelling accounts in the absence of a strictly logical protocol for validation. It is the lack of empirical corroboration that separates applied field models from "proper" structural explanations - such as those of Darcy's law - that are commonly employed in process-oriented research. If it is important that hydrology be a predictive science, then the gap between model and prototype must be narrowed by constraints imposed by theoretical and experimental benchmarks. These benchmarks are the business of process-oriented research, which serves as the backbone of a productive long-term hydrologic strategy. The hope is that the non-uniqueness of models can be sufficiently constrained that the practical usefulness of models can overwhelm their technically underdetermined status. Developments similar to the fracture mixing research described in Chapter 6 have generated a cautious optimism among process modelers devoted to nurturing the symbiosis of experiment and theory. This optimism follows an historical thread typified by the admonition of Max Planck: "*Man muss optimist sein!* Science demands also the believing spirit... The pure rationalist has no place here".

Table of Contents

Part Two: Experiments of Light

List of Figures, Tables

Figure

Table_

1

Progress: The Very Idea

It is, I think, particularly in periods of acknowledged crisis that scientists have turned to philosophical analysis as a device for unlocking the riddles of their field.

- Thomas S. Kuhn, *The Structure of Scientific Revolutions*[1]

1.0 Introduction

Recent developments in environmental regulation and restoration suggest the timeliness of an effort to place hydrologic methodology and its hierarchy of models in a thoroughly public perspective. Hydrologists play an important role in describing and quantifying the movement and behavior of water and contaminants in the environment. As a result, their professional judgments - often filtered through adversarial legal proceedings - are central elements in controversies over water supply, protection or remediation, as are the hydrologic models on which they appear to rely.[2] The sources and consequences of any uncertainty in hydrology thus deserve serious attention both from hydrologists themselves and also from regulatory, legal and private interests.

While environmental degradation in the form of contaminated groundwater supplies has spurred an accelerating disquiet in the public at large, a certain uneasiness has echoed of late within the community of hydrologists.[3] Their misgivings often coalesce in issues of justification, confirmation or validation - essentially a matter of assessing or establishing the reliability of scientific data, models or conclusions. Applied hydrologists are often concerned with prediction, and thus questions about the reliability of their predictions stand out in the validation debate. Such questions can interweave a confusing mix of considerations that can be variously addressed on technical, legal or philosophical grounds. The thoughtful hydrologist may well find all of these epistemic standards disquieting as applied to his art. Some degree of uncertainty is associated with virtually every step of a hydrologic investigation and analysis, but it is not obvious whether or how hydrologists must adapt their usual internal justification practices when they turn their attention to public works. Occasionally, the waters have been muddied by the misapprehensions of hydrologists themselves.

Confusion about validation results in part from the youth and incoherence of the discipline, amplified by the pressure of public considerations. Hydrology is a relatively new quantitative discipline, consisting of an interdisciplinary melange still seeking to answer very basic questions about water resources. Differing and even conflicting motivations and expectations are not surprising under the circumstances. More traditional sciences have evolved more naturally or gradually into their present positions in society, with well-documented histories of their competence to provide certain needed answers or services. This does not appear to be the case in hydrology, especially in that portion of the science that contends with contaminant transport in the subsurface. Rather, standards of environmental compliance are set by statute and regulation without seeming regard for the capacity of the science to respond.[4] Vast sums of money have suddenly been made available to pursue a legally-driven mandate with little or no basis in the history of the science. Referring to the story of the Emperor's New Clothes, the prominent hydrologist Mary P. Anderson highlights the incompleteness of hydrologic theory by saying: "In the case of solute transport models the Emperor's underwear, the theory upon which the model is based, is full of holes".[5] With public expenditures comes a correspondingly large public expectation of performance. There is, however, reason to doubt that hydrology is a sufficiently mature science to consistently contribute to the resolution of significant public questions as these are currently framed and debated; there is thus reason to doubt that expense, certainty and protection of the public trust are necessarily correlated. The scale of environmental threats to the public health - and the likely cost of addressing those threats effectively - emphasize the need to catalog and analyze any growing pains within the relatively immature science of hydrology.

The general comments made so far pertain equally to surface and groundwater science, due to their contiguous interests in the water cycle and many shared technical and legal issues. Our focus throughout this study, however, is on groundwater. Occasionally the comments and practices of surface water hydrologists will be used to illuminate that discussion. The significance of the public concerns involved make it critical to ascertain in general and in the particular case how groundwater scientists know what they claim to know. This study will use a combination of general discussion and specific case studies to illustrate several important theoretical and technical issues within hydrology. Many of these issues bear on the question of what degree of confidence should be associated with hydrologic claims. Exploration of the sources of both doubt and confidence in pursuit of practical, public concerns necessarily leads to a focus on methodology; in particular, modeling methodology will emerge as the centerpiece of our larger discussion of how compelling accounts are constructed in groundwater hydrology. Our most general goal is a realistic description of scientific progress within an immature discipline beset by practical expectations.

1.1 Methods in Science

Commentators on scientific methodology are normally at pains to distinguish, relate and rank four things: observation, experimentation, models and theory. The ironic view of the scientific method, however, sees it as a thing more honored in the breach than in practice, or maybe more the province of reminiscing emeritus professors than of working scientists. Nobel laureate biologist Peter Medawar has advanced this point of view (though as something of a strawman), and spoofs the whole business:

> ... they [scientists] are not in the habit of thinking about matters of methodological policy. Ask a scientist what he conceives the scientific method to be, and he will adopt an expression that is at once solemn and shifty-eyed: solemn, because he feels he ought to declare an opinion; shifty-eyed, because he is wondering how to conceal the fact that he has no opinion to declare. If taunted he would probably mumble something about "Induction" and "Establishing the Laws of Nature", but if anyone working in a laboratory professed to be trying to establish Laws of Nature by induction we should begin to think he was overdue for leave.[6]

Medawar himself is quick to point out that lack of discussion does not mean lack of method. Construing the idea of experiment very broadly to include both direct intervention and thought experiments, he proceeds to identify four distinct methodologies. He names each one after a prominent philosopher/scientist whom he rather loosely considers an early apologist or exemplar.

The first of Medawar's categories consists of what he calls demonstrative or **Aristotelian** experiments. In Aristotle's hands, these are intended to demonstrate the validity of preconceived ideas, by first deducing their consequences and then displaying their agreement with our experience. They are thus classic thought experiments reasoning from prior principles (theory) to expected behavior. The diagnostic phrase is: "Given this and that in general, then such and such must be the case in particular".

The secondary titles of Medawar's methods are, of course, only honorifics, and therefore more suggestive than precise. For example, Aristotle would have rejected the use of approximations in the derivation of expected behavior from fundamental principles. He and his sympathizers were and are suspicious of the mathematization of physical problems, since the practice of imposing mathematical formalisms on nature must be "inevitably incomplete *as physics*, because it has left behind the qualitative richness of Nature".[7] The general argumentative posture described by Medawar as Aristotelian is more typically referred to as *causal*, and this is the term that will be used here. Our somewhat broadened category of causal methods includes all top-down arguments from the more general to the more specific; within this category, both approximation and mathematics play major roles.

Medawar's second category recognizes inductive or **Baconian** experiments. These are said to consist of contrived experiences without either explicit theoretical import or critical intent. They are, however, guided by specific questions or expectations, and thus are more sophisticated than simple observations, since "we cannot [simply] browse over the field of nature like cows at pasture".[8] The diagnostic phrase is: "I wonder what would happen if we did this..." Sometimes such procedures and any conclusions drawn from them are simply called *empirical*, but we will retain Medawar's term for reasons that will become apparent, though modifying its implications somewhat.

A third type of investigation relies on crucial or **Galilean** experiments. As described by Medawar, these seek to discriminate between hypotheses. The consequences of competing theories are logically deduced, and compared to actual events. His diagnostic phrase is: "Which of these possibilities is correct?" The procedure is causal (deductive), but the consequences are discriminating. Crucial tests presumably rule out certain avenues of investigation or explanation. We prefer a slightly broader reference that also recognizes crucial demonstrations. These provide evidence that must be considered in any future theory; they can enforce an exploratory and explanatory preference. Since the idea of crucial experiments and demonstrations can be conveyed clearly without reference to Galileo (and can be illuminated profitably with reference to Bacon, among others) we will follow the generic practice and term both types of activity *crucial*.

Finally, deductive or **Kantian** experiments explore the logical consequences of varying the fundamental postulates presumed to lie behind our experience. We might find that our usual postulates are not necessary, after all, as was the case in the development of non-euclidean geometries. The diagnostic phrase is: "Let's see what happens if we proceed from a fundamentally different basis". Truly Kantian games involve far more than simple hypothesis substitution, and are rarely played in hydrology; the repercussions of abandoning conservation of mass, momentum and energy (the fundamental principles underlying hydrology) will go unexamined here.

Despite his broadsides about the irrelevance of the scientific method for typical working scientists, Medawar concludes:

> It could be said - has been said - that there is a distinctive methodology of science which scientists practice unwittingly, like the chap in Moliere who found that all his life, unknowingly, he had been speaking prose.[9]

It is the purpose here to foster a more articulate discussion of methods in hydrology. Having left aside Kantian substitution, the remaining styles of inquiry - causal, Baconian and crucial - form the basis of an examination of method in hydrology.

Causal derivations will be illustrated in several contexts throughout this study. They explicitly include any such derivations for which no actual experience is available for comparison. The pitfalls *and promise* associated with this situation will also be explored, both in connection with predictive applied modeling (Chapters 2-4) and in theoretical support for experimental findings (Chapters 5-6). The primary operational contrast to such methods is provided by empirical or Baconian experiments, as discussed in regard to Darcy's law and solute transport at fracture junction (Chapters 5-6). Crucial experiments - either as tests or demonstations - occupy a potentially illuminating but somewhat contested middle ground; these are occasionally considered in Chapters 4-7.

It would be unfortunate if readers inferred that use of one method or another entails wholesale membership in any particular philosophical camp. The use of the generic titles *causal* and *crucial* minimizes unamplified historical allusions. Setting aside Kantian analysis altogether leaves unaccounted for only our fondness for the term *Baconian experimentation*. A selective review of the works of Francis Bacon that prompt retaining Medawar's honorific can quicken a more detailed taxonomy of hydrologic models. Maybe because he did not consider himself a philosopher, but instead a scientist, Bacon's comments can serve as a remarkably pertinent introduction to the question of methodology in hydrology. A brief historical account of this strangely malleable Baconian science is purposely framed to let the author speak for himself;[10] readers are invited to judge for themselves any relevance to the state of affairs in hydrology.

1.2 Baconian Science

Truth emerges more readily from error than from confusion.
- Francis Bacon[11]

In 1620, Francis Bacon published his *Great Instauration*, the title of which conveys his conviction that science was in need of a thorough restoration; this piece was more or less a preface to his *Novum Organum*, in which he discussed the particulars of his program. The typical presuppositions, goals and methods of the day were all found wanting - so much so, that "the entire fabric of human reason which we employ in the inquisition of nature, is badly put together and put up, and like some magnificent structure without any foundation".[12]

For our purposes, Bacon's criticisms fall into two principal areas. One of his themes concerned what he perceived to be the shortcomings of random, unthoughtful experimentation that was followed by overly hasty leaps to very broad, general formulations:

> But the manner of making experiments which men now use is blind and stupid. And therefore, wandering and straying as they now do with no settled course, and taking counsel from things only as they fall out, they fetch a wide circuit and meet with many matters, but make little progress.[13]

He believed that experimentation needed to be systematic; properly pursued, it would lead to lesser, intermediate empirical connections, "seeing that the nature of things betrays itself more readily under the vexations of art than in its natural freedom".[14] Expanding insights were expected in turn to guide further experimentation,[15] even as their accumulation made possible the pursuit of more general laws.

Secondly, Bacon also attacked what he viewed as overly academic speculation unconnected to actual experience.[16] The dominant scholastic Aristotelians favored deduction of intermediate axioms from what they believed to be fundamental principles. In the absence of systematic experimental support for these beliefs, however, Bacon considered the presupposition of any general "laws" to be gross speculation that could not lead to valid intermediate insights.[17] It is perhaps useful to notice that the historical context of Bacon's work included not only the Copernican (1473-1543) revolution in astronomy, but also, beginning in 1610, the publication of the new telescopic investigations of Galileo Galilei. These events not only contributed to the unsettling of both the specific orthodoxy of geocentrism, but also helped to undermine the general appeal of the orthodox Aristotelian scholasticism. Ironically, Bacon embodied the spirit of the latter, broader challenge, but adhered to the former, specific inheritance.

Bacon thus objected on the one hand to unwarranted inductions and on the other to spurious deductions, and largely distributed the blame for a relatively unproductive science between these failings:

> the sciences to which we are accustomed have certain general positions, ...but as soon as they come to particulars... when they should produce fruit and works, then arise contentions and barking disputations, which are the end of the matter and all the issue they can yield; ...insomuch that many times not only what was asserted once is asserted still, but what was a question once is a question still...[18]

As chronicled by J.B.Bury in *The Idea of Progress: An Inquiry into its Growth and Origin*, the Middle Ages were infertile ground for the idea of permanent progress.[19] Bury attributes the generally backward-looking stance to a belief in the superiority of the ancient philosophers and to a sense of general helplessness before divine fiat. Bury agrees with Bacon that the theories of the ancients were typically founded on opinion, rather than on rigorous experiment. As a result, by Bacon's time, "science has remained stationary for the last two thousand years; whereas mechanical arts, which are founded on nature and experience, grow and increase".[20] The effect on the development of science, and the scale of the *instauration* problem, were described by Bacon as follows:

> ... it seems that men have not been happy hitherto either in the trust which they have placed in others or in their own industry with regard to the sciences; especially as neither the demonstrations nor the experiments as yet known are much to be relied upon. But the universe to the eye of the human understanding is framed like a labyrinth; presenting as it does on every side so many ambiguities of way, such deceitful resemblances of objects and signs, natures so irregular in their lines, and so knotted and tangled. And then the way is still to be made by the uncertain light of the sense, sometimes shining out, sometimes clouded over, through the woods of experience and particulars; while those who offer themselves as guides are (as was said) themselves also puzzled, and increase the number of errors and wanderers. In circumstances so difficult, neither the natural force of man's judgment nor even any accidental felicity offers any chance of success. No excellence of wit, no repetition of chance experiments, can overcome such difficulties as these. Our steps must be guided by a clue...[21]

Despite this gloomy assessment of science in his time, Bacon found cause for hope in a middle way that is evoked today in references to "Baconian experimentation". Although he carefully exempted theological questions from examination by his method, his view was otherwise inclusive and was not restricted to the natural sciences. His method was intended to directly interrogate the world, but in ways and directions suggested by whatever larger understanding was available. He was optimistic that guiding clues might be found by following the cautiously inductive method he sketched.[22] He championed the "interpretation of nature" as "reasoning which is elicited from facts by a just and methodical process", but decried the "anticipation of nature" as something "rash or premature".[23]

He saw no hope of progress in "groping in the dark" nor in the "repetition of chance experiments"; investigators must therefore attend to the careful and selective observation of phenomena. At the same time, Bacon had no fundamental objections to the eventual construction of the most far-reaching generalizations; these were in fact the ultimate goal of what he called *True Induction*: "A method rightly ordered leads by an unbroken route through the woods of experience to the open ground of axioms".[24] Dismissing "the induction that proceeds from simple enumeration" as "childish", Bacon claimed that a true induction would "analyze nature by proper rejections and exclusions; and then, after a sufficient number of negatives, come to a conclusion on the affirmative instances",[25] a far more insightful view of induction than is commonly advanced even today. Furthermore, many such axioms ought to result in what we might call testable deductions:

> In establishing axioms by this kind of induction, we must also examine and try whether the axiom so established be framed to the measure of those particulars only from which it is derived, or whether it be larger and wider. And if it be larger and wider, we must observe whether by indicating to us new particulars it confirm that wideness and largeness as by a collateral security...[26]

Proper axioms of this sort (stemming from proper observations) would recognize deeply seated causal forces within nature, which operate in "a process perfectly continuous, which for the most part escapes the sense":

> For seeing that every natural action depends on things infinitely small, or at least too small to strike the sense, no one can hope to govern or change nature until he has duly comprehended and observed them.[27]

Despite his expectation that use of such methods would eventually result in the "dawn of a solid hope", Bacon simply found graspings after high-level axioms premature, given the state of science in his time:

> I, who am well aware that no judgment can be passed on uncommon or remarkable things, much less anything new brought to light, unless the causes of common things, and the causes of those causes, be first duly examined and found out, am of necessity compelled to admit the commonest things into

> my history. Nay, in my judgment philosophy has been hindered by nothing more than this - that things of familiar and frequent occurrence do not arrest and detain the thoughts of men, but are received in passing without any inquiry into their causes...[28]

Since his method rested on a cumulative theory of knowledge, more and more general laws were expected to become substantiated on the basis of preliminary work. In due course, the accumulation of experience and knowledge was expected to yield reliable general laws. Thus Bury says of Bacon's respectful rebellion from ancient philosophy and method, "Time is the great discoverer, and truth is the daughter of time, not of authority".[29] Taking the same long view, Bacon even included a prescription for reproducibility and peer review with which every modern scientist can concur:

> ...whenever I come to a new experiment of any subtlety (though it be in my own opinion certain and approved), I nevertheless subjoin a clear account of the manner in which I made it; that men knowing exactly how each point was made out, may see whether there be any error connected with it, and may arouse themselves to devise proofs more trustworthy and exquisite, if such can be found; and finally, I interpose everywhere admonitions and scruples and cautions, with a religious care to eject, repress, and as it were exorcise every kind of phantasm [unfounded idea].[30]

Neither Bacon's emphasis on proper preparation and manageable projects, nor his corresponding deflation of schemes to quickly extract natural laws, is evidence of a retreat from practical goals. In fact, Bacon objected to simple observation, hasty empiricisms and unfounded theoretical deductions precisely because of "the unfruitfulness of the way".[31] He was convinced that "an unseasonable and premature tarrying over such things"[32] would ultimately hinder rather than advance his goal of laying "the foundation, not of any sect or doctrine, but of human utility and power".[33] As a result, although the very purpose of science was to advance the human condition, he thought this best achieved by proceeding methodically and without pretense, but instead with due patience in building up a thorough science. Patience is required, since:

> ... no one successfully investigates the nature of a thing in the thing itself; the inquiry must be enlarged, so as to become more general. And even when they seek to educe some science or theory

> from their experiments, they nevertheless almost always turn aside with overhasty and unseasonable eagerness to practice; not only for the use and fruits of the practice, but from impatience to obtain in the shape of some new work an assurance for themselves that it is worth their while to go on; and also to show themselves off to the world, and so raise the credit of the business in which they are engaged. Thus, like Atalanta, they go aside to pick up the golden apple,... and let the victory escape them.[34]

Bacon sent a copy of *Novum Organum* to King James I, as part of an endless campaign for personal political position and influence. The King "likened it, in a well-used formula, to the peace of God, since it passed all understanding".[35] Present day readers of this thesis, however, with Bacon's goals and something of his method at hand, should have no trouble understanding the following comment, in which Bacon notes that even God took one day just to shed some light on His problem:

> ...in the true course of experience, and in carrying it on to the effecting of new works, the divine wisdom and order must be our pattern. Now God on the first day of creation created light only, giving to that work an entire day, in which no material substance was created. So must we likewise from experience of every kind first endeavor to discover true causes and axioms; and seek for experiments of Light, not for experiments of Fruit. For axioms rightly discovered and established supply practice with its instruments, not one by one, but in clusters, and draw after them trains and troops of works.[36]

1.3 An Aside on Philosophical and Scientific Realism

Like other philosophical isms, the term "realism" covers a variety of sins.
- Larry Lauden [37]

Philosophers are continually reconsidering the most basic questions of their trade. This can be disconcerting for scientists adrift on the philosophical tide. Among the perennial questions in the philosophy of science is that of realism. At the most basic level, the question is whether anything exists independent of human perception and thought, and if so, how that might be established. Those who answer that nothing is real, and that we only entertain mental constructions, are known as anti-realists, at least with regard to entities. Most scientists will find this a bizarre notion, since realism with regard to entities is fundamental to our notion of objective science. The manipulation of entities in supposedly reproducible experiments typically presupposes a bluntly realist attitude toward those entities.[38] No less than Albert Einstein, always interested in the philosophy of science, did not waste any time in repudiating subjective idealism, a form of anti-realism that takes the outer world to be a derivative of consciousness: "No physicist believes that. Otherwise he wouldn't be a physicist... Why would anybody go to the trouble of gazing at the stars if he did not believe that the stars were really there?"[39]

A question scientists should find more engaging concerns the relation between scientific theory and the physical world. Although we will take realism with regard to entities as a given, realism with regard to theories is more problematical. Realists in this area believe 1) that laws and mature theories, particularly as expressed in mathematical form, actually correspond to certain physical realities; or, 2) that current formulations that are recognizably incomplete or inaccurate will eventually be improved and will converge to something that can be considered the truth. For example, Bacon's unfettered realism is implicit in the earlier discussion of his cumulative theory of knowledge; he took the proliferation of thoughtful experimentation as more or less equivalent to gains in knowledge, allowing ever grander statements about nature. Bacon's realism is made explicit in the following statement:

> For I am building in the human understanding a true model of the world, such as it is in fact, not such as a man's own reason would have it to be; a thing which cannot be done without a very diligent dissection and anatomy of the world ...Truth therefore and utility are here the very same things.[40]

Anti-realists with regard to theories (or, *instrumentalists*), on the other hand, view such "laws" as only working theories that function more or less reliably for prediction, technological advances, etc.; workable theories are not equated with the truth, and no final convergence to the truth is expected (possible).[41] Realists hold to the shrinking or extinction of the distinction between theory and reality as a good definition of scientific progress. Anti-realists are content with comparatively more effective tools. The issue of convergent realism in hydrology will be at least implicit throughout our discussion of methodology in hydrology, and will be explored explicitly in Chapter 5 in connection with the discovery and derivations of Darcy's law.

Naturally, there are stronger and weaker versions of both realist and anti-realist positions. For our purposes, the realism debate turns on the *purpose* of scientific investigations: 1) convergence, however slow and difficult, to a true description of natural processes and a one-to-one correspondence between theory and reality; or 2) non-convergence, stemming from a recognition that such a correspondence is not only unlikely but impossible, and resulting in a effort to achieve a reasonably good fit between theory and reality in order to reliably support applied science.

1.4 Motivation for a Study of Methodology

The realist/anti-realist debate reflects very distinct orientations toward scientific progress. The confusion of these positions particularly plagues discussions of the validation of hydrologic models. Hydrologists might reply that none of this matters - that final statements deferred indefinitely (realism) might as well be final statements denied (anti-realism). But in that case, they are not speaking prose, they are speaking nonsense. The different approaches to validation necessarily entail positions on realism, and therefore a coherent position on realism will focus hydrologists' role in deliberations on such issues as the siting and operation of nuclear waste repositories. Realism with regard to theories requires direct comparisons of theory and reality. Anti-realism with regard to theories obviously implies that there can be no strict validation, since no final correspondence is expected. Thus the anti-realists are more inclined to stress the comparative nature of theory acceptance, since for them there is no final solution.

A stable framework within which to consider such questions would be a useful thing, since the questions hydrologists typically address have changed radically within a single generation. Traditional hydrologic concerns were and are dominated by the solution of reasonably tractable water supply problems. Thus Freeze and Cherry, writing in 1979, preface their influential textbook *Groundwater* with the following comment:

> We perceive a trend in the study and practice of groundwater hydrology. We see a science that is emerging from its geological roots and its early hydraulic applications into a full-fledged environmental science... If this book had been written a decade ago, it would have dealt almost

> entirely with groundwater as a resource... But groundwater is more than a resource. It is an important feature of the natural environment; it leads to environmental problems, and may in some cases offer a medium for environmental solutions.[42]

In contrast to traditional hydrologic applications, contaminant hydrology is an applied science that has come of age in a more regulated and litigated context. At the same time, the public prominence of water supply issues has intensified in the West. It is not imagined that hydrologists ever operated within a halycon independence. Current trends, however, emphasize less and less problems suggested by some internal logic of the science; agenda are instead set by pressing engineering problems dictated by social concerns. Research directions and even theory validation methods have become entrained in a politically and socially embattled engineering program.[43]

Hydrology often lacks anything resembling the autonomy that Thomas Kuhn (writing in 1962, mostly about physics) claims is the normal situation in science. Kuhn suggests that the more usual autonomy of science is responsible in part for a "very special efficiency":

> Just because he is working only for an audience of colleagues, an audience that shares his own values and beliefs, the scientist can take a single set of standards for granted. He need not worry about what some other group or school will think and can therefore dispose of one problem and get on to the next more quickly than those who work for a more heterodox group. Even more important, the insulation of the scientific community from society permits the individual scientist to concentrate his attention upon problems that he has good reason to believe he will be able to solve. Unlike the engineer, and many doctors, and most theologians, the scientist need not choose problems because they urgently need solution and without regard for the tools available to solve them.[44]

While the loss of relative autonomy in hydrology is variously viewed as either beneficial or detrimental to both the science and to long-term public interests, it certainly calls for an coherent research philosophy.[45] Hydrologists are habitually constrained or impelled by external considerations and inclined to view at least some of these as impediments to good science. As a result, they should be professionally sensitized to the need for public accommodation, explanation and defense; hence, for an articulation of their goals and methods. This requires, among other things, that the rationale for procedures and decisions be made intelligible and compelling to outsiders.

Renewed federal emphasis has lately been placed on "strategic research". Strategic research is intended to directly and promptly address specific problems of high priority, however these are defined. As a result, nationally defined goals for academic research are being proposed that can be expected to reach across all disciplinary boundaries.[46] Both funding and procedural decisions might well be reconsidered in light of the growing impatience with environmental remediation efforts.[47] The new state of affairs has been protested by various researchers, who generally doubt both the long- and short-term efficacy of this orientation. A case might be made that this trend is, as Bacon put it, an example of stepping aside for *Atalanta's golden apples* and losing sight of the ultimate goal.[48] It is not the intention here to debate this policy decision directly, only to note that the current funding climate ensures that hydrologic justifications will come under closer scrutiny than heretofore. The situation again calls for a certain self-examination by partisans.

Hydrologists have traditionally shown little interest in the philosophy of science, a field with a great deal to say about methodology. In this they have exhibited the same sound intuition as most scientists; the greater bulk of the writing in the philosophy of science is not directed to their edification, nor toward a description of how science actually proceeds. Philosophers' interest in science is often secondary to their interest in how knowledge has, does or could progress.[49] They are fascinated by science as the purest form of truth-seeking or knowledge-gathering available for their inspection; one, moreover, in which progress is evident. Thus Ian Hacking can say of Lakatos: "He thought science was our model for objectivity".[50] The conclusions drawn by philsophers, however, have traditionally been largely irrelevant to the immediate concerns of scientists.

The claim that most philosophy of science does not describe real science can be dangerously self-defeating in a regulatory and normative context. The fact that the argumentative form (methods) of science appears generally successful by almost any definition[51] has not inhibited outsiders from engaging in continual criticism; some of these commentators appear determined to poke as many holes as possible in the traditional picture of science as objective and progressive. The situation within applied groundwater modeling, a less uniformly successful enterprise, does nothing to deflect this criticism. If hydrologists do not operate according to either the theoretical structures or the methodologies discussed by favored philosophers, perhaps - some will argue - they should. These are not only academic conflicts. Expanded avenues for public and academic involvement in environmental decisionmaking have attracted a significant contingent from the philosophical community who have injected new arguments into old debates over research, funding and policy.[52] These complicating pressures are among the most powerful arguments in favor of this explicit investigation of methodology in hydrology.

In their discussion of model validation, Konikow and Bredehoeft state baldly that "society's actions will be based upon our professional judgements."[53] This is not wholly, necessarily, nor permanently the case. Loss of influence due to input from non-scientific parties might affect everything from the micro-scale of individual funding decisions to the macro-scale of siting decisions for waste disposal sites. A coherent description of hydrologic methodology and its relation to progress has an important place within such debates.

Medawar's burlesque with which we opened this discussion captures the seeming lack of relevance of the philosophy of science to science practitioners. The situation occasionally prompts observers to suggest that the real need may be to educate outsiders on how science operates; an expanded lexicon and perspective can facilitate these conversations. Hydrologists would do well to be able to make their cases in the present regulatory and funding atmosphere. The standard indifference of scientists has largely left the argumentative field to those concerned with the nature of argument and truth abstracted to some degree from practice (philosophers), or to those naturally more influenced by public expediency (legislators and regulators). That this is not a new situation is evidenced by a comment of H.L. Fairchild in 1904, noted approvingly by Vic Baker: "Geologists have been too generous in allowing other people to make their philosophy for them".[54] In what follows, the description of methodology is addressed to the imbalance of these debates - an imbalance that seems particularly misguided given the applied nature of hydrologic science. As Thomas Kuhn observed, it is in times of crisis that scientists may find the philosophy of science most edifying. There are many kinds of crises.

Some of Medawar's comments reinforce the common opinion that all such considerations may be largely academic within the scientific community. A survey of current events suggests, however, that if hydrologists are interested in defending or defining the extra-scientific margins of their practice, they cannot remain indifferent to these issues. Furthermore, and more positively, if "science is what scientists do", then a credible description of methodology within hydrology can actually be useful within the community. Our own discussion, while recognizing and engaging a wide-ranging series of observers on the margins of science, will remain firmly anchored in hydrologic practice. The role of models is a unifying theme in the description of what hydrologists do.

1.5 Models

Modern hydrologic investigations might incorporate any or all of the three principal methodologies sketched in the opening section; models play essential roles within both unalloyed and hybrid applications of causal, Baconian or crucial methods. Since model is a term used in so many conflicting ways, it is best to define immediately what is meant by the term here: models are geometrically defined spaces, within which certain properties hold and certain processes operate. They are designed to be analogous, if not identical, to prototypal physical systems of interest. The role of models is thus a co-axial issue with realism, since idealization necessarily introduces a distinction between model and prototype that realists will find at least occasionally troublesome.

An early objective in the construction of hydrologic models is to reduce the complexity recognized in the prototype. This filtration is necessary due to 1) the complexity of materials and processes; 2) limited appreciation of the composite effect of these materials and processes; 3) inadequate measurement and characterization capabilities; and 4) insufficient numerical capacity for the solution of more inclusive problem statements. Hydrologic models can therefore be treated as the result or goal of idealization exercises.

Models differ in the nearness of their connection to specific practical concerns. *Internal* or *process-oriented* models can be usefully contrasted to *external* or *directly applied* models. The internal pursuits of theoreticians and experimentalists emphasize understanding generic processes through detailed examination of modeled relationships between possibly contributing factors. Therefore process-oriented researchers typically make use of controlled experiments, either in theory or in practice, to isolate the effects of materials or processes of interest. These investigations might be undertaken as formal mathematical exercises, as bench-scale laboratory studies, or as numerical simulations. Control of all extraneous factors is the essential component. A successful process model is one that captures, compares, contrasts or quantifies the interplay of forces and conditions in a revealing way under specified conditions. The internal logic of the science is generally brought to bear to determine the degree of "success", a procedure that will be considered later.

The process-oriented researcher is thus more interested in *functions* that summarize *phenomena* than in *parameters* that merely reflect the details of case-specific *data*. These phenomena are potentially recognizable - and these functions potentially useful - in many applications. The details that might allow theoretical advances and experimental insights to be applied to specific external problems are not, however, of particular interest. Application-specific parameter values are considered inappropriate to fundamental research for several reasons.[55] First, the process-oriented models employed may bear only the most superficial resemblance to any actual physical system; how such models might still be revealing will be explored below in a discussions of Darcy's law (Chapter 5) and solute transport through fracture junctions (Chapter 6). Secondly, it is often the case that the called-for numbers are known to be simply unmeasurable in the current state of affairs; characterization of flow through porous or fractured media is generally incomplete in terms of both the materials involved and the processes operative within them. Lastly, and most importantly, the goal of process-oriented research is not to solve particular problems, but only to advance general insight into possibly important processes and their interrelationships. It is, therefore, from the outset a separate and usually secondary consideration whether the model might ever be directly utilized in a field problem. Thus the parameter values necessary for field applications are generally not considered part of a process-oriented model, since these values are functions of not only the operative processes but also of the materials present.[56]

Process-oriented certainly does not mean *pure* in the sense of near total abstraction from practical concerns; it should be taken to mean *application-deferred*. In an applied science like hydrology, there is little room (and certainly no funding!) for research totally abstracted from practice. Despite the tangential interest of process modelers, applied modelers utilize available generic process insights to augment site-specific information in order to reach conclusions with practical implications. The phrase *directly applied* will be used occasionally in describing the latter efforts, in an effort to re-emphasize the connection (remote as it can be) between the indirectly applied process models and site-specific investigations. Clearly, some field sites will be influenced by multiple, perhaps poorly understood factors that cannot be controlled, in which case individual theoretical and experimental process insights may not be readily available nor easily imported.

Directly applied modeling is more concerned with site-specific problem-solving than with theoretically intriguing but currently impractical research. Such models are frequently used for prescriptive or predictive purposes. As a result, field parameters and other site-specific factors largely by-passed in internal research often assume critical importance in applied models. Accordingly, the term model is properly broadened to include not only the structure and governing equations, but also the measured or estimated field parameters; in short, *the entire body of geometry, processes, parameters, approximations and assumptions that contributes to model results*. As will be seen below, this broader definition is not universally accepted. The organizing principle adopted here is to identify the *conceptual model* with all of the uncertain or idiosyncratic decisions that are essential to the modeler's conclusions. In the end, we believe this principle can relieve a certain amount of confusion in the discussion of model validation. A small but important distinction will be maintained between models and theories, since the latter also incorporate fundamental laws that are neither uncertain nor idiosyncratic. Within theories, models are combined with relevant fundamental laws. Thus process modelers require only arbitrary structures and selected physical, chemical and/or biological processes (and fundamental laws); predictive applied modelers resort to arbitrary structures, assumed processes, estimated parameter values (and fundamental laws).

We began by identifying four methods by their distinctive argumentative structures, and then reduced the field to Baconian, causal and crucial inquiries. Corresponding models can be categorized by their method of construction. The dependence of method upon model is so complete that our discussion of method in science is essentially a discussion of 1) how models are chosen; 2) the different and sometimes complementary roles the three types of models can play; and 3) how the appropriateness of models might be judged. We will treat method as essentially a guide to the construction, interaction, evaluation, and succession of models. A few preliminary comments can be made at this time.

First, in the case of Baconian induction, the model emerges as a generalization of the vagaries of data and illustrates basic relationships between certain physical factors; in short, the model is the *result* of generalizing-*upward* inquiry. The robustness of an inductive model depends mostly on the available data and the severity of the controls under which they were collected. The historical details of Darcy's law (Chapter 5) and an account of work on solute mixing at fracture junctions (Chapter 6) illustrate the satisfying use of such models in the discovery and corroboration of hydrologic processes. Secondly, despite Bacon's concerning deduction, causal arguments have two legitimate uses in the view espoused in Chapter 5; models are employed in 1) semi-rigorous confirmations that empirical results "make sense" under the larger umbrella of established fundamental laws; and 2) *a priori* indications of plausibly expected behavior, a guide that points toward needed empirical confirmation of theoretically derived expectations. The added uncertainty associated with predictive causal models unsupported by direct experiment is examined in Chapters 2, 3 and 4. Thirdly, crucial experiments may serve at any point as either arbiter between competing theories and their models, or as a further indication of profitable research directions; this view is elaborated in Chapter 6 and in the concluding Chapter 7.

The central argument of this thesis is concerned with the form that a profitable symbiosis between different aspects of causal, Baconian and crucial models might take; or, in examining what Henri Poincarè said: "Logic, which alone can give certainty, is the instrument of demonstration; intuition is the instrument of invention".[57] Events will show that the situation is not so simple, but more creative. The main thing that separates hydrologists from, say, paleontologists - a group that struggles more or less privately with procedural and epistemological issues in the historical sciences - is the intensity of public interest in the results of scientific inquiry. Accordingly, we first take up certain aspects of these practical affairs, and consider the prospects for what Bacon called *Experiments of Fruit*.

Predictive groundwater models are often the most visible public displays of the current state of the art in hydrology. Hydrologists have perhaps been too confident over the short-term capabilities of applied hydrologic modeling, and too sanguine over the prospects for long-term progress. Grayson, *et al.*, consider the performance of their own physically-based (surface water) basin-scale model, and comment on the prevalence of public optimism:

> With the profusion of hydrologic models linked to geographic information systems (GIS), there is an extensive literature that answers the question... 'How many hydrologists have shared such unrealistic expectations?" The answer is a resounding "too many". Similar problems are arising in related fields as the forthcoming NATO Advanced Research Workshop entitled "Ecosystem Modelling - Delineating the Possible From the Impossible" attests.[58]

What strikes some as grossly opportunistic over-promising may only reflect the genuine if unschooled optimism of certain modelers. Neither deception nor naivetè is likely to fare well in the present scrutiny of environmental science; delineating the reasonable from the unreasonable is a worthwhile project. The search for effective methodological rules and theory evaluation protocols motivates in part the efforts of hydrologists to find some firm philosophical ground. This ground will be seen to lie in a careful consideration of the roles played by experiment, model and theory.

1.6 Notes:

1. Kuhn, Thomas S. (1962), *The Structure of Scientific Revolutions*, University of Chicago Press, 2nd edition (1970), p.88.

2. See, *eg.*, National Academy of Sciences (1990b) (O'Melia, C.R., chair): *Keeping Pace with Science and Engineering, Case Studies in Environmental Regulation*, National Academy Press, which begins by noting that "While technical understanding generally will not exclusively determine the outcome of the regulatory process, this information is central to assessing the risks and devising alternative mitigating strategies, if any, from which the decision-making process may choose. No one would argue that environmental regulation, if not determined by the best scientific and engineering understanding, should not at least be based on it" (p.*v*).

3. An overview of public concerns can be found in the *New York Times*, 5 May 1992: C,1:1. V. Klemes has thrown down the gauntlet in several places. See, *eg.*, Klemes, V. (1988a) "Dilettantism in Hydrology: Transition or Destiny?", in *Water Resources Research*, **22**:9, pp.177s-188s; and Klemes, V. (1988b), "A Hydrological Perspective", in the *Journal of Hydrology*, **100**, pp.3-28. Earl Bardsley, "A Moribund Science?", *New Zealand Journal of Hydrology*, **31**:1, pp.1-4, writes of a "current malaise" in which "a whole vocabulary of dishonesty has arisen to mask these dubious modeling endeavors with a cloak of scientific respectability" (p.3). Several essays in O'Kane, P. (ed.) (1992), *Advances in Theoretical Hydrology: A Tribute to James Dooge*, Elsevier, address different constructions of hydrology and hydrological issues. The authors reveal both epistemological and organizational anxieties: there are doubts voiced about the relevance of hydrological theory to specific technological demands, in regard to both surface water and groundwater. In addition, the interdisciplinary vs. multidisciplinary debate is explored. Essays of interest include those by Nash, J.E., "Hydrology and Hydrologists - Reflections; Philip, J.R., "Hydrology in the Real World"; and Abbott, M.B., "The Theory of the Hydrologic Model, or: The Struggle for the Soul of Hydrology". Some of these same authors participated in an IAHS/UNESCO panel on hydrologic education issues - Nash, J.E., Eagleson, P.S., Philip, J.R. and Van Der Molen, W.H. (1990), "The Education of Hydrologists", in *Hydrological Sciences Journal*, **35**:6, pp.597-607: "There is a widespread unease among hydrologists that in the past forty years the practice of hydrology as a technology for the solution of the problems arising in water resources development has not progressed adequately. Most of the problems recognized at that time remain with us today and, when they arise in individual cases, their solution is usually approached in an empirical way drawing little on the principles of natural science. Perhaps because of the failure of technologists to found their technology on a sound scientific basis, the science too has been slow to develop -thus in a sense seeming to justify its neglect by the technologist. It would seem that this vicious circle can be broken only by a thorough review of the science and technology of hydrology and a reassessment of the manner of education of hydrological scientists and technologists" (pp.598-599).

4. See, *eg.*, Burges, S. (1993), text of an address to the UNNY Symposium, University of Waterloo, Waterloo, Ontario, July 1993; also National Academy of Sciences (1990a), (Schwartz, F.W., chair): *Ground Water Models: Scientific and Regulatory Applications*, National Academy Press, p.80; also Rogers, P.P. (1983): book review of *Analyzing Natural Systems*, in *EOS*, **64**:25, p.419.

5. Anderson, M.P. (1983), "Ground-Water Modeling - The Emperor Has No Clothes", in *Ground Water*, **21**:6, pp.666-669. Anderson is currently president of the Hydrology Section of the American Geophysical Union (AGU).

6. Medawar, Peter B. (1969), *Induction and Intuition in Scientific Thought*, American Philosophical Society, p.11.

7. McMullin, E. (1985), "Galilean Idealization", in *Studies in the History of the Philosophy of Science*, **16**:3, Pergamon Press Ltd., p.249-250.

8. Medawar, P.B. (1969), p.29.

9. Medawar, P.B. (1969), p.9.

10. Bacon was a lawyer. The various readings of Bacon are summarized in Bacon, F. (1620b*), *Novum Organum*, Urbach, P. and Gibson, J. (eds. and trans.), Open Court (1994), pp.xiii-xiv: "Bacon's reputation as a man and as a philosopher has fluctuated dramatically. After his near deification in the 17th century as creator of the new experimental approach to science, he was largely eclipsed in the eighteenth, scarcely mentioned by the great philosophers of that era; and then the 19th century saw an enthusiastic revival of scholarly interest in Bacon". Urbach and Gibson then detail some of the more prominent and unusual readings of Bacon. Another source for perspective on Bacon is Hamlin, C. (1990), *A Science of Impurity: Water Analysis in 19th Century Britain*, University of California Press, pp.61-65: "Mineral Analysis was Baconian in a number of aspects. First, it relied on an army of fact-gatherers... Second, mineral water chemists saw their enterprise as Baconian inasmuch as they equated scientific progress with more facts... No conceptual rearrangement was foreseen... Third, [they] saw themselves as building a foundation for technical and medical progress... Fourth,...analyses were intended to contribute knowledge that both described... and provided a basis for generalization... Finally... a part of the Baconian ethos was that all information was grist for the mill..."(p.62-63). The accuracy of this reading of Bacon can be judged by the material in Section 1.1, below.

11. Bacon, Francis (1620b), *Novum Organum*, in Burtt, E.A. (ed.) (1939), *The English Philosophers From Bacon to Mill*, The Modern Library, cited in Kuhn, T.S. (1962), p.18. The Modern Library edition (1935) does not contain the entirety of the second book of *Novum Organum* aphorisms. Simple Roman numerals refer to the number of an aphorism in the first book; Pagination refers to the Modern Library Edition of 1939. When an aphorism is in the second book, it is preceded by an Arabic *2:*; pagination given is then from Bacon, F. (1620b*), Urbach and Gibson (eds. and trans.), Open Court (1994).

12. Bacon, F. (1620a), *The Great Instauration*, preface, in Burtt E.A. (ed.) (1939), *The English Philosophers From Bacon to Mill*, The Modern Library, p.5. "Instauration" means restoration after decay.

13. Bacon, F. (1620b), *lxx*, p.48. See also *c*, p.69-70: "For experience, when it wanders in its own track, is, as I have already remarked, mere groping in the dark, and confounds men rather than instructs them. But when it shall proceed in accordance with a fixed law, in regular order, then may better things be hoped of knowledge"; or *cxxv*, p.83.

14. Bacon, F. (1620a), p.20.

15. See, *eg.*, Bacon, F. (1620b), *cxvii*, p.77.

16. See, *eg.*, Bacon, F. (1620b), *xix*, p.31; or *civ*, p.71.

17. Bacon, F. (1620b), *lxxxvi*, p.61: "As the matter now is, it is nothing strange if men do not seek to advance in things delivered to them as long since perfect and complete".

18. Bacon, F. (1620a), preface, p.7.

19. Quinton, A. (1980), "Bacon", in *Renaissance Thinkers*, Oxford University Press (1993), p.146: "Bacon is the most confident, explicit and influential of the first exponents of the idea of progress. J.B.Bury, the historian of that idea, could find only slight and soon obliterated traces of the belief in progress before Bacon: a momentary scintillation in Democritus that was not perceived by his Epicurean successors, a rather stronger hint in Bacon's medieval namesake, Roger [Bacon]. Bacon's progressivism is the outcome of two strains in his thought. The first of these is his more or less unprecedented notion of knowledge as cumulative. The second is his insistence that knowledge is for practical use, specifically for 'the relief of man's estate'".

20. Bury, J.B. (1932), *The Idea of Progress: An Inquiry into its Origin and Growth*, Dover Publications, p.53. Bury is paraphrasing Bacon, F. (1620b), *lxxiv*, p.52: "For what is founded on nature grows and increases; while what is founded on opinion varies but increases not. If therefore those doctrines had not plainly been like a plant torn up from its roots, but had remained attached to the womb of nature and continued to draw nourishment from her, that could never have come to pass which we have seen now for twice a thousand years; namely, that the sciences stand where they did and remain almost in the same condition; receiving no noticeable increase, but on the contrary, thriving most under their first founder, and then declining. Whereas in the mechanical arts, which are founded on nature and the light of experience, we see the contrary happen, for these (as long as they are popular) are continually thriving and growing, as having in them the breath of life; at first rude, then convenient, afterwards adorned, and at all times advancing".

21. Bacon, F. (1620a), preface, p.10-11.

22. Bacon, F. (1620b), *lxx*, p.48: "But the best demonstration by far is experience, if it go not beyond the actual experiment. For if it be transferred to other cases which are deemed similar, unless such transfer be made by a just and orderly process, it is a fallacious thing". His views on how the Scholastics "mixed things human and divine quite improperly" can be found in *lxxxxix*, p.63.

23. Bacon, F. (1620b), *xxv*, p.32: "The axioms now in use, having been suggested by a scanty and manipular experience and a few particulars of the most general occurrance, are made for the most part just large enough to fit and take these in: and therefore it is no wonder if they do not lead to new particulars [implications about other situations that might be checked]". See also *lxxiii*, p.51: "Now, from all these systems of the Greeks, and their ramifications through particular sciences there can hardly after the lapse of so many years be adduced a single experiment which tends to relieve and benefit the condition of man, and which can with truth be referred to the speculations and theories of philosophy. And... the experimental part of medicine was first discovered, and afterwards men philosophized about it, and hunted for and assigned causes..."

24. Bacon, F. (1620b), *lxxxii*, p.57. See also the description in *cv-cvi*, p.71.

25. Bacon, F. (1620b), *cv*, p.71.

26. Bacon, F. (1620b), *cvi*, p.72.

27. Bacon, F. (1620b), 2:*vi*, p.92.

28. Bacon, F. (1620b), *cxix*, pp.78-9.

29. Bury, J.B (1932), p.54. Those interested in how knowledge progresses might find it interesting that Bury is echoing a much older sentiment. Writing adaptations of Greek plays around 250 B.C., Plautus says in "Trinummus", line 367, *Non Aetate, verum ingenio apiscitur sapientia*, or: "Wisdom is acquired by character, not by age".

30. Bacon, F. (1620a), p.21.

31. Quinton, A. (1980), "Bacon", in *Renaissance Thinkers*, Oxford University Press (1993), p.118.

32. Bacon, F. (1620b), *cxvii*, p.78.

33. Bacon, F. (1620a), preface, p.13. Bacon considered the *finis scientiarum* (the goal of science) to be "the endowment of human life with new inventions and riches", and evaluated the worth of the different efforts on this basis.

34. Bacon, F. (1620b), *lxx*, p.49. The "enlarging" of the inquiry is another reference to *true induction*, as described earlier. See also *ciii*, p.70: "Our road does not lie on a level, but rises and falls; first rising to axioms, then falling to works".

35. Quinton, A. (1980), p.122.

36. Bacon, F. (1620b), *lxx*, p.49.

37. Lauden, L. (1984), *Science and Values: The Aims of Science and Their Role in Scientific Debate*, University of California Press, p.105.

38. See the discussion of this point, and the possible contrast between theorists and experimenters, in Hacking, I. (1983), *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*, Cambridge University Press, pp.262-266. Hacking contrasts the psychology of theorists and experimenters, suggesting that the latter are most naturally inclined to realism about entities because they *manipulate* them.

39. Planck, Max (1933), *Where is Science Going*, The Humanities Press, pp.212-3 (Epilogue). Scientists are of course free to regard philosophy as intriguing, irritating or incomprehensible by turns. It should be noted, however, that at times the most celebrated of scientists, such as Max Planck, have been seriously engaged in the development of the philosophy of science. For example, Ernst Mach and Ludwig Boltzmann each held the chair for the philosophy of the inductive sciences at the University of Vienna. Albert Einstein was given to prefacing his technical papers with this sort of thing (see, *eg.*, the beginning of the 1905 paper on special relativity). In the 1920s, the so-called Vienna Circle brought together some of the great names in philosophy and science. The Circle included Riemann, Helmholtz, Hertz, and the mathematicians Poincaré and David Hilbert.

40. Bacon, F. (1620b), *cxxiv*, p.82. Bacon's view of gains in knowledge as something fated is alluded to in *xciii*, p.66.

41. See, *eg.*, Hacking, I. (1983), p.63.

42. Freeze, R.A. and Cherry, J.A. (1979), *Groundwater*, Englewood Cliffs: Prentice-Hall, pp.*xv*-2.

43. See Chapter 2, below, "Dark Grey Boxes"; also see Shrader-Frechette, K.S. (1993), *Burying Uncertainty: Risk and the Case Against Geological Disposal of Nuclear Waste*, University of California Press.

44. Kuhn, T.S. (1962), p.164.

45. There is a vast literature in the area of Science and Technology Studies, for example, that generally treats the "democratization" of science as an important political goal. This position implies not only educational reform, but a general demystification of science with a reduced reliance on experts,

and a greater reliance on an informed public to set goals and evaluate progress.

46. Many articles and letters in *EOS*, 1994-1995 address this issue within the geophysical sciences. Also, "Roundtable: Physics in Transition", in *Physics Today*, **46**:2, February 1993, p.46, contains the following discussion. "[Judith Bostock]: 'Congress believes what scientists say they can do. We know that science can't work the miracles that are promised. The day Congress figures that out for itself, we are in deep trouble... The Clinton Administration has made it clear that it plans to shift money in the science budget from fundamental research to applied research as a way of stimulating industrial innovation and productivity. American scientists will have to live under such strictures. It is unrealistic to think that scientists will be able to say, as they did in the past, "Just give us the funds and let us decide what to do"'. [Richard Zare]: 'Judy, when you talk about productive [science], are you talking about just end results... or are you talking about relevance to society?' [Bostock]: 'Relevant to society. That is what Congress is beginning to define as productive science'. [Daniel Kleppner]: 'What you're talking about is directed science'. [Bostock]: 'Yes'."

Besides talk of re-directing research, budget analysts also expect a significant scaling back of research support. See, *eg.*, "Budget Update: Future is Looking Gloomier for Science", in *EOS*, **76**:14, 4 April 1995; and Irwin Goodwin (1995), "Clinton's R&D Budget Defers Pain to Unkindest Cuts by Republicans", in *Physics Today*, **48**:4, April 1995, provides both a detailed research budget for the current year and proposals for the future. Although the situation is highly uncertain, the analysis suggests that if House budget leaders have their way: "DOE's research programs are almost sure to be whacked, possibly by as much as $1.2 billion. The most likely candidates for the chopping block: environmental cleanup and rehabilitation, magnetic fusion, renewable energy sources and high-energy physics" (p.67). More generally, "government spending on R&D is likely to be reduced by as much as 25% over the next five years" (p.67) according to George E. Brown, previous chair of the House Committee on Science, Space and Technology, and generally considered a strong supporter of basic science initiatives. This article also includes line item information from the National Science Foundation budget.

47. See, *eg.*, Albuquerque Journal, 1 Feb 1995.

48. The connection between personal motivation and public results also tends to get trampled in such discussions. In Planck, M. (1933), p.211, Albert Einstein says: "What the scientist aims at is to secure a logically consistent transcript of nature. Logic is for him what the laws of proportion and perspective are to the painter, and I believe with Henri Poincare that science is worth pursuing because it reveals the beauty of nature. And here I will say that the scientist finds his reward in what Henri Poincaré calls the joy of comprehension, and not in the possibilities to which any discovery of his may lead". Practitioners of an applied and constrained science will not find a very sympathetic climate for these sentiments, however important they may be as personal motives.

49. Note Lauden, L. (1984), p.40: "What is remarkable is that, despite massive evidence to the contrary, some philosophers of science still tend to imagine that methodology is protoscientific, that is, prior to and independent of the kind of empirical inquiry which science itself represents". On the other hand, Flew, A. (1979), *A Dictionary of Philosophy*, St. Martin's Press, p.319-320, notes straight-forwardly: "Organized empirical science provides the most impressive result of human rationality and is one of the best accredited candidates for knowledge. The philosophy of science seeks to show wherein this rationality lies; what is distinctive about its explanations and theoretical constructions; what marks it off from guesswork and pseudo-science and makes its predictions and technologies worthy of confidence; above all whether its theories can be taken to reveal the truth about a hidden objective reality".

50. Hacking, I. (1983), p.112. See also Popper, K.R. (1959), *The Logic of Scientific Discovery*, Routledge (1992).

51. Such diverse observers as Thomas Kuhn and Karl Popper agree that science is progressive almost by definition. See, *eg.*, Kuhn, T.S. (1962), pp.162-167: "... the phrases 'scientific progress' and even 'scientific objectivity' may come in part to seem redundant... Does a field make progress because it is a science, or is it a science because it makes progress?" (p.162); see also, Popper, K.R. (1959), p.49, where Popper describes himself as part of a group who "take it as their task to analyze the characteristic ability of science to advance"; or, Popper, K.R. (1963), *Conjectures and Refutations: The Growth of Scientific Knowledge*, Routledge, p.216: "The history of science, like the history of all human ideas, is a history of irresponsible dreams, of obstinacy, and of error. But science is one of the very few human activities - and perhaps the only one - in which errors are systematically criticized and fairly often, in time, corrected. That is why we can say that, in science, we often learn from our mistakes, and why we can speak clearly and sensibly about making progress there".

52. See, *eg.*, Shrader-Frechette, K.S. (1993), pp.103-59, which concludes: "If the reasoning in this chapter is correct, then it is impossible, with our current data base, for proponents of permanent, geological disposal to avoid three invalid inferences in scientific and ethical reasoning: the appeal to ignorance, affirming the consequent, and the consent inference. Consequently, our conclusion dictates a change in policy..." (p.159).

53. Bredehoeft, J. and Konikow, L.: "Ground-water Models: Validate or Invalidate?", in *Ground Water*, **31**:2, Mar-Apr 1993, p.179. The authors also suggest that the whole idea of validation as commonly written into regulations is an outmoded notion left over from the pre-Popper Positivists. On a more philosophical level, note Lauden, L. (1984), p.54: "Indeed, it has sometimes been argued, although it remains a notorious point of contention, that an agent has no more priveleged access to what his goals are than do those third parties who study carefully his overt behavior."

54. Baker, V. (1993), in *GSA Today*, Nov 1993, pp.285-7: book review of Schumm, S.A. (1991), *To Interpret the Earth: Ten Ways to be Wrong*, Cambridge University Press. The comment from Fairchild is in the *Geological Society of America Bulletin*, 1904.

55. It should be noted that the coefficients said to be of little interest in process research are the site-specific properties and parameters, such as hydraulic conductivities, dispersion coefficients, etc., *not* the geometric coefficients that are *general* and appear, for example, when Poiseuille flow between parallel plates is integrated to yield the "cubic law" for mean velocity: $\bar{v} = (b^2/12)(g/\nu)(\partial\phi/\partial x)$. See Chapter 6 for an application of this law in process research.

56. It may be of some interest that Galileo "never wrote an equation in his life". In constructing his physics of motion, he instead relied on relational observations. As a result, "the physical constants that loom so large in our thinking simply cancell[ed] out in proportionalities of the kind that Galileo used exclusively in his physics". The advantages and disadvantages of this approach are discussed in the introduction by Stillman Drake to Galileo Galilei, *Two New Sciences*, Wall and Thompson (1974), pp.*xxx-xxxii*.

57. Poincare, H. (1905), *The Value of Science*, Halsted, G.B. (trans.), p.23.

58. Grayson, R.B., Moore, I.D. and McMahon, T.A. (1994), "Reply", in *Water Resources Research*, **30**:3, pp.855-856.

Part One:

Experiments of Fruit

Keep reason under its own control.

- Marcus Aurelius, *Meditations*: IX,7

2

Dark Grey Boxes:

Uncertainty in Applied Groundwater Modeling[1]

It was, I think, one of Pirandello's characters who said that a fact is like a sack - it won't stand up until you've put something in it.

- E.H. Carr: *What is History?*[2]

2.0 Modeling: Basic Choices

Applied groundwater modelers adopt a causal model approach: specified conditions and laws of interaction lead to predictions of expected behavior. Within the literature of the field, modelers are less voluble about this fundamental choice and more so about the details of impediments such as parameter estimation. Modelers of surface water, on the other hand, generally do not use causal models, preferring empirical statistical correlations between measured inputs and outputs. Recent promotions of causal models for surface flows have, therefore, sparked a revealing methodological debate. Since surface water modeling is arguably in a greater state of flux than even the modeling of groundwater, the methodological issues stand out in high relief; they can be surveyed with profit by groundwater scientists.

Without minimizing the difficulties faced by surface water modelers, some of the contrasts between their efforts and those of groundwater modelers should be briefly acknowledged. The surface water modeler 1) can more likely measure whatever he thinks is important in the system; 2) can more likely directly observe events and results relevant to his model; 3) has relatively rapid feedback from what might be considered continuous model evaluations, such as seasonal floods; and therefore 4) tends to rely on these empirical correlations, rather than causal models; but 5) models a "more open" system in which meteorological and atmospheric physics play large roles.

The reliance on observations and measurements of system input and output to construct probable empirical relations is the essence of black box hydrology. Practitioners make relatively few claims to understanding the system, and instead measure what are hoped to be reliable indicators of future system response. A simple example would be rainfall-runoff relations. There may be wide disagreement among modelers over what the indicators ought to be in a given basin - and over the specific relationship of inputs to outputs - yielding a plethora of potential empirical relationships. As a result, Vit Klemes championed the development of causal models in surface water hydrology:

> The unifying framework for all these *ad hoc* models can hardly be anything else than a physically consistent model of the catchment mechanisms, i.e., a causal theory of the hydrologic cycle... This line of inquiry seems to be the most promising way out of the unenviable present situation that has been aptly compared (Dooge, 1978) to "a riot of growth reflecting a variety of scale, colour and type and... a cacophany of noise... confronting... a traveler lost in the jungle.[3]

The image borrowed from James Dooge recalls Francis Bacon's dolorous depiction of science in his day. Klemes also seems to argue with Bacon that *our steps must be guided by a clue*, which to Klemes (unlike Bacon) means "a causal theory of hydrology".

Within groundwater modeling circles, the argument is rarely over whether to use causal models. Causal models are commonly applied in the field to explore and predict the behavior of groundwater systems, in order to guide or support both private and public water resource decisions. Discussions instead revolve around how best to represent the prototype with a model. As mentioned in Chapter 1, some hydrologists prefer to maintain a distinction between the conceptual model and the local parameters. This topic will also be taken up below in detail, but the term model is defined at the outset to include the entire body of geometry, processes, assumptions and parameters that contribute to model results. We will return to the surface water modelers' methodological debate in due course, but first we need to attend to the details of constructing a causal groundwater model. After some introductory discussion of modeling procedure, the present paper takes a general approach to the questions surrounding the justification or validation of model results. Chapter 3 provides a detailed case study of judgments made under uncertainty in support of larger public purposes, with particular attention to any problems in establishing criteria by which decisions and conclusions might be justified. Details of the link between internal and external standards of validation will be explored in later chapters.

Groundwater modeling methodologies have been described in detail by various authors, though with some disagreement over motivation and terminology.[4] Terminology is especially tangled in the areas of "calibration", "verification", and "validation". Later on we will explore whether the garbled lexicon is actually a sign of a more deep-seated conflict. However the steps are labeled, models are generally constructed in something like the following procedure, although it should be stressed that modeling of complicated systems is properly a reiterative process in which choices are frequently reconsidered; several aspects of the procedure may also be pursued simultaneously.[5]

1) problem definition - determination of performance criteria for the model, what the model is expected to accomplish, and within what constraints of time, funding, etc.;

2) a review of existing databases and reports on the actual or somehow similar systems to get an initial idea of important processes and the plausible range of values for various model parameters;

3) development of a qualitative and partial conceptual model (relevant geometry, material properties and processes) that serves to describe the system and guide any new data collection activity;

4) representation of the conceptual model in a mathematical construct, complete with governing equations and boundary conditions;

5) (sometimes) approximation of the mathematical terms in a form suitable for numerical approximation methods;

6) assignment or distribution of model parameter values, and (sometimes) pursuit of a field sampling or testing program;

7) initial "solution", either analytically or numerically;

8) calibration of model output to a portion of the available historical data by adjustment of model parameter values and/or model geometry;

9) second-phase history-matching - "verification" of model results by accurately predicting another portion of the historical data held in reserve for this purpose;

10) sensitivity analysis to determine critical model components and possibly to guide further data collection;

11) "validation" (hypotheses-testing);

12) use of the model for its intended purpose(s), such as a prediction of future system response to new and/or existing stresses, subject to new data and reformulation of any of the preceding steps.

13) performance of a post-audit some time later to determine the accuracy of model predictions, by comparison to actual field developments;

14) development of a better model, by inclusion of new information and reconsideration of any or all of these steps.

Hydrologic modelers routinely replace complicated hydrogeologic systems with what are surmised to be appropriate conceptual models and their equivalent mathematical statements.[6] The prototypal systems are invariably simplified through this procedure after suffering a loss of information, whether the loss is intentional or not. The important information is primarily in three areas: geometry, processes and parameters. The material of the original is mainly of interest as it affects any of these three things. The relative simplicity of the model often reflects the degree to which details of the prototype are unknown. Many nuances of the original can be lost in the translation from prototype to model. The geometry of the prototype is normally severely simpified; processes are often reduced in type, order and in the degree of interdependence; and the actual distribution of parameter values (heterogeneity) may be averaged in some fashion.

Simplicity can also be enforced by factors having nothing to do with limited hydrogeological insight. Computational convenience often militates against more inclusive models. Hydrologists working within an applied field-scale framework must often evaluate model choices partly in terms of a trade-off between tractability (which translates directly into economy of time and money) and accuracy. Modelers must constantly weigh what can safely be faked, lumped, averaged, minimized or ignored. Whatever the motivation, simplification obviously opens a certain distance between any model and the prototype; a degree of uncertainty may be introduced by the diluted analogy between the two.

Clearly, nevertheless, an "appropriate" conceptual model should preserve the contributions of essential features and forces of the physical system. The difficulty lies in identifying, separating and quantifying these essentials in constructing the model. What are these system essentials? The conceptual model consists of the system geometry, reasonable boundaries and associated flow and transport conditions, a plausible list of physical, chemical and biological processes contributing to the system response, and representative values or distributions of important system parameters. This essence of the system depends in part on the particular issues under study, which in turn are determined by the modeler's purpose. Even given these performance criteria, there is a certain latitude in the modeled description. Bear and Veruijt point out both the need for simplification and also its immediate consequence - the possibility of developing more than one model of the same system:

> The real system is very complicated and there is no need to elaborate on the need to simplify it... Because the model is a simplified version of the real system, *there exists no unique model* for a given groundwater system. Different sets of simplifying assumptions will result in different models, each approximating the investigated system in a different way.[7]

The same system can be modeled in arbitrarily many different ways; choices of model geometries, boundary conditions, processes and parameters can be combined in many ways, none of which is uniquely correct. Non-uniqueness may thus introduce additional uncertainty regarding any one model. The key is that every adequate model design must somehow preserve the essential behavior of the original, real system. Competing models will do so in different ways, with varying degrees of success in covering the range of data collected. Discussions of what constitutes an overall adequate fit is thus stock in trade for hydrologists, besides being fodder for legal or regulatory wrangling. At root, *the conceptual model is really an hypothesis about what is truly important in the physical system in light of the model's purpose and any accompanying external constraints.* As such, it is open to challenge and revision, even "falsification" and replacement.

Hydrologists are still trying to establish the minimum requirements for confident site description. It is not necessarily at all obvious which processes are truly important, for instance, nor in what ways they may interact with possibly poorly known material geometries and properties. Time and funding constraints typically apply before any theoretical considerations come into play. The corresponding weakness in the causal model is identified by Klemes in the debate among the surface water modelers:

> The attendant danger of this difficulty [shortfalls in basic research] is the temptation of reductionism: to make shortcuts and to fill the void between the data and the goals with logically plausible assumptions that are sometimes correct but often wrong and, more often than not, individually untestable.[8]

The ability to construct an appropriate model naturally depends in part on the local database, but as this is invariably incomplete, the strength of the analogy between model and prototype must depend to a degree on the experience and expertise of the modeler. Limited as this insight usually is, the typical applied modeling exercise is thus characterized less by a "*temptation* of reductionism" than by the *necessity* for the same.

Perhaps the biggest problem in narrowing the field of plausible models is a general inability to dependably describe the geologic systems, materials and responses of interest. At any scale significant to public activity, the composition and behavior of porous media and their contents are at least somewhat mysterious. The aquifer characteristics of interest usually include the hydraulic conductivity and the anisotropy ratio (measures of the ease with which water moves through the system in different directions), and the specific storage and specific yield (measures of the productivity of the system). Additional parameters must be specified when the problem in hand involves contaminant transport. Field data is typically sparse and sometimes inconsistent. Modelers will sometimes apply extrapolation and interpolation techniques to extend the limited field data and thereby synthetically generate spatially variable field parameters across the entire field of interest.[9] On other occasions, computational convenience (and possibly simplicity in the prototype) strongly favors an extrapolation methodology that is nothing more than choosing an "effective" value to be applied everywhere within a complex system. This single value may not actually represent any portion of that system; it is hoped that it captures the average behavior of the entire system. The scale and purpose of the model plays a large role in such decisions. Larger models (in time or space) usually lead to more sweeping assumptions to maintain tractability; this practice may be justified because effects tend to average out over either time or distance. Measured or synthesized, the simplified model requires estimates of various parameters. These estimates are almost always underdetermined by available data. Approximation therefore joins simplification as the major sources of model uncertainty. Research into geophysical methods attacks a portion of the parameter estimation problem directly. Indirectly, some progress has been made in dealing with the uncertainty in model parameter inputs through sensitivity analysis, uncertainty propagation, and other formal methods.

On the other hand, an epidemic of environmental problems occurring under a great variety of hydrogeologic conditions has spread thin the investigations of hydrologists. As a result of the diversity in settings and processes, detailed research in a particular location often has limited usefulness elsewhere, lessons learned are difficult to generalize, and progress in directly applied groundwater modeling is slow. More and more detailed study of a particular system, even if it leads to a better and better predictive model for that site, often simply ties that model more and more closely to some unique combination of prevailing conditions at that site. In this we can see the continued timeliness of Bacon's remark that

> In establishing axioms by this kind of induction, we must also examine and try whether the axiom so established be framed to the measure of those particulars only from which it is derived, or whether it be larger and wider.[10]

The construction of "larger and wider" models is complicated by the diversity of hydrogeologic systems, limits on allowable computational burdens, and the relative isolation of applied modelers. The insights of those working elsewhere are not easily adapted/imported to close the "void between the data and the goal".

Several factors thus combine to prolong the time required, the cost, the uncertainty and also the legal vulnerability of groundwater investigations. These include the complexity of the systems of interest, limited and inconsistent data, computational compromises, arbitrary choices among multiple modeling options, fragmented research, and the usual inability to generalize without significant loss of defensible relevance and predictive accuracy. The geometry of the original is often grossly simplified, though in ways the modeler judges will not distort the behavior "too much". In combination, these factors make a realist interpretation or application of a necessarily simplified model impossible; the model is clearly not a replica of the prototype, even with respect to essentials. They also complicate instrumentalist confidence in applied hydrologic models. C.F. Tsang admits:

> One quickly comes to the realization that how well modeling results can predict future behavior of real systems depends very much on the state of our knowledge of various physical and chemical processes that take place in complex geological systems which can be characterized only in a limited way.[11]

The net result is a sometimes large and often apparently irreducible uncertainty regarding idealized and approximated geometries, processes and parameters.[12] Both the logic and implications of a reductionist science will be examined in Chapter 4, when we focus on the logic of model testing.

Except in the simplest analytical models, all decisions regarding geometry, processes and parameters - however arrived at - are loaded into a computer code for a numerical solution of the mathematical formulation surmised to be appropriate. Standard codes typically incorporate versions of Darcy's law into equations stating the conservation of mass to generate governing equations for the spatial variability of hydraulic head. Additional assumptions may thus be implicitly introduced. These equations with accompanying initial and boundary conditions then constitute a mathematical model applicable to a specific problem.[13] In this process, Darcy's law is taken as an experimentally verified relationship, the empirical foundation and limited scope of which are considered in Chapter 5.

Directly applied hydrologic models clearly employ causal modeling tactics, in which the implications for system behavior are deduced (calculated) from certain initial conditions. It is equally evident that the conditions specified in the model will not exactly mimic those in the field; nor is the model solution likely to predict exactly the eventual behavior of the prototype. The remainder of the modeling procedure attempts to fine-tune and pass judgment on the soundness of the proposed model and its accompanying solution. The possibility of multiple competing models of the same system naturally raises the question of how one discriminates between non-unique and inexact representations. Of major interest in our developing taxonomy of hydrologic models is what implications the choice of a causal argumentative structure carries for this *validation*, regarding either its procedure or possibility.

2.1 Inspection: Validation of Groundwater Models

> Even in the midst of the strangest experiences we still do the same: we make up the major part of the experience and can scarcely be forced *not* to contemplate some event as its "inventors". All this means: basically and from time immemorial we are *accustomed to lying*. Or to put it more virtuously and hypocritically, in short, more pleasantly: one is much more of an artist than one knows.
>
> - Friedrich Nietzsche, *Beyond Good and Evil* [14]

It is an interesting fact that the precision or objectivity of the modeling art is contrary to the hierarchy of needs: the greatest sophistication is available for numerical solutions, while the least is at the conceptual modeling stage. The former are largely mathematical in nature, and benefit from both mathematical precision and the fact that shared problems are easily communicated and attacked in unison. Conceptual models, on the other hand, are essentially geological in nature, with troublesome admixtures of chemistry, biology, etc. As such they are necessarily descriptive, and suffer from ambiguities of observation and the isolation enforced by their site-specific nature. Conceptual descriptions are eventually reduced to a deceptively mathematical language within a numerical code. The relative precision of the numerical solutions comes into play only after iterative translations from prototype to conceptual model to governing equations. The overall result is thus dependent on the quality of the weakest and least objective of these links. Typically, this is the conceptual model itself.

Conceptualization uncertainty must be confronted in attempts to determine if the model is "working", a process often referred to as "validation". An effort is also usually made to improve and/or assess the model through the history-matching exercises of "calibration" and "verification". There is some sentiment to avoid all three of these terms, due in part to the fact that there never was complete agreement on what was meant by them.[15] Some also object to what they consider unwarranted connotations of model reliability.

"Calibration" is the least controversial of the three and has been retained here in its usual meaning of adjusting model components to achieve a better fit with data. The latter stages of model development commonly feature a comparison of known or assumed steady state conditions in the past to the model simulation of an equilibrium condition. When equilibrium data is lacking, transient simulations may be compared to transient records. Model geometry, fluxes or parameters may be iteratively adjusted within reason to fine-tune this fit; bounded automated parameter optimization may also be used.[16]

"Verification" usually consists of a single-blind test of a model's ability to "predict" historical data; this second stage of history-matching uses the calibrated model to predict known transient field data. When the required second set of data is available, the model is run forward in time from a known condition (preferably steady-state) with known stresses to see if model results reasonably match recorded heads and fluxes.[17] Once again, adjustments may be made in geometry, fluxes or parameters to improve the match between model results and historical records. Storage parameters are not involved in steady-state simulations; they can only be adjusted during these transient runs. This procedure is probably best described as what it is - namely, more history-matching - while avoiding the claims implicit in "verification".[18] Any model changes at this stage should trigger a second calibration cycle with the new values.

The major cause of either erroneous output or resistance to history-matching is often not obvious, however: models incorporate many interacting specifications that complicate a judgment on how flaws in output should be attributed. As a result, model "falsification" is not usually a wholesale affair, and the modeling procedure outlined and discussed earlier allows for frequent adjustment and reconsideration of various component bits. In due course, re-iterated history-matching usually results in a model judged to be acceptable in its fit to available steady-state and transient data. Anderson and Woessner advocate meeting "predetermined" calibration and verification "targets" for acceptable error.[19]

The term "validation" has been retained here to designate the demonstration of model success.

The semantic objections have been noted, but this choice is motivated by the term's ubiquity in the language of groundwater regulations.[20] The uncertainty associated with model testing and acceptance will be explored in a detailed case study in Chapter 3, and on logical grounds in Chapter 4. As will be discussed later, nested or coupled causes within the model not only make it difficult to pinpoint the source(s) of model error; they also make it unclear on what grounds any model can be shown to be "right".[21] In practice, model acceptance is generally the result of an iterative improvements in a conceptual model that lead to more satisfactory history-matching.

In practice, the ability of a model to reproduce historical data is often taken as a surrogate for a more comprehensive model test, or, as some would have it, verification implies validation. Thus de Marsily *et al.* say in some give-and-take with Konikow and Bredehoeft: "As long as [the parameters and structure of a model] reproduce the observed behavior of the system, we can use them to make predictions". They recognize no larger issues, and appear unconcerned that errors in a driving conceptualization might permit history-matching but still cause a model to be misleading:

> In no way has any groundwater flow model claimed that it can predict climate variability in the future! If, in the case of the Coachella Valley, the ten years following the model calibration are more humid than those used for the calibration, why should this in any way invalidate the model? ...It is clearly not the fault of the model if just one assumed climatic sequence was considered.[22]

Even though this *model* was not intended to predict climate variability, the *modeler* predicted that very thing when he assumed humidity would remain steady. It makes no difference if this simplification was intentional or an oversight. Regulatory agencies and private clients cannot be expected to accept de Marsily's "So what?" as an appropriate response to model errors or oversights that cause models to lose significant predictive accuracy. Many modelers thus view history-matching as a necessary but not sufficient validation test, and severe limits are commonly placed on how far into the future a history-matching model might be trusted.[23]

A committee of prominent modelers assembled by the National Research Council has pointed out the statistical shortcomings of simple history-matching:

> Traditional statistical methods are not particularly useful in ground water modeling studies, however, for several reasons. First, there are rarely enough measurements in ground water applications to provide a statistically rigorous test of a model's explanatory capabilities. These measurements are typically available at scattered well locations, which are spaced further apart than characteristic scales of variability. Second, the conditions prevailing when the measurements were collected may not reflect those that the model is designed to simulate. Finally, most classical statistical tests are based on assumptions that are not necessarily met in complex subsurface environments. These tests typically assume that the model's structure is perfect, and they are based solely on an analysis of the effects of measurement error. In reality, natural heterogeneity and deficiencies in model structure are likely to be far more important than measurement error.[24]

In the same exchange with Konikow and Bredehoeft over "validation", de Marsily points out that one of their models worked very well "taken from the shelf with no changes" after ten years and with "only the actual climatic forcing functions and the observed withdrawals introduced". In short, the model could successfully history-match another set of data after being re-calibrated. On this basis, de Marsily and his co-authors "claim that this is a 'validation' of the model", despite the problem that "of course, if we had looked back at the real 'predictions' of the model, made ten years ago, the results would have been very different".[25] They take the model to be distinct from the parameter values or other details it requires. Thus they do not view an incorrect estimate of future climatic conditions or pumping rates as a model error, but only as, respectively, evidence that further, different research needs to be done, or someone else's mistake.[26]

This approach and waiver of responsibility is consistent with the status of process-oriented models that lack or downplay certain application-specific details. The de-emphasis on parameter estimation and essential forcing functions comes at the cost, however, of being generally divorced from site-specific purposes. The limited goals of a process-oriented model are policed by correspondingly diffuse validation requirements. de Marsily has blurred the roles of process-oriented models and predictive ones: his model is at best a good process model, and at worst a poor predictive one. de Marsily *et al.* have inappropriately attempted to extend a weak validation protocol from one category to the other.

Sometimes inaccurately incorporated future stresses are the result of changing water management decisions, such as modified groundwater pumping schemes or contaminant loading rates.[27] Modelers typically operate with specific scenarios in mind, often having been supplied these by water managers. Stress specification errors can obviously be another complicating factor in the usefulness of groundwater models. Changes from original specifications can certainly cause model results to stray from actual outcomes, but there are no fundamental difficulties associated with these model "errors". There is no conceptual or theoretical barrier to projecting many different development schemes with their contemplated stresses; this is purely a logistical problem of sufficient computer run-time and storage. We explicitly remove all such manager-induced errors from our discussion of modeling protocol and problems. In this view, it is perfectly reasonable to insert in de Marsily's ten year old model the "observed withdrawals" prior to an assessment of model performance. This will not repair the original modeling efforts, but it certainly leads to a more meaningful post-audit that can judge, among other things, the accuracy and impact of the original climatic assumptions.

Modelers typically experience substantial difficulties even when perfectly prescient managers are on the job; in this specific case, these troubles are in the form of the "climatic forcing functions". de Marsily and his co-authors do not, however, give any information on what portion of the original prediction error is attributable to the climatic forcing functions that remain, in our view, their responsibility. Here we are interested in a broader and *a priori* view of validation as some assurance that model projections reasonably indicate the eventual response of the original physical system to contemplated stresses, thereby ensuring that initial performance criteria can be met and that the model can produce useful information. In short, the scope of validation must match the scope of the model's application. In this broader view of applied models and validation, it is not enough to say that if we knew all of the relevant field parameters and conditions, our process-oriented model would give useful information. *Predictive tools* must argue or assume that parameters and conditions are either constant or distributed, steady or predictable; moreover, specific values must be assigned within the model (via any of many methods). In this view, projections of climatic conditions are the modeler's responsibility, as are the direct consequences, such as recharge and evapotranspiration rates.

Validation is sometimes used in a very limited sense to mean only that a computer *code* operates as designed, *ie.*, that a numerical approximation algorithm is accurate within specifications. While a validated code that can also reproduce the known history is a standard beginning, validation of the model as a whole must assess the appropriateness of the bundled hypotheses (approximations, assumptions and guesses regarding the system essentials) embedded in the conceptual model. Even though some aspects of a proposed predictive model may not rest on well understood processes or confidently measured field information, validation must still offer an estimate of the overall performance of the model in pursuit of the stated objectives. This is not too unlike the sense in which regulatory agencies often use the term. The United States Department of Energy favors a definition of validation as "a process whose objective is to ascertain that the code or model indeed reflects the behavior of the real world", while to the U.S. Nuclear Regulatory Commission it is a way of gaining "assurance that a model, as embodied in a computer code, is a correct representation of the process or system for which it is intended". The International Atomic Energy Agency maintains that "... a model cannot be considered validated until sufficient testing has been performed to ensure an acceptable level of predictive accuracy".[28] By any of these criteria, de Marsily's model clearly fails; even granting a dispensation for the future withdrawal rates, predictive accuracy requires an ability to predict future precipitation and recharge.

These standards, motivated in the main by the problem of siting nuclear waste repositories, are obviously the subject of some dispute in the current validation debate. Certainly there is cause for concern if they are taken to smear the difference between literally capturing the full details of the real behavior (a realist goal), versus simply producing a good (possibly lumped parameter) model that functions well as a predictor (an instrumentalist's goal). On the other hand, models are often used to make or support decisions involving major expenditures and substantial environmental risks; on the face of it, therefore, it does not seem unreasonable to require that models should "reflect the behavior of the real world", with "an acceptable level of predictive accuracy", at least with regard to the stated modeling objectives. Otherwise, what good are they?

To gain perspective on the questions and objections prompted by regulatory definitions of validation, we must consider in greater detail the basis on which the model rests. These remarks can then serve to frame the detailed consideration of model testing issues illustrated in the case study of Chapter

3. Bredehoeft and Konikow describe the construction of an appropriate conceptual model as "an *a priori* decision on the part of the modeler", a description that emphasizes the discretion afforded the modeler. They also say:

> The new computer methods, both for parameter estimation and for establishing confidence bounds for predictions, suggest to the uninitiated that the modeling process can be done with a minimum of human judgment. There are a number of important steps in modeling where professional judgments are critical.[29]

Since the critical steps are somewhat idiosyncratic, the final verdict can be no less so. Maloszewski and Zuber note that, "Contrary to calibration, the validation process is a qualitative one based on the modeller's judgment".[30] Even the International Atomic Energy Agency says that "the acceptable level of accuracy is judgmental and will vary".

Similar thoughts have prompted many prominent modelers to adopt a reflexive or self-referential standard for validation that cannot be easily explored by public interest or legal proceedings. C.F. Tsang identifies expertise as the critical input (italics added):

> Some of the modeling steps, such as the design of conceptual models... will be difficult to validate in an objective way, since their validation depends on the depth and breadth of scientific knowledge of the modelers involved... *A thorough understanding represents actually the major part of validation.*[31]

In this view, models can at best be provisionally validated in some limited sense by a concurring consensus of experts. The situation is again complicated by the inherently site-specific nature of applied models, making validation a case-by-case enterprise in which few will match the insight of the principal investigators. There is, moreover, a shortage of qualified modelers with expertise not only in numerical methods but also in hydrogeologic field science. A systematic peer review plan is therefore generally infeasible, although selected, high-profile sites might benefit from special review.

Some writers have challenged the whole idea of validation, arguing on philosophical and semantic grounds that a claim of validation can never be sustained.[32] A careful consideration of this view will be taken up below (Chapter 4), but Tsang summarizes the gist of various authors' comments succinctly: "Based on the above discussion it may be apparent that almost by definition, one can never have a validated computer model without further qualifying phrases."[33] Naturally, this conclusion depends on just how validation is defined. Konikow and Bredehoeft would banish the terms verification and validation from the discussion altogether; they feel these terms give "the uninitiated" a very misleading idea of the degree to which supposedly verified or validated models capture reality.[34]

Many applied (external) models relate directly to public concerns through procedures requiring public participation. Abandoning or diluting the idea of routine validation of models for any purpose has profound implications. For one thing, it may mean that only a comparison of model results and actual system response can determine or improve the goodness of fit from one to the other, as in de Marsily's exercise. This is essentially the purpose of what are known as post-audits. If we consider models as proposed explanations - or at least descriptions - of important structures and driving processes at a site, then post-audits and other revisitings might seem to offer confirmation with greater confidence of one explanation or another. A forthright discussion of modeling uncertainties by various practitioners has been joined in recent years by the occasional audit of modeling performance. Both Superfund sites and many projects requiring Environmental Impact Statements under the National Environmental Policy Act (NEPA) have been revisited.[35]

Given the heavy reliance on the experience-based insight of the modeler, observers of an immature science may well wonder to what extent the strengths and weaknesses of an individual modeler influence either his conceptual grasp of a situation, his site methodology or his interpretation of results. Audits indicate that the results of predictive modeling are mixed at best, and frequently seriously flawed. Viewing models as a summary of a generally inadequate grasp of the subsurface, such a finding is not unexpected. The relative rareness of post-audits focused on the hydrology of predictive models, however, has led some observers to chafe modelers as the practitioners of a new risk-free "indoor sport".[36] The next chapter covers this game wire to wire. Post-audits will considered further in a discussion of the logic of model testing in Chapter 4.

We began by surveying the rationales advanced by surface water modelers for the use of causal models within their discipline. At one point Vit Klemes forcefully promoted such models. But just a few years later, he is grumbling about the new developments in terms whose spirit is a general summary of the situation in directly applied groundwater modeling, as well:

> The so-called hydrological conceptual models have been conceived in the spirit of good science, trying to improve on the black box models by introducing hydrological mechanisms and processes into the modelling of rainfall-runoff relations. The problem was that most of these mechanisms were not well understood; nor were their interactions, either in general or in specific conditions of different river basins. Thus the resulting models are concoctions of a few facts and many artificial constructions, in which the individual processes are represented by postulated rather than by established mechanisms and interactions and by assumptions arrived at by simplistic reasoning rather than serious hydrologic research. In effect they are, for the most part, just complex assemblages of black or, at best, dark grey boxes.[37]

2.2 Conclusion

Over the past ten years the practice of hydrogeology, and in particular its use of applied groundwater models, has been complicated by the pressure exerted by issues of public policy and environmental law. Major concerns include the siting, characterization and remediation of facilities that can or did cause groundwater contamination. As might be expected, the effects of any technical uncertainty are exacerbated within the legal processes that drive so much of the activity. Applications of theoretical developments within hydrology to full-scale systems are thus accompanied by special complications. Our motivations for exploring methodology in hydrology recognized these extra-scientific complications. Even as hydrologists feel their way within the limits of their science, their attempts at prophecy often become central elements in arguments between partisan observers who have their own, possibly conflicting, standards of performance and proof. Our stated goal was a methodological description of how in general and in the particular case how groundwater scientists know what they claim to know.

It should already be apparent from this preliminary discussion of applied modeling that selection among, and interpretation of, available facts are the crucial activities; together, these constitute the essential something that must fill out the facts to give them shape, substance and weighted meaning.[38] The historian Carr says further:[39]

> Ignorance is the first requisite of the historian, ignorance which simplifies and clarifies, which selects and omits... The modern historian must cultivate this necessary ignorance for himself... He has the dual task of discovering the few significant facts... and of discarding the many insignificant facts.[40]

Predictive groundwater modeling has been described as artful insight reliant on both process-oriented science and on field technologies; the model is sometimes said to merely incorporate, expand and illustrate the educated opinion of the modeler.[41] To the extent that this is true, the real bases on which models depend will remain largely opaque to the "uninitiated" public. The doubtful view of validation suggests that even within the hydrologic community models cannot be trusted until after their predictions are borne out, in which case the predictive stature of models is negligible. If the comments reported so far are representative of the real situation, major challenges face hydrologists attempting the usual *Experiments of Fruit*.

We began by noting the misgivings of many hydrologists over the status of their science. In the context of applied modeling and regulatory expectations, these misgivings suggest that hydrology may not be a reliably predictive science. This unpleasant tentative conclusion may stem mostly from present-day procedural, practical or technological considerations; these problems are quite possibly temporary in nature. Following a look at these problems in Chapter 3, a more detailed examination of validation issues will be taken up in Chapter 4, despite the warning from Konikow and Bredehoeft that "much professional effort is being devoted to validation; it is costing more than a semantic ambiguity is worth."[42] This later re-visiting will emphasize more resistant difficulties stemming from the logic of model testing. As will become apparent, hydrologists are not quibbling about terminology; the semantic differences seemingly endemic to the discussion of hydrologic validation result from a basic murkiness about the goals of the science.

Thus far, there is some indication that fundamental difficulties may accompany the direct application of hydrologic theory. In particular, the prospect of limited data substantiating multiple, mutually inconsistent models appears to generate misgivings about the validation process. This is best observed by means of a well-documented case study. The preference of water managers for a single reliable model is understandable enough; a story of their frustration in the presence of more than one "equally good" model is told in the next chapter. The history of two models from the United States Geological Survey will further illuminate our scrutiny of the applied modeler's craft; in particular, it will give substance to both our outline of modeling procedure and the nature of validation misgivings.

2.3 Notes:

1. An early version of this chapter appeared as Hofmann, P. (1994), "The Case of the Immodest Modeler: Uncertainty and Realism in Groundwater Modeling", *New Mexico Conference on the Environment*, New Mexico Environment Dept., Albuquerque, NM, April, 1994. The title phrase is taken from Klemes, V. (1988), "A Hydrological Perspective", in *Journal of Hydrology*, **100**, pp.3-28:

   > The so-called hydrological conceptual models have been conceived in the spirit of good science, trying to improve on the black box models by introducing hydrological mechanisms and processes into the modelling of rainfall-runoff relations. The problem was that most of these mechanisms were not well understood; nor were their interactions, either in general or in specific conditions of different river basins. Thus the resulting models are concoctions of a few facts and many artificial constructions, in which the individual processes are represented by postulated rather than by established mechanisms and interactions and by assumptions arrived at by simplistic reasoning rather than serious hydrologic research. In effect they are, for the most part, just complex assemblages of black or, at best, dark grey boxes. (p.12)

2. Carr, E.H. (1961), *What is History?*, Vintage Books, p.9.

3. Klemes, V. (1982), p.103.

4. Anderson, Mary P. and Woessner, W.M. (1992a), *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*, Academic Press, pp.6-9; Anderson, M.P. and Woessner, W.M. (1992b), "The Role of the Postaudit in Model Validation", in *Advances in Water Resources*, **15**, pp.167-173, provides a flow chart of model development; Maloszewski, P. and Zuber, A. (1992), "On the Calibration and Validation of Mathematical Models for the Interpretation of Tracer Experiments in Groundwater", in *Advances in Water Resources*, **15**:1, pp.47-62. Mercer, J.W. and Faust, C.R. (1981), *Groundwater Modeling*, National Water Well Assn.; Bear, J., *et al.* (1992), "Fundamentals of Ground-water Modeling", in *EPA Ground Water Issue*, EPA/540/S-92/005, April 1992.

5. Carr, E.H. (1961), pp.34-35: "As any working historian knows, if he stops to reflect on what he is doing as he thinks and writes, the historian is engaged in a continuous process of moulding his facts to his interpretation and his interpretation to his facts. It is impossible to assign primacy to one over the other."

6. This sentence is a kind of test for attitudes toward modeling. Choose the most appropriate verb (arranged from the most confident to the nearly criminal) to complete the sentence:

   "Complicated hydrogeologic systems are routinely simplified by replacing them with what are
   known
   expected
   thought
   believed
   professed
   reckoned
   asserted
   advanced

presented
introduced
proffered
suggested
supposed
hypothesized
conjectured
surmised
counted on
implied
hinted
purported
fancied
imagined
hoped
guessed
alleged
insinuated

to be appropriate conceptual models and their equivalent mathematical statements..."

7. Bear, J. and Veruijt (1987), *Modeling Groundwater Flow and Pollution*, D.Reidel Publishing Co., p.12.

8. Klemes, V. (1982), "Empirical and Causal Models in Hydrology", in National Academy Press (1982), *Studies in Geophysics: Scientific Basis of Water-Resource Management*, p.102.

9. See the discussion in National Academy of Sciences (1990a), p.84.

10. Bacon, F. (1620b), *cv*, p.71.

11. Tsang, C.F. (1991), "The Modeling Process and Model Validation", in *Ground Water*, **29**:6, pp.825-31.

12. See, *eg.*, Hassanizadeh, S.M. and Carrera, J. (1992), "Editorial", in *Advances in Water Resources*, **15**:1, pp.1-3. The authors introduce this issue of *AWR* devoted to validation issues and perspectives by taking note of "...complications associated with spatial and temporal effects. Geo-hydrological media are mostly heterogeneous and virtually impossible to characterize sufficiently; processes that are not relevant at small scales become dominant at larger scales; new processes emerge as a result of a change of scale" (p.1). Hydrology does share some problems with the other historical sciences, such as paleontology and evolutionary biology. These have to do with the trouble inherent in the study of incomplete records of unique physical systems that resist generalization. Consider Gould, J. (1989), *Wonderful Life: The Burges Shale and the Nature of History*, Norton, p.59: "Puzzles mount upon puzzles the more we consider details...". Operational differences in the various sciences are often rooted in the unevenness of the social significance of the conclusions reached; later, in discussing the prospects for explaining the history of life on earth, Gould points out candidly: "This goal, once achieved, brings no particular earthly benefit" (p.84).

13. See, *eg.*, Wang, H.F. and Anderson, M.P. (1982): *Introduction to Groundwater Modeling: Finite Difference and Finite Element Methods*, W.H. Freeman and Co. Typical governing equations are noted in Chapters 3,4, pp.41-88. See also Anderson, M.P. and Woessner, W.W. (1992a), *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*, Academic Press; general

equations from both the "aquifer viewpoint" and the "flow system viewpoint" are given on pp.12-20; the governing equations for special conditions of unsaturated or multi-phase flow, solute transport, fractured media and density-dependent flow of miscible fluids are addressed on pp.321-335.

14. Nietzsche, F. (1886): *Beyond Good and Evil: Prelude to a Philosophy of the Future*, Vintage Books Edition (1966), p.105.

15. See, *eg.*, National Academy Press (1990a): "Model validation is a term that means different things to different people, largely because it is rarely defined with any precision. This general concept has both technical and policy origins" (p.230); and: "Several different groups in ground water modeling have in the past defined and used the terms validate and verify to mean different things. In fact, there is no consensus among ground water hydrologists, either on the definition of these two terms or on how to achieve validation and verification" (p.237).

16. Wang, H.F. and Anderson, M.P. (1982), *Introduction to Groundwater Modeling: Finite Difference and Finite Element Methods*, W.H. Freeman and Co., p.109: "It is not unreasonable to adjust the input data because these data are imperfectly known, and there will be a certain range of values that may be valid. It is not uncommon to make from twenty to fifty trial-and-error simulations before an acceptable calibration is achieved. Calibration is really a way of solving the inverse problem..." Hence the interest over the last ten years in automated parameter optimization techniques. It remains to be seen if such sophisticated techniques are only another example of Thoreau's aphorism: "All our inventions are but improved means to an unimproved end". Anderson and Woessner (1992a, pp.223-246), discuss calibration at length. They introduce the subject by noting that: "Calibration is accomplished by finding a set of parameters, boundary conditions, and stresses that produce simulated heads and fluxes that match field-measured values within a preestablished range of error" (p.223). They discuss the occasional need for a "transient calibration" (p.226). They later indicate that "Parameter estimation is essentially synonymous with model calibration" (p.231), suggesting that geometry is relatively fixed at this stage. Naturally, if calibration is impossible without resorting to fantastic parameter values, more fundamental choices must be reconsidered. Their discussion covers trial-and-error and automated methods, calibration representation guidelines, methods of evaluating calibration results (mean error, absolute mean error, and root mean squared error), and the effects of these choices on model adjustments.

17. Anderson, M.P. and Woessner, W.W. (1992a), pp.255-256.

18. See, *eg.*, Oreskes, N. *et al.* (1994): "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences", in *Science*, **263**, 4 Feb 1994, pp.641-646, in which the over-promising implicit in these terms is debunked at length.

19. Anderson, M.P. and Woessner, W.W. (1992a), pp.226-7, 255, 283: "A calibration target is defined as a calibration value and its associated error", *eg.*, "10.12m ± 0.23m". A verification target is defined similarly for the transient simulation. Both calibration and verification errors will not be uniform everywhere within the model, and different standards may be applied to different areas depending on the objectives of the modeler.

20. Having freed up the term verification from lesser roles (using the term history-matching instead), later on we will occasionally use verification as synonymous with validation, mainly because *verifiable* and *verifiability* are words; there are no equivalent forms of *valid*.

21. Anderson, M.P. and Woessner, W.W. (1992a), pp.226-229, 231-236, discusses the different methods of calibration: trial-and-error and automated. However, "Even in a quantitative evaluation, however, the judgment of when the fit between model and reality is good enough is a subjective

one. To date, there is no standard protocol for evaluating the calibration process, although the need for a standard methodology is recognized as an important part of quality assurance in code application" (p.236, quoting the National Research Council, 1990a). Since calibration can be problematical, surely validation can be expected to be even more difficult. This issue will be taken up at length in Chapter 4, below.

22. de Marsily, G., Combes, P. and Goblet, P. (1992), "Comment on 'Groundwater Models cannot be Validated', by L.F. Konikow and J.D. Bredehoeft", in *Advances in Water Resources,* **15**, pp.367-369.

23. See, *eg.*, Bredehoeft, J.D. and Konikow, L.F. (1993a), "Ground-water Models: Validate or Invalidate?", in *Ground Water*, **31**:2, pp.178-9, where the authors suggest a rule of thumb that models should not be projected further into the future than they can successfully history-match into the past (p.178).

24. National Academy Press (1990a), p.231.

25. de Marsily, G. *et al.* (1992), pp.368-369.

26. This approach to models is not uncommon. A distinction is often made between, on the one hand, a numerical approximation of the important processes, and on the other hand, the inclusion of necessary field information in a predictive model. See Peters-Lidard, C.D. and Wood, E.F. (1994), "Estimating Storm Areal Average Rainfall Intensity in Field Experiments", in *Water Resources Research*, **30**, 7, pp.2119-2131: "... we present a 'theoretical' error variance model which could provide a priori estimates of the areal mean precipitation errors provided its parameters can be estimated" (p.2120).

27. This topic is briefly discussed in Anderson, M.P. and Woessner, W.W. (1992a), pp.257-259.

28. *Radioactive Waste Management Glossary* (1988), 2nd edition, International Atomic Energy Agency Technical Document, IAEA-TECDOC-447; *DOE Environmental Assessment - Yucca Mountain Site, Nevada Research and Development Area* (1986), DOE/RW-0073, Vol.2, U.S.DOE, Office of Civilian Radioactive Waste Mgt; *NRC A Revised Modelling Strategy Document for High Level Waste Performance Assessment* (1984), U.S. NRC, Office of Nuclear Regulatory Research.

29. Bredehoeft, J.D. and Konikow, L.F. (1993a), pp.178-9.

30. Maloszewski, P. and Zuber, A. (1992), p.48.

31. Tsang, C.F. (1991), pp.829-30, italics added. See also National Academy of Sciences (1990a), pp.82-3.

32. See, *eg.*, Oreskes, N. *et al.* (1994): "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences", in *Science*, **263**, 4 Feb 1994, pp.641-646.

33. Tsang, C.F. (1991), p.826.

34. Bredehoeft, J.D. and Konikow, L.F. (1993a), p.179.

35. There have been only a handful of specifically hydrologic case studies performed. See, *eg.*, Konikow, L.F. and Bredehoeft, J.D. (1992a), "Groundwater Models Cannot be Validated", in *Advances in Water Resources*, **15**:1, pp.75-83; Anderson, M.P. and Woessner, W.W. (1992a), pp.286-294; and Anderson, M.P. and Woessner, W.W. (1992b), pp.167-173. More generally, see

Culhane, P.J. *et al.* (1987), *Forecasts and Environmental Decisionmaking: the Content and Predictive Accuracy of Environmental Impact Statements*, Westview Press.

36. Beven, K. (1993), "Prophecy, Reality and Uncertainty in Distributed Hydrological Modelling", in *Advances in Water Resources*, **16**:1, pp.41-51, quotes L. von Bertalanffy (1966) as saying: "Nowadays model building has become a fashionable and generously supported indoor sport" (p.41).

37. Klemes, V. (1988), "A Hydrological Perspective", in *Journal of Hydrology*, **100**, pp.3-28 (p.12).

38. Carr, E.H. (1961), p.22, elaborates: "...the main work of the historian is not to record, but to evaluate; for, if he does not evaluate, how can he know what is worth recording?" Compare Bacon's dismissal of uncritical observation and induction, p.4-5, above.

39. Yeats said: "The best lack all conviction, while the worst are full of passionate intensity".

40. Carr, E.H. (1961), pp.13-4. The repeated references to this well-known collection of essays are intended to reinforce the parallels between modeling in the physical sciences and what are more easily accepted as preliminary and speculative efforts in the humanities and social sciences.

41. In a famous environmental trial related to the contamination history of Woburn, MA, the well-known hydrologist George Pinder once declared that there was no way his computer model or its results could change his basic opinion of the site. He frequently reiterated that neither his conceptual model nor his final opinion depended on computer output. And thus his "opinion would not be affected whether my drawings [contaminant concentration contour plots, etc.] were done by a computer or by a Disney artist. (see Pinder, G.: United States District Court, District of Massachusetts, Civil Action 82-1672-S, Anne Anderson et al. v. W.R.Grace and Co. et al.: deposition, day 3, 10 Jan 1986, p.140-143.)

    Pinder used his models to explore the system as part of an "intellectual exercise", and then later in an illustrative way at trial. It is a well-known and regrettable phenomenon that computer-generated plots and other output are more compelling to the layman - and even to too many professionals - than the unvarnished opinions they embody and illustrate:

    > Modelers must contend with the practical reality that model results, more than other expressions of professional judgement, have the capacity to appear more certain, more precise, and more authoritative than they really are [National Academy of Sciences (1990a), p.11].

    It is sometimes remarked that a certain danger lies in the ease with which positions essentially unsupported by the evidence can be promoted through the use of computer models and graphics. Some have even suggested that the development of user-friendly modeling codes (still in its infancy) does the discipline a disservice, since such codes lower the barriers to entry by the uninitiated. This initiation refers to the likelihood that Tsang's "thorough understanding" informs the modeling effort.

    > Indeed, it could be argued that the lack of a user-friendly interface may be a useful safety feature to help prevent inappropriate use of the models by nonqualified personnel [National Academy of Sciences (1990a), pp.18-19].

42. Konikow, L.F. and Bredehoeft, J.D. (1992a), p.82.

3

Apples and Oranges:

Modeling the Tesuque Aquifer Near Pojoaque, New Mexico

If gold ruste, what shal iren do?
- Chaucer, *Canterbury Tales* [1]

It is difficult to aim anything but imprecations accurately by moonlight.
- Edgar Rice Burroughs, *A Princess of Mars* [2]

3.0 Introduction

The general misgivings about groundwater modeling have not, of course, prevented models from occupying a prominent place in water resource planning and remediation. Jacob Bear has pointed out the various types and sources of uncertainty; in 1992 he is forced, nevertheless, to acknowledge the practical situation:

> Often the question is raised as to whether, in view of all these uncertainties, which always exist in any real-world problem, models should still be regarded as reliable tools for providing predictions of real-world behavior - *there is no alternative!* [3]

Earlier, Jay Lehr was not completely sure. Despite his self-described status as modeling convert and expert, he writes in 1980: "Today, computers simulate everything- except our ability to to understand what in hell they're all about. Few people really understand enough to criticize, so the proverbial emperor's nakedness continues to go unannounced". And hence: "To model or not to model - that is the question".[4] Mary Anderson followed in 1983, saying: "Let me be one of the people in the crowd to shout, 'Be careful! The Emperor has no clothes!'."[5]

It has already been suggested that hydrologists might expect to face difficulties if forced to choose a single guiding conceptual model with legal or economic ramifications. According to Bear (p.33), single, unevolving model approaches reject potential insights. The possibility exists therefore for conflict with the "convergent legalism" of managers, in which a single and possibly static model often dominates the policy-making discussion. As Anderson points out, "it is easy to lose sight of the limitations of a model when results are brought into the emotionally charged arena in which many regulatory decisions are made".[6] The technical strategies and difficulties of modelers may be given scant attention by managers when a single model is by default "regarded as a reliable tool for providing predictions of real-world behavior".

Our discussion of applied models will benefit, therefore, from a closer look at a specific case in which the existence of more than one model severely upset the operational realism of managers and greatly complicated the decision-making procedure. When available, different models - "each approximating the investigated system in a different way" - may not simply fail to converge to a single, optimal model, a possibility considered carefully in Chapter 5. Sparse data can sometimes support conceptualizations so fundamentally different as to be incommensurate, *i.e.*, they cannot even be directly compared due to fundamentally alien approaches. For example, the point of contention between two models may be much more than the best approximation of some parameter (such as the hydraulic conductivity) to be inserted into both models; one model might be fundamentally more sensitive to this choice than the other, due to structural differences. The typical managerial hunt for a single preferred model then impels a highly uncertain choice among very different modeling approaches, each with its own insights and blind spots.

What Vit Klemes calls the "void between the data and the goal" (p.34) - and the regulatory response - figures prominently in the present case study of essentially incommensurate models. Events in this case even challenge the general optimism of hydrologists in the area of groundwater flow models that avoid the admitted difficulties of contaminant transport - an optimism reflected in Anderson's 1983 comments on modeling:

> In a ground-water flow model, we are fortunate in that the theoretical framework upon which the model is built has been well verified and the meanings of the parameters are well understood. In this case the model (Emperor) has a good set of underwear and the directions for tailoring his clothes are clear.[7]

Our understanding of hydrologic methodology and the misgivings associated with its uncertainty can be greatly advanced by an examination of the details of model construction and validation for practical purposes. This case has forced managers to confront the strategies and limitations of modeling directly.

3.1 Background

In the early 1970s, the U.S. Bureau of Indian Affairs (BIA) proposed an expansion of the existing irrigation system in the Pojoaque River basin, located in north central New Mexico just north of Santa Fe (see Figure 3.1). The plan calls for new withdrawals of about 45 cfs of groundwater from 708 wells installed near the Pojoaque River and its tributaries, including the Rio Nambe, Rio en Medio, Rio Chupadero and the Rio Tesuque, as well as diversions of about 13 cfs of surface water from the streams. Estimates of return flow through infiltration, if accurate, would reduce the net new withdrawals to about 28 cfs; total groundwater pumping (existing and additional) would then be just under 40 cfs. It is proposed to irrigate 11,337 acres of tribal land with diverted surface flows augmented with groundwater during dry seasons; another 2,628 acres of non-tribal land in the basin are to share in the new supply. The acreage brought under cultivation does not completely mirror the distribution of the new supply; 88% is for irrigation of tribal land, 11% for irrigation of non-tribal land, and 1% for expanded municipal supplies (tribal acreage is 82% of the total). Well installation and diversion permits have not been approved by the New Mexico State Engineer Office as of March, 1995.

A consistent regulatory concern of the State Engineer is that within an appropriated basin, applicants must demonstrate that proposed new water uses will not adversely affect users with prior rights within the basin. Since the Pojoaque River basin is fully appropriated, the expansion of irrigation must not impact existing water rights. The hydraulic connection, if any, between the targeted productive pumping zones and the surface water bodies is thus the principal technical question in the Pojoaque River basin. The water rights situation in this basin is ultimately complicated by larger questions. Rights of prior appropriation under the BIA plan are part of a much larger legal war over the status of Indian water rights throughout the American West. The tangled legal background of how the battle goes for the Pueblos of New Mexico in particular will be surveyed briefly in the concluding sections of this paper. These non-scientific concerns add emphasis to our rationale for examining hydrologic methodology; methodological debates can dominate the margins of hydrologic practice where water resource decisions are made.[8] For the moment it will suffice that it is important for various

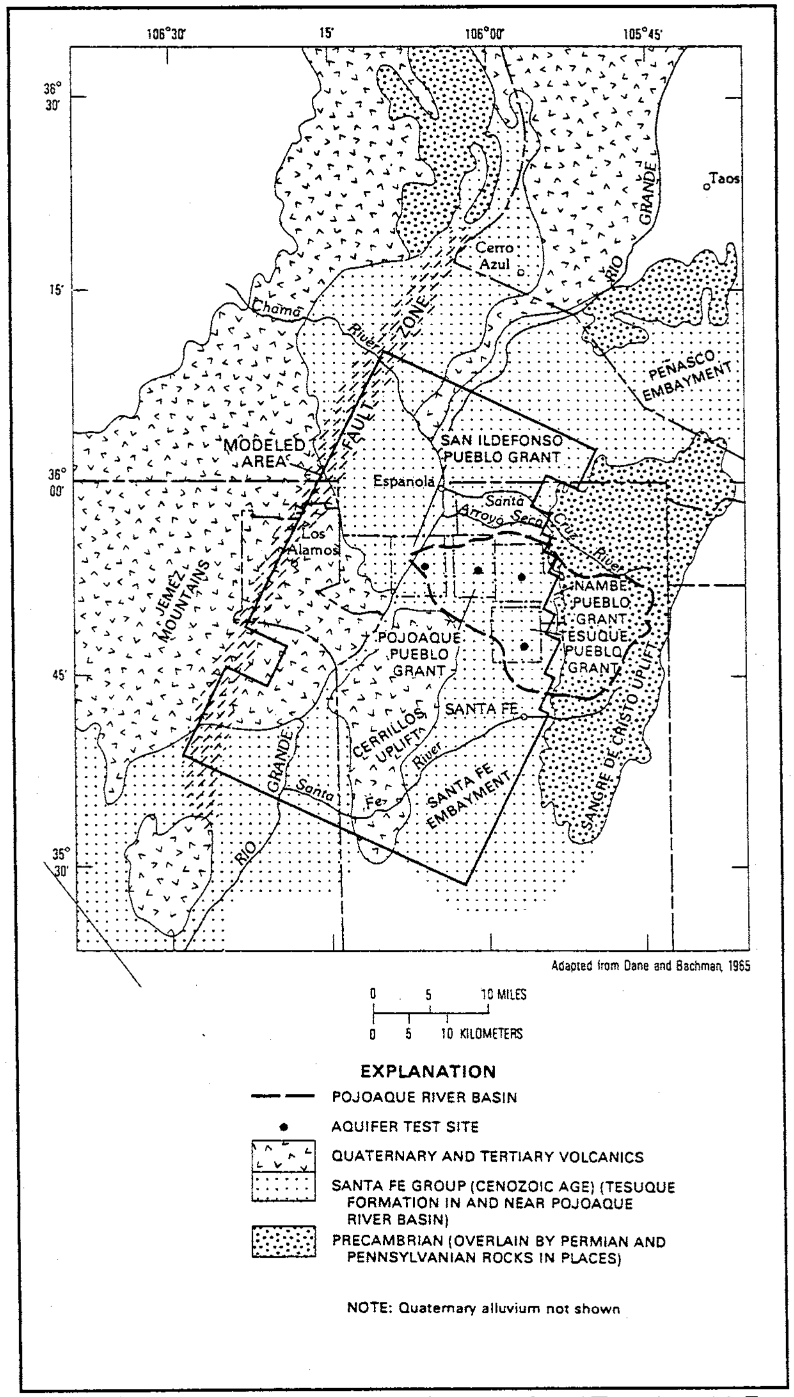


Figure 3.1: Location of the study area, showing the outline of Hearne's model. From Hearne (1985a), p.2.

reasons to assess the potential impact of the BIA plan on surface and groundwater flow within the Pojoaque River basin.

Some of the interested parties are actually outside the Pojoaque River basin, as is apparent from a list of the major existing wells within the larger Espanola Basin. Wellfields developed in Los Alamos Canyon (1947), Guaje Canyon (1950), and on the Pajarito Mesa (1965) all supply the city of Los Alamos, while the Buckman (1972) and Santa Fe (1950s) Well Fields supply a portion of the municipal water of Santa Fe. These wells accounted for nearly all of the withdrawals at the time of the proposal (11.24 cfs in 1977). The situation in 1976 and details of the proposed changes are gathered together in Table 3.2, page 73 below. Table 3.2 also indicates the return flow from infiltration of irrigation (and hence the net withdrawals) expected by the developers of the proposed plan.

3.2 Hearne Model: Complicated Structure and Simple Parameters

An investigation was undertaken by the United States Geological Survey (USGS) in an attempt to assess the likely effect of the proposed development on both groundwater levels and streamflow within the basin. An Open File Report by Glenn A. Hearne appeared in January, 1981, later bound as a Water Supply Paper in 1985.[9] The motivation for the study is given as: "The U.S. Geological Survey was requested to evaluate the effect of ground-water withdrawals on ground-water levels and streamflow".[10] As discussed above, we will defer any further discussion of the legal context and examine how Hearne pursues his stated purpose.

Introductory comments in the Water-Supply Paper concern the objectives and possibilities of modeling, and are couched in language we have since accepted for our own. Hearne is interested in the "analogy between the prototype and the model" in three areas: structure (geometry), aquifer characteristics, and boundaries. The only processes of interest are physical, since this is not a water quality issue. His view of this analogy is reminiscent of our earlier discussion:

> Although a digital model can assist in analysing a system and can make predictions for use in management decisions, the user must realize that the model is only an approximate representation of

> the prototype system. The validity of the predictions made by the model depend on the closeness of this representation. The state of the art of digital modeling does not permit a statement on the confidence limits bounding the projections made by the model. This still needs to be done subjectively.[11]

Hearne's report is exemplary in its detailed documentation of the choices made in the modeling process. This can be attributed in part to the report being geared to meet the challenge of expected legal scrutiny.[12] Pertinent older reports and current investigations - both geological and hydrological - are carefully considered to first construct what appears to be a reasonable conceptualization of the site, and later to establish boundary conditions and to estimate the needed field parameters. The rationale for all modeling decisions is readily available to the interested reader in a useful form.

Hearne begins with a compact description of what he calls the "structure of the prototype", that is, of the general geology of the Pojoaque River basin. The Santa Fe Group in the Espanola Basin consists of the Puye, Ancha and Tesuque formations (in order of increasing age). A generally westward dip causes these to outcrop successively from west to east. The basin has been studied for various purposes over the decades,[13] but in the course of Hearne's description, certain problems are noted. The thickness of the Tesuque formation and the nature of the underlying rocks remain unknown,[14] as does the spatial variability of both the geologic architecture and the associated hydraulic properties. The Puye on the west consists of latite and andesite gravel, sand and tuff;[15] it is said by Manley to be derived from the Jemez mountains and to contain numerous laharic and pyroclastic deposits.[16] The Ancha is also a fanglomerate, contemporaneous with the Puye (2.9 my and 2.7 my, respectively), but derived from the Precambrian granitic material of the Sangre de Cristos on the east.[17]

The underlying Tesuque formation is said by various authors to be at least 3,700 feet deep, with some putting the thickness much greater, especially near the Rio Grande.[18] Hearne summarizes the geology as "interbedded layers of gravel, sand, silt, and clay with some intercalated volcanic ash beds". Purtyman and Johansen describe the intertonguing of the Tesuque with the Puye and other formations, but Hearne decides to include these formations with the Tesuque aquifer system. He reasons that "although these formations may be quite important locally, their potential for affecting the geohydrology beneath the Pojoaque River basin is slight".[19]

On the other hand, Hearne identifies two of what he regards as essentials of the modeled system: "the dip of the beds and the lack of continuity of the individual beds".[20] Dip is generally to the west and northwest, and has been reported as anywhere from near zero to as much as 30° for individual fault blocks. The average *bed* dip was reported by Kelley as between 5° and 10°. Geophysical tests were conducted in four test wells installed four to five miles apart and bracketing the proposed irrigation site. All of these sites are east of the Rio Grande (they are indicated by black dots in Figure 3.1). These tests gave bed dips ranging from 1.5° to 7° and strikes varying from due north to N.35°E. Given dipping and probably intertonguing beds of variable permeability, the "ability to transmit water is likely to vary greatly within a relatively short distance". Hearne concludes reasonably enough that "water flows parallel to the beds much more readily than perpendicular to the beds".[21] The quantification of this anisotropic behavior is addressed later. The final general geologic note concerns the greater grain size and better sorting of the alluvium bordering the Rio Grande and its local tributaries. Up to two miles wide along the main river and up to 100 feet thick, this sediment should display significantly different hydraulic behavior than the rest of the formation.

Hearne transforms a generalized geologic cross-section into a typical model cross-section, both of which are depicted in Figure 3.2. He then applies his understanding of the prototype - within the constraints noted - to construct a finite difference model.[22] Overall, the modeled area is about 25 miles wide by about 40 miles long. The two essentials identified - dipping, discontinuous beds - are preserved in the model, though with some simplification. A curving strike was considered, but operational problems proved insurmountable.[23] The model consists of 22 dipping rows, 22 columns and 22 layers, arranged in a "slumping stack of pancakes". The grid blocks vary in size, with the smallest being one mile square in the xy plane and 650 feet thick; these blocks are concentrated in the area of greatest interest around the Pueblos. West of the Rio Grande, the east-west block dimension gradually increases from the minimum of one mile; north and south of the Pueblos, the north-south block dimension also increases with distance away from the area of interest. In the far field the largest blocks have dimensions $x = 4.5$ miles $= y$, with a thickness of 1950 feet (southwest and northwest corners of the model). Hearne takes the strike to

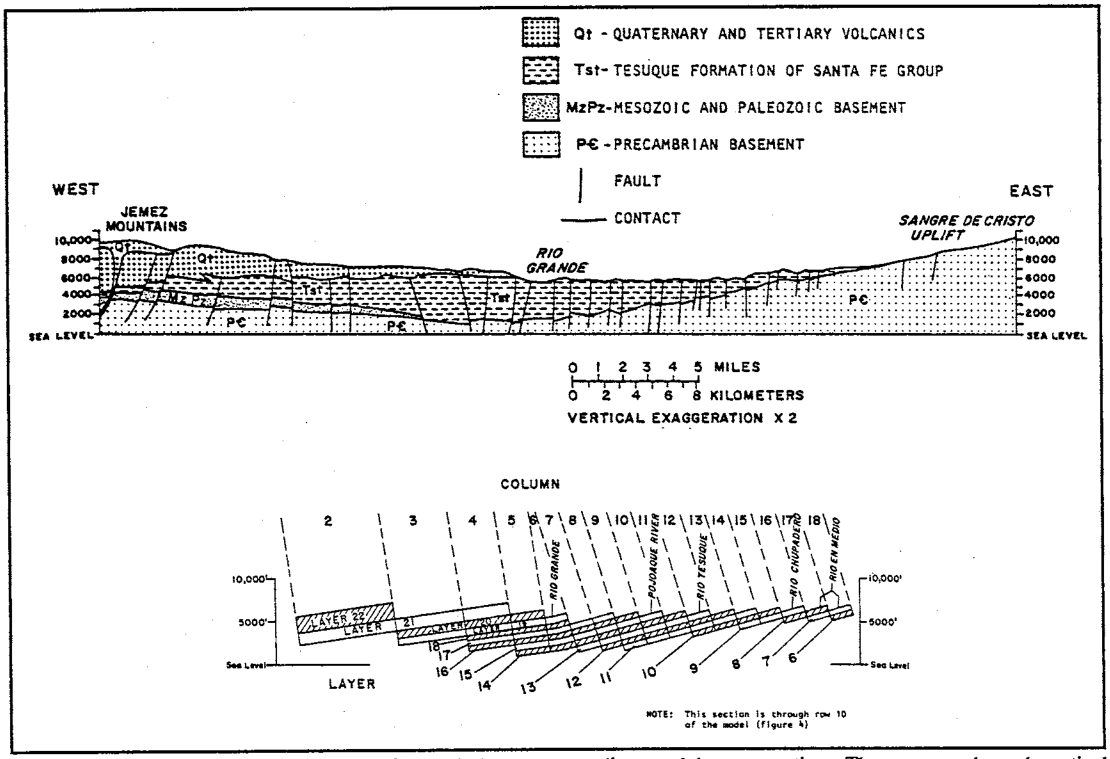


Figure 3.2: Hearne's geologic cross-section and the corresponding model cross-section. The same scale and vertical exaggeration applies to both figures. Both sections from Hearne (1985a), pp.5,6.

be N. 25° E. The dip is thus to the west northwest, at a uniform 8° on the east of the Rio Grande, and 4° on the west side. He orients the *xy* plane of the model within these bedding planes, with the *z* direction orthogonal to this plane, thereby simplifying the mathematics of the model. The greatest overall model thickness occurs beneath the Rio Grande, reflecting at least some of the geological reports (see Figures 3.2, 3.3). The governing equation within the model is given by:

$$\frac{\partial}{\partial x}\left(K_x \frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial h}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z \frac{\partial h}{\partial z}\right) = S_s \frac{\partial h}{\partial t} + W(x,y,z,t)$$

where K_x, K_y, K_z ≡ hydraulic conductivity values in the x,y,z directions, respectively [L/t];

h ≡ the hydraulic head [L];

S_s ≡ specific storage [L^{-1}];

W ≡ pumping or recharge volume per unit aquifer volume per unit time [t^{-1}];

t ≡ time [t]. [24]

Some of these decisions reflect Hearne's conviction that modeling the recognized complexity of the real system "is beyond the present capabilities of modeling techniques".[25] His conceptual model is a further simplification of the already limited geological data. His discussion makes it clear that some choices are motivated as much by the need for mathematical convenience as by some imagined fidelity to the real system - the size of the grid blocks being only the most obvious example. The situation bears out in practice the thought of Chia-Shen Chen, who once said: "To conceptualize is to simplify to the level that your mathematical tools can be applied".[26] Models are not always entirely *willingly* simplified.

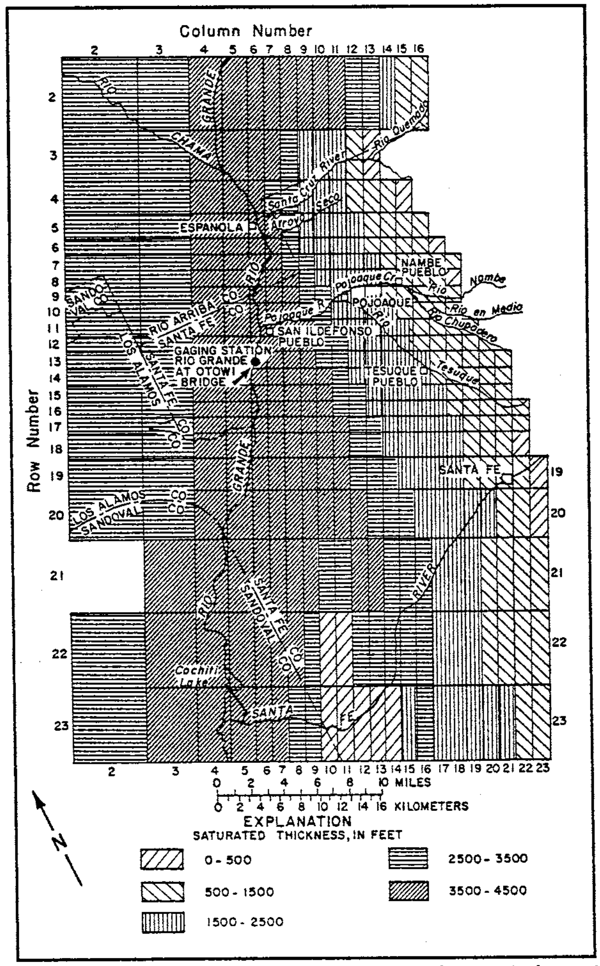


Figure 3.3: Plan view of Hearne's model, showing block size and the saturated thickness. From Hearne (1985a), p.7.

On the one hand, simplification reduces the number of geometric and parametric decisions required; on the other hand, these decisions are effectively replaced by broader assumptions. At some point this may raise the modeler's level of discomfort as the analogy between prototype and model becomes harder to defend. For example, larger grid blocks reduce the size of the matrix equation for head, but result in more averaging of behavior within the system, possibly obscuring important details. Cell thickness likewise introduces both computational convenience and vertical averaging. Similarly, Hearne's maximum system thickness is to the small end of the plausible range, but conserves modeling resources; he comments on the possible importance of the unknown thickness, and plans to later test his model's sensitivity to this decision.

The second leg of the analogy consists of representative values for the needed aquifer characteristics. These must be estimated and entered for each of thousands of discrete cells in the model.

In this phase Hearne relies on previous field work, including pump tests conducted by C.V. Theis in 1962. Uncertainty over the thickness of the water producing layers complicates these parameter calculations. Hearne also notes that "no aquifer tests of the Tesuque aquifer system have been conducted long enough to determine the specific yield".[27] As mentioned earlier, besides the existing municipal wellfields, new test wells were developed (see Figure 3.1), one each within the Pueblo Grants of San Ildefonso, Pojoaque, Nambe, and Tesuque. However,

> At each of the four sites where wells were constructed to test the aquifer characteristics, a sequence of interbedded gravel, sand, silt, and clay was penetrated... Attempts to correlate individual beds from one site to another using geophysical logs were unsuccessful.[28]

These and similar observations[29] prompt Hearne to neatly express both the central dilemma of modelers, and also one of the standard strategies (italics added):

> *The data available for estimating these characteristics indicate that the characteristics vary spatially within the aquifer system. However, the data are not adequate to describe this variation as a general pattern.* At most places in the Pojoaque River basin, a well several hundred feet deep will intersect at least one sandy unit that transmits water readily. This sandy unit will be overlain and underlain by units less able to transmit water that serve to isolate this unit from other sandy units. However, it is doubtful if either the transmissive units or the confining units are very extensive. Within a few hundred feet, the unit may thin to extinction, be terminated by a fault that positions it adjacent to a unit of different character, or change in character from a sandy unit to a silty unit that does not transmit water as readily.
>
> *The approach adopted for this report was to estimate, from the available data, the most likely average value for each aquifer characteristic throughout the entire basin.*[30]

Hearne has available the results of an aquifer test he conducted in 1975 after the BIA proposed its irrigation plan. As described above, this plan identified the Tesuque aquifer as the source of irrigation water. The pump test was intended to help determine if the new withdrawals would adversely impact fully appropriated surface flows. Hearne's initial report on the pump test is dated 1980, and the Water-Supply Paper (2206) was published in 1985.[31] An account of his procedure and analysis can give further insight into the reasoning behind his developing basin-scale analogy.

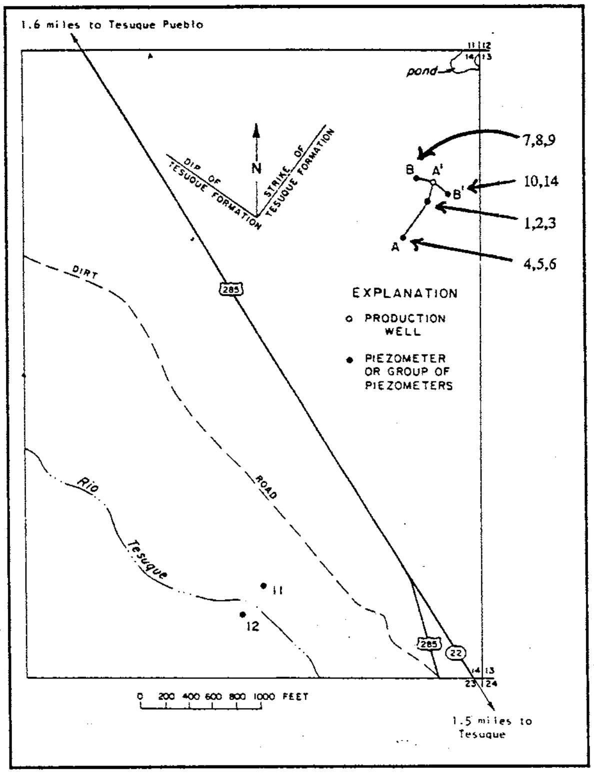


Figure 3.4: Location and arrangement of the Tesuque Pueblo pump test. A' is the production well; piezometers as numbered (see Fig.3.5). Santa Fe is to the south on Hwy. 285. Adapted from Hearne (1985b), p.4.

The production well for the 1975 pump test was located on the Tesuque Pueblo Grant (see Figure 3.4). The particular well used was chosen because "[the site] appeared to be typical of the Tesuque aquifer system, and there was an existing well capable of withdrawing enough water to test the characteristics of the system".[32] About 200 feet of screen was installed in the sand units from 300-820 feet below the surface (the *production zone*). This well had not been pumped for two years prior to the test, since its use in a step-drawdown test in February of 1974. 13 piezometers were gathered in four groups around the production well. Within the deepest of each group, gamma, density, neutron and caliper logs were taken prior to casing the holes; within the production well, a temperature log replaced the caliper test. Well completion depths varied in elevation over about 800 feet (see Figure 3.5). At least one well in each group was screened below the production zone of the pumping well, and at least one piezometer in each group was completed toward the bottom of the beds comprising this zone. The two groups southwest of the pumping well also had one piezometer screened in the beds toward the top of the zone. One piezometer downdip was completed above these beds. In every case, "It was thought that the response in these beds would be most sensitive to the hydraulic conductivity normal to the bedding plane".[33] Finally, two piezometers were installed in the alluvium near the Rio Tesuque 3750 feet southwest of the production well, presumably to test for indications of stream capture.

Hydraulic gradients parallel and normal to the beds were estimated to be, respectively: $\partial h/\partial x = 0.02$, and $0.012 < \partial h/\partial z < 0.13$. Static head levels were also analyzed to provide hydraulic gradients,

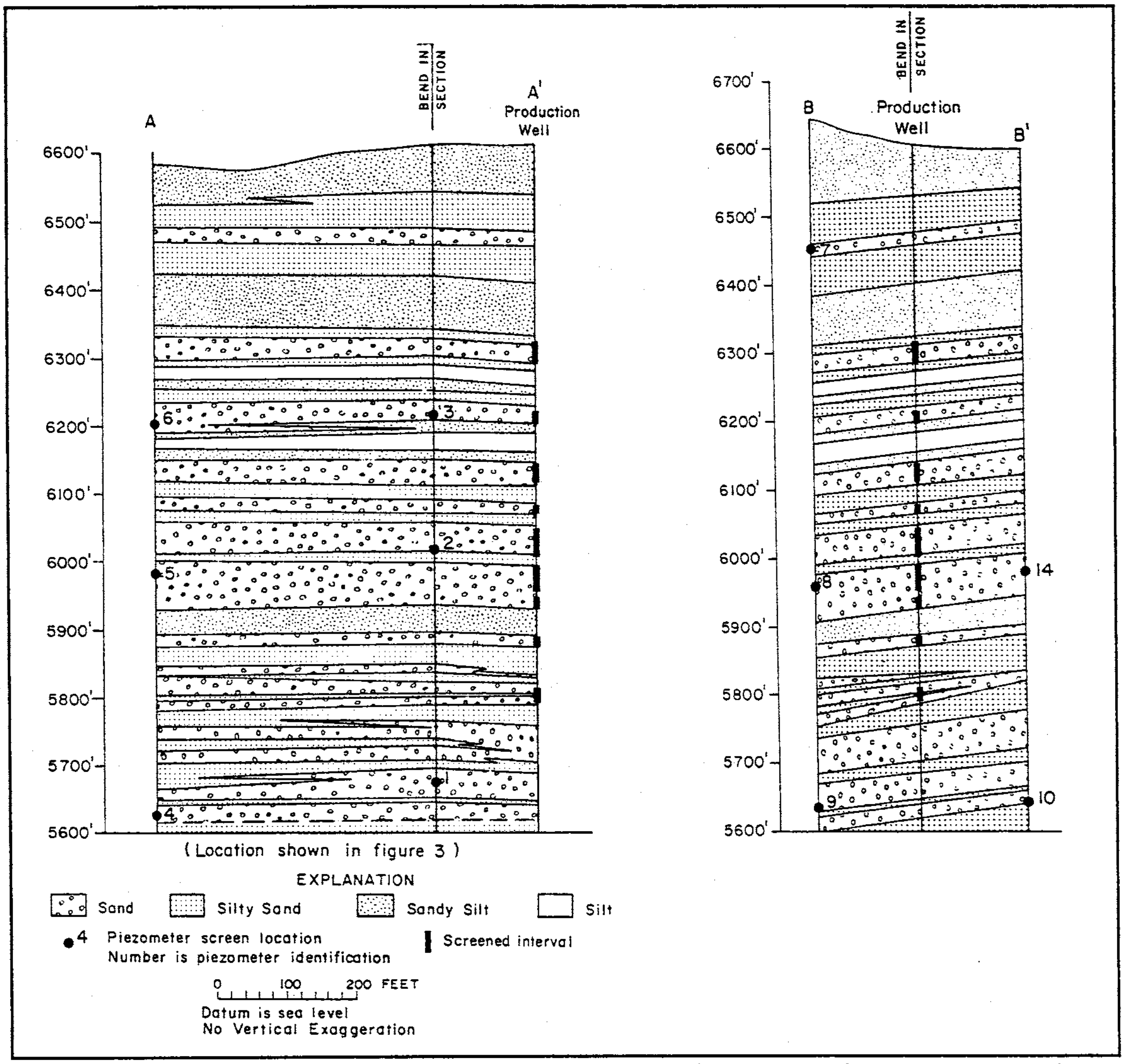


Figure 3.5: Sections A-A' and B-B' from the pump test (see Fig.3.4), showing the screened intervals of the production and monitoring wells. Adapted from Hearne (1985b), pp.6,7.

and hence another measure of anisotropy in the system. The resulting anisotropy ratio ranged from 0.004 to 0.2. The equation used to make these calculations assumes essentially horizontal flow in the plane containing the dip.[34] Extrapolating the result to the rest of the Tesuque aquifer, and assuming an average hydraulic conductivity parallel to the beds of 1 ft/day, average conductivity normal to the beds is just the anisotropy ratio, or 0.004 ft/day.

The pump test took place from 16 December to 29 December, 1975. Drawdown under a stress of 320 gal/min was monitored, as was the later recovery of head levels. A water quality sample was taken from the production well during the test; this sample was "believed to be representative of the water contained in beds of the Tesuque aquifer".[35] Piezometer 4 was located 550 feet southwest of the production well; completed below the production bed, it flowed with an artesian head, and also differed significantly in water quality. These facts indicated to Hearne "the extent to which the units are hydraulically separated by the intervening silts and clays".[36]

The piezometers screened above and below the production zone showed minimal head changes during the test: any response was "apparently less than a few tenths of a foot and was masked by noise in the system".[37] Piezometers screened about 100 feet below the top of the 500 foot thick production zone responded to pumping, but less so than those screened 200 feet lower. Both the apparent thickness of the sandy bed and the length of the screened interval were greater at this lower elevation, as shown in Figure 3.5. Monitored heads in the upper part of the production zone actually rose during part of the pumping period, leading Hearne to speculate: "water appears to have been flowing from underlying beds in the production zone through the production well" to the beds 200 feet higher in elevation.[38] Be that as it may, the two distant piezometers straddling the Rio Tesuque showed "no apparent correlation with the aquifer test",[39] a fact more readily interpreted as evidence of little or no hydraulic connection, due to the dipping beds.

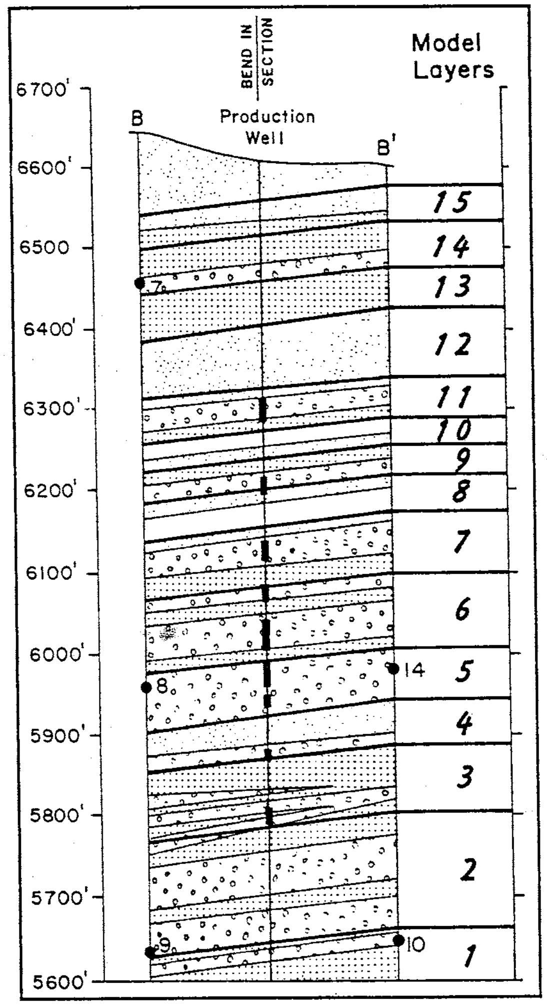


Figure 3.6: Simplification from well logs to modeled layers at the Tesuque pump test. From Hearne (1985b), p.16.

Hearne then constructed a finite difference model of the local area to further explore system response to the pump test. This model bears some resemblance to the BIA model that we have been discussing, except that it was on a much smaller scale, the strike was oriented N.35°E., and the dip was a uniform 7°. Limiting the model size to 4 mi^2 was justified by the negligible effects there of "pump testing" the model, as were no-flow conditions on the northwest and southwest boundaries (specified-flux boundaries were designated elsewhere).[40] Model parameters were assigned to each of 30 beds based on the borehole tests mentioned above. These beds were then grouped into 15 model layers (see Figure 3.6). In a microcosm of the larger modeling problem of interest here, Hearne again stresses the importance of a credible model/prototype analogy:

> The model of the aquifer test... cannot incorporate the full complexity of the... system and the stress applied to that system. However, the validity of the model depends in part on the extent to which its description reflects these complexities... Each cell [of the model] must be described in terms of the estimated ability of the corresponding beds of the Tesuque aquifer system to store and transmit water. This is accomplished by assuming a degree of homogeneity that almost certainly does not exist in the aquifer system itself. The assumption is necessary because the specific nature of the heterogeneity of the complex system is largely unknown. The success of the model rests on the dual assumptions that the effect on the response of the aquifer system of a single nonuniformity is negligable and that many nonuniformities randomly located throughout the system produce a homogeneous system.[41]

This is a good summary of the argument for the use of effective parameter values. Hearne calibrated his pump test model to steady conditions - under which simulated flow was from southeast to northwest - and then adjusted parameters further to better match the response to pumping. Some of these calibrated values end up larger or smaller than those determined directly from aquifer test analysis. Hearne also justified the adjustments on a more theoretical basis: "On the larger scale of a basin model, the effective hydraulic conductivity is probably less" than the test showed, due to the "discontinuity of the more permeable beds". The "effective hydraulic conductivity normal to the beds may be higher", on the other hand, due to the discontinuity of the less permeable beds.[42] Hearne appears inclined to believe the anisotropy ratio resulting from model calibration (0.00005) may be representative of the smaller aquifer test site; but that the value derived from field-measured steady-state gradients (0.004) is more likely representative of the basin as a whole. On this basis, the original thickness of the pump test model was reduced after early simulations showed the strongly anisotropic system was insensitive to events in the lowermost layers. The elimination of these layers reduced "costs of computer time and storage capacity".[43] It is interesting that in the end Hearne finds the static field measurements more revealing in some ways than the pump test. Leaving aside any further details of this earlier experience, let us return to the larger scale BIA model with which we began.

Hearne now considers several clues as to the aquifer characteristics. These include other well tests that C.V. Theis (1962) and R.L. Griggs (1964), for example, conducted on the other side of the Rio Grande, his own more recent investigations, laboratory studies, and sediment samples (none, apparently, from deeper than 100 feet). On this basis, he assigns both a plausible range and a likely value for four parameters: the hydraulic conductivity parallel to the beds; the anisotropy ratio (hydraulic conductivity normal to the bedding plane divided by that within the plane); the specific storage; and the specific yield of the unconfined portions of the aquifer. Hearne's standard simulation will make use of his most likely average values, as shown in Table 3.1.

Table 3.1: Aquifer characteristics of the Hearne model, Tesuque aquifer[44]

Aquifer Characteristic	Lower limit of plausible range	Most likely average value	Upper limit of plausible range
Hydraulic conductivity (parallel to beds)	0.5	1.0	2.0
Anisotropy ratio	0.001	0.003	0.01
Specific storage [ft^{-1}]	1×10^{-6}	2×10^{-6}	1×10^{-5}
Specific yield [/]	0.10	0.15	0.20

The third leg of the model/prototype analogy consists of boundary conditions. Hearne's discussion of boundaries in the prototype and in his model illustrates Bredehoeft and Konikow's comment on the role of *a priori* expectations in the construction of conceptual models. Hearne had, of course, some experience with the real system, as detailed above; here we refer to expectations prior to the running of the model. For example, Hearne assigns a constant head boundary condition to all grid blocks through which the Rio Grande or the Santa Cruz Rivers flow, linking the head to the land elevation as estimated from topographic maps. He clearly expects the maintenance of these head levels to be a steady feature of the regional flow system (see Figure 3.7).

The Pojoaque River and its tributaries, however, are *expected* to respond differently, and they are represented in the model as head-dependent regions. (It is, of course, possible for head to remain more or less constant in the vicinity of a head-dependent boundary). Routing the streams through adjacent grid blocks requires calculating a leakance coefficient; Hearne equates a lumped parameter model of flow through the channel bottom to the Darcy flow through blocks adjacent to the streams. This calculation is buttressed with the results of three seepage tests performed along the Rio Grande in 1967-8. Highly variable tests results are averaged to show a mean gain of 0.3 cfs per mile. Expected gain at the Rio Grande would later emerge as an important model evaluation tool. Hearne puts the average as "about 0.5 cfs per mile or less" and puts the upper limit as "certainly no more than 1 cfs".[45]

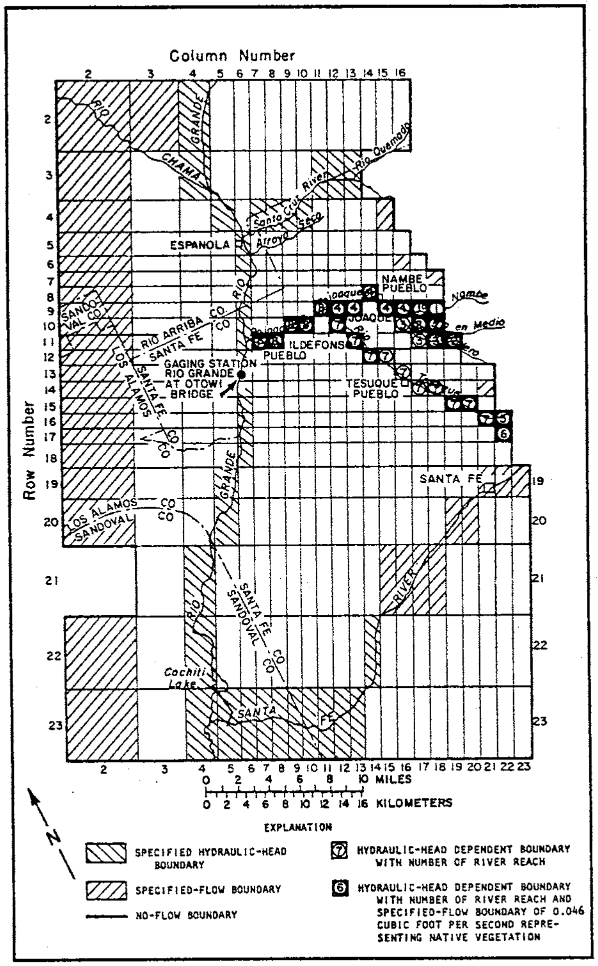


Figure 3.7: Boundary conditions. From Hearne (1985a), p.14.

More generally, Hearne considers stream discharge records for several rivers in the basin, and on this basis argues the reasonableness of assigning constant head, constant flux, or head-dependent boundaries to various grid blocks through which the rivers pass. Sometimes the model itself is used in a preliminary or tentative way to guide these decisions. Tentative results, for example, cause Hearne to allow the Santa Fe river to flow from grid blocks of constant flux to downstream blocks of constant head. The constant rates in the upper river were also chosen based on the same preliminary steady-state simulations. Similarly, in setting up a tractable model, Hearne announces that the boundary at the contact between the Tesuque formation and the Sangre de Cristo Mountains is "sufficiently distant that the effect on the response to the proposed withdrawals is negligible".[46] He bases this choice on the fact that "reasonable" steady state conditions were achieved taking this boundary as a hydraulic boundary (no flow), although it is not, strictly speaking, a geological boundary.

Anisotropy is sometimes used, as in this effective parameter model, to mimic the overall effects of heterogeneity. Although mathematically advantageous, secondary problems can arise in handling processes that are themselves heterogeneity-dependent. For example, Hearne resorts to special conditions to include storm runoff and vegetation effects in his model. In particular, aquifer recharge from runoff was limited to the perimeter of the model, and the estimated transpiration losses were evenly divided along a certain reach of the Pojoaque River. Recharge is not discussed as a separate issue, but Hearne adds the expected inflow from the mountains to the headwater blocks of the rivers.

With these parameters and distributions in hand, Hearne is ready to begin fine-tuning his model. He anticipates certain objections by saying: "The terms 'calibration' and 'verification' have been avoided because of their misleading connotations of control, accuracy, and certainty".[47] He begins with the two standard history-matching exercises. The historic "steady state" is said to have existed in 1946, before major groundwater withdrawals began to supply the municipalities. The minor historical pumping within the Pojoaque Basin for irrigation is ignored: "The effect... on the ground-water system is assumed to be negligible".[48] Hearne "calibrates" his model by first running it as constructed, with no pumping or other non-historic stresses, to find its steady state or predevelopment condition. The simulated water budget, in the form of fluxes at boundaries and head levels in wells, is then compared to the measured or inferred values in the basin some 40 years earlier. The 71 wells used for this purpose are mostly in a square region that includes the four Pueblo Grants.[49]

A rough scaling from his plotted water levels (see Figure 3.8) shows the mean difference between measured and simulated heads to be between 60 and 70 feet. Some of this difference is ascribed to the size of the grid blocks, since wells do not necessarily or even usually map onto grid block centers:

> For example, wells 13, 14, and 15 are so close together that they are represented by the same cell even though the measured hydraulic heads range from 6,235 to 6,391 feet. A slight change in the locations of wells 33 and 54 would have placed them in cells in which the simulated hydraulic heads of 6,805 and 6,020 feet, respectively, compare more favorably with the measured hydraulic heads of 6,808 and 5,989 feet, respectively.[50]

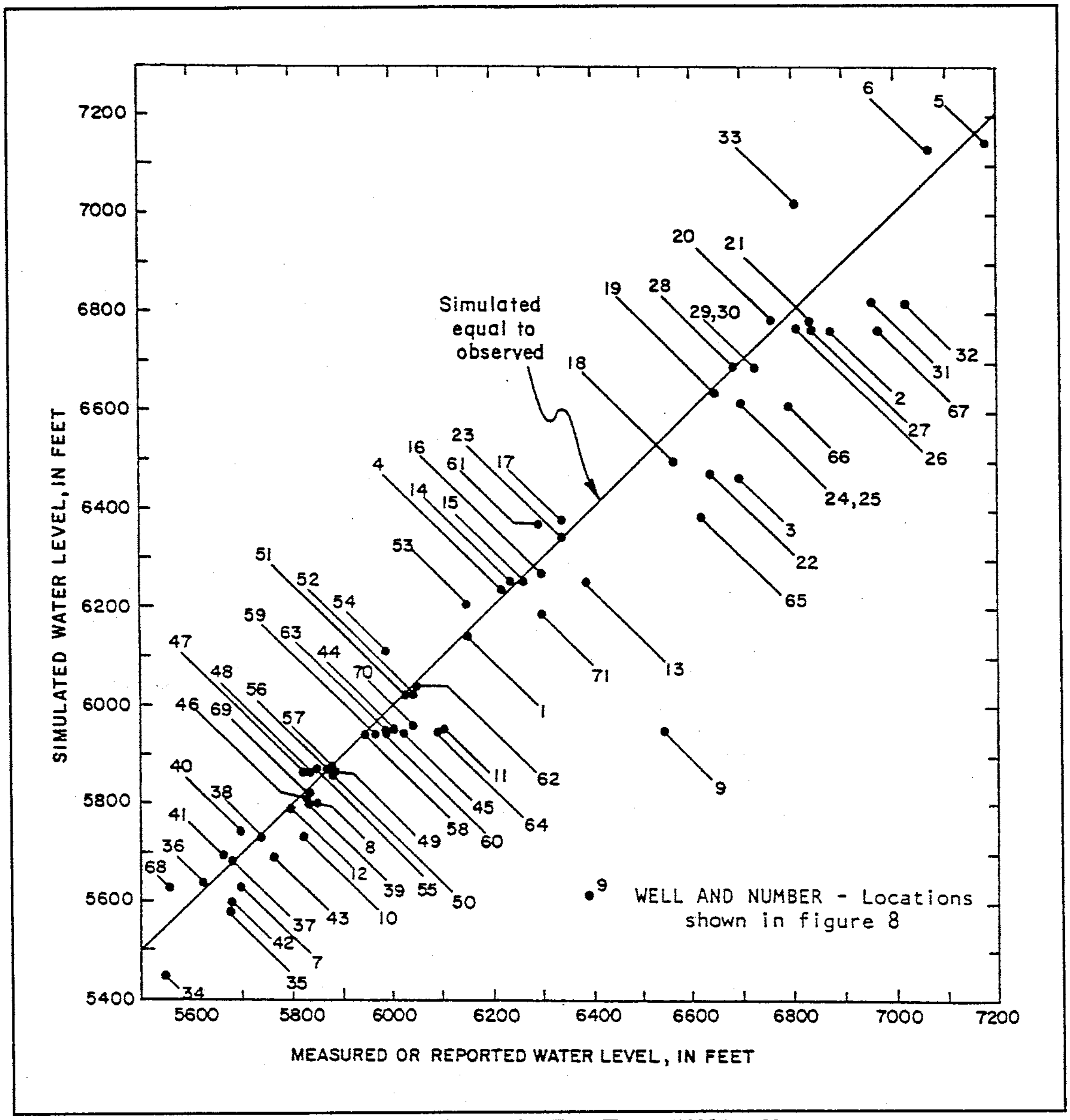


Figure 3.8: Results of the steady-state history-matching exercise. From Hearne (1985a), p.22.

Portions of the error are also attributed to wells outside model boundaries being compared to the nearest cell within the model, or perched conditions that are misleading in terms of apparent water levels. In short, the size of the grid blocks complicates an evaluation of the history-matching exercise. The mismatch may be particularly severe near the mountains, where head levels change very rapidly. And: "Finally, some of the variability in this relationship is undoubtedly due to the model's not representing the detailed heterogeneity of the Tesuque aquifer system".[51]

It is only at this point in Hearne's report that we learn that the structure and values adopted or derived earlier - and presented as eminently plausible - were not the first or only ones considered. Preliminary models with different structures and boundaries were discarded because "the steady-state condition that they simulated was too dissimilar to that established in the prototype".[52] The structural differences between these models are discussed very briefly. Among other things, orienting the strike due north and placing the northern model boundary south of the Santa Cruz River were tried and rejected. The principal criterion advanced as decisively in favor of the chosen model structure and extent is that "the steady state condition was acceptable without adjusting the aquifer characteristics".[53] Moreover, the fit between simulated and recorded steady-state head levels did not significantly improve for the "preliminary models" even *with* parameter adjustments. Hearne prefers a model that does not require him to ignore or wildly distort his best professional judgement on the aquifer properties. A considerable investment of time and effort has led him to certain conclusions about the plausible range of field parameters; it is evidently much more uncertain just how the geometry and boundaries of the model might best capture the essential behavior of the actual system. It is therefore the latter part of the model that Hearne is most inclined to adjust.

The chosen model could be viewed as a demonstration of how a specific structure could combine with reasonable boundaries and properties to produce the historical situation in the basin; *i.e.*, it could be viewed as a causal argument. There is, however, no suggestion that either 1) this is the actual configuration of the aquifer; nor that 2) this is a uniquely accurate analog to the prototype. The situation does serve to illustrate the iterative modeling procedure: in calibrating a model based on his best professional judgment, Hearne has evidently been forced to reconsider his reasoning. Certain assumptions or approximations evidently no longer seemed tenable, and were modified or discarded, even though there was no reason yet to think that the model was working, nor any exact indication how these particular choices affected the results. The simple explanation is usually that preliminary results can show that the assembled model is clearly *not* working, and thus *some*thing must change. Choice of parts to revise is clearly a matter of professional judgment; in this case Hearne seems inclined to use his best estimate of effective parameters to test his choices of model geometry and scale.

Since the model's stated purpose is intimately involved with possible hydraulic connections between groundwater and surface water, Hearne must re-evaluate steady-state gains and losses. He revisits the historical discharge data for all the streams, and reviews previous estimates of gains from or losses to groundwater. For example, he argues that Spiegel and Baldwin's estimate of Rio Grande gains of 25 cfs between the gaging stations at Otowi and Cochiti are probably a little high; Hearne's model shows a steady-state discharge to the river of 22.06 cfs, an amount that agrees with his previous analysis.[54] Similarly, Hearne concludes that "the net flow between the Pojoaque River and its tributaries and the Tesuque aquifer system is probably less than 5 cfs and could be either to of from the aquifer". His steady state simulations result in flow rates to or from the aquifer within this range: "The simulated flow between ground water and surface water is compatible with observed data. A more precise adjustment of the model is not justified by available data".[55] For all of these reasons, including the head-matching exercise in Figure 3.8, at some point he "considered the comparison between simulated and historical water levels to be acceptable" for steady-state conditions.[56]

The second stage in evaluating the model introduces the historic water withdrawals from 1946 to 1977, and assumes these rates remained constant from 1977 to 1980 (data were not yet available for the latter years). This is a second phase history-matching effort (which Hearne refuses to call "verification"). The goal is to see if the model, now calibrated to an acceptable steady state condition in 1946, can also produce a reasonable match to known head levels in existing wells over the next thirty years or so. The first thing is to once again review the actual recorded response to stress. The significant withdrawals have been from the municipal fields; once again, "the effect... on the ground-water system is assumed to be negligible" from the minor historical pumping within the Pojoaque Basin;[57] the non-municipal pumping has mainly been for irrigation, and mainly confined to areas near the Rio Grande, Rio Pojoaque, and Rio Tesuque.

The volume of pumping within the basin has increased fairly continuously; the number of wells has increased, and the pumping volume at most wells has increased over time. The non-pumping water levels within the basin wells have also declined - somewhat erratically for the Los Alamos field and more steadily for the others. Unfortunately, the stresses that Hearne incorporates into his history-matching are rather distant from the actual area of interest (see Figure 3.1). As a result, the grid blocks are very large in the vicinity of some of the history-matched wells - up to three miles across. This is expected to damp the drawdown response.[58] In each of eight test locations (the four well fields and the four Pueblo grants), Hearne compares without much comment the actual water levels to the nearest grid block center. These usually show the same general behavior, though the short-term fluctuations are smoothed, and the model underestimates the drawdown in every case.

Turning to an issue of particular importance to the model's purpose, Hearne notes that the effects on surface water flows from historical municipal pumping "have not been quantified".[59] Little history matching is therefore available for this aspect of model performance. Simulated changes in head in the Pojoaque basin due to historic pumping, on the other hand, "vary from negligible along the mountain front [a recharge boundary] to a few feet just east of the Rio Grande".[60] These slight changes, both up and down, are essentially in the noise of the model. Another unmeasured influence is exerted by historical diversions for irrigation; in 1980 sufficient surface water was diverted to irrigate 3,700 acres.[61] In some respects, therefore, the historical record is inadequate for "verification" purposes. There is no indication in the report of the extent of calibration/"verification" cycling to achieve the reported fit.

Table 3.2: Sources of Water in Pojoaque River Basin under the BIA-proposed irrigation plan. Data gathered from Hearne (1985), p.32; all units are in [cfs (Acre·Ft/year)].

		Tribal	Non-Tribal	Municipal
Historic use (1976)	groundwater	negligible	negligible	11.24 (8,140)
Proposed additions	groundwater	37.45 (27,113)	8.43 (6,103)	0.46 (333)
	Rio Nambe	8.59 (6,219)	1.86 (1,347)	
	Rio en Medio	0.84 (608)	0.11 (80)	
	Rio Chupadera	0.17 (123)	0.13 (94)	
	Rio Tesuque	1.08 (782)	0.36 (261)	
Expected return flow from *all* applied water		26%: 12.39 (8,970)	50%: 5.44 (3,938)	28%: 0.13 (94)
Net new withdrawals		25.06 (18,143)	2.99 (2,165)	0.33 (239)
Total proposed groundwater withdrawals: 11.24 + (25.06+2.99+0.33) = 39.63 cfs (28,691 A·ft/yr)				

The third stage of analysis, which Hearne considers part 2 of stage 2, is an exploration of two scenarios. The first (Hearne's *null hypothesis*) assumes the recent pumping rates at the well fields will remain constant for the next fifty years. These withdrawals total 2.651 x 10^9 gallons per year, or 11.24 cfs. The second (Hearne's *alternative hypothesis*) adds a version of the proposed pumping and surface water diversions to the established uses. In the latter case, Hearne plans to stress the model by introducing pumping wells and reducing streamflows in accord with the proposed surface water diversions. The difference of these two hypothetical results is exactly the modeled effect of the BIA irrigation plan. Slight but potentially important differences between the actual and simulated proposal enter because model sources and sinks are necessarily located at node centers, and may not correspond to exact geographical sites. Hearne also adjusts the proposed withdrawals and diversions to give certain areas credit for the expected return flow. The modeled sources of the new water supply, as detailed in the proposal and modified by Hearne's calculations, are summarized in Table 3.2.

The existing well fields are geographically dispersed enough to be represented within his model at at least two nodes, and at as many as four. Withdrawals are incorporated as specified-head boundary conditions at the center of the nearest node, between 300 and 950 feet below the water table. Return flow, on the other hand, is distributed to node centers at the water table, about 350 feet below ground surface (see Figure 3.9). Hearne presents the results at various points of interest. These include the municipal wellfields and the Pojoaque River basin. The apportioning of the drawdown between continued historic pumping and the new plan on both the municipal well fields and on streamflow is shown in Table 3.3. Not surprisingly, the projected drawdowns at the existing wellfields are primarily the result of continuing the historic withdrawals. The proposal only calls for the municipal pumping rate of 11.24 cfs to be increased by 0.46 cfs (or 0.33 cfs if return flow is considered); this increase would be distributed among some 18 active wells. The remainder of the 28.39 cfs increase would be pumped from the 708 new wells completed along the Pojoaque River and its tributaries. The details of the BIA plan show these wells irrigating as little as one

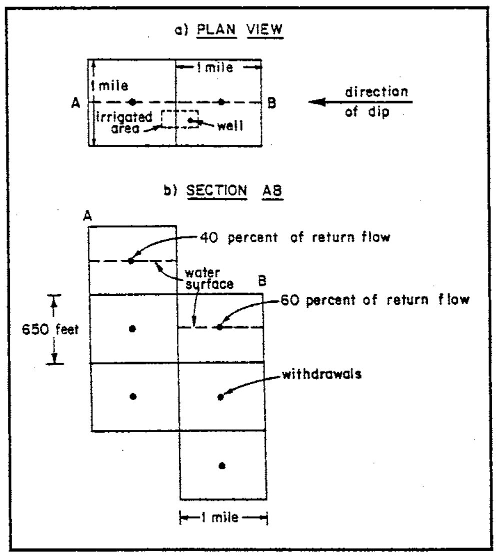


Figure 3.9: Distribution of withdrawals and return flow in Hearne's model. From Hearne (1985a), p.33.

acre each and as much as 138 acres. Here the simulated effects of the proposed plan are much more dramatic. The maximum decline in unconfined surface cells is about 143 feet at a node near the Rio Tesuque. Withdrawals are concentrated in the underlying confined cells, however. Maximum decline in head is 335 feet, just north of (and below) the most affected surface cell (see Figure 3.10).

Table 3.3: Summary of Hearne model results for the year 2030; data gathered from Hearne (1985a), pp.44-50. A: contrasting head declines due to continued historic pumping v. additional drawdown due to increased pumpage in the alternative hypothesis. B: contrasting effect on streamflow from null hypothesis v. the alternative hypothesis.

<table>
<tr><td colspan="8">A: Head changes due to continued historic pumping [ft] : additional drawdown due to BIA proposed stresses [ft]</td></tr>
<tr><td>Los Alamos:

6.8: 6.2
15.5: 3.4
15.1: 3.8
7.4: 5.5</td><td>Guaje Canyon:

31.5: 0.6
14.2: 4.1</td><td>Pajarito Mesa:

47.9: 0.5
46.8: 0.4</td><td>Buckman:

48.7: 13.1
46.3: 17.2</td><td>San Ildefonso Pueblo Grant:

8.3: 84.0</td><td>Pojoaque Pueblo Grant:

0.8: 210.6</td><td>Nambe Pueblo Grant:

- : 160</td><td>Tesuque Pueblo Grant:

- : 184</td></tr>
<tr><td colspan="8">B: (steady-state simulation of streamflow in 1980), Change in streamflow due to continued historic pumping [cfs] : change due to BIA plan [cfs]</td></tr>
<tr><td colspan="2">Rio Grande

(21.04), -0.97 : -0.90</td><td colspan="2">Santa Cruz River

(2.56), +0.09 : +1.08</td><td colspan="2">Santa Fe River

(4.34), -0.06 : -0.13</td><td colspan="2">Pojoaque River and tributaries

(0.09), +0.22 : +2.40</td></tr>
</table>

Referring again to Table 3.3, the standard simulation shows streamflow, on the other hand, to be relatively unaffected by the new stresses. Two of the streams are actually expected to gain flow by 2030, and overall 86% of the new water supply comes from groundwater. Flow in the Rio Grande in 2030 is projected to decrease by 10.13 cfs without irrigation development (2.46 cfs due to stream capture and 7.67 from diverted tributary flow), and to decline by 18.77 cfs (5.63 capture and 13.14 diverted) if the BIA plan is implemented. Though highly variable, the mean flow in the Rio Grande at the Otowi station is about 1000 cfs, so either of these scenarios projects less than a 2 per cent decrease from the mean. The model suggests, however, that the contribution from captured streamflow will increase slowly over time, a change that would make sense in the prototype due to increased vertical gradients in the highly transmissive riparian sediments. By the year 2080, modeled depletion of storage has declined to 80% of pumpage. Hearne concludes: "According to the digital model simulation, storage will probably be a significant source of water for several centuries...[but] changes in head may make it impossible to

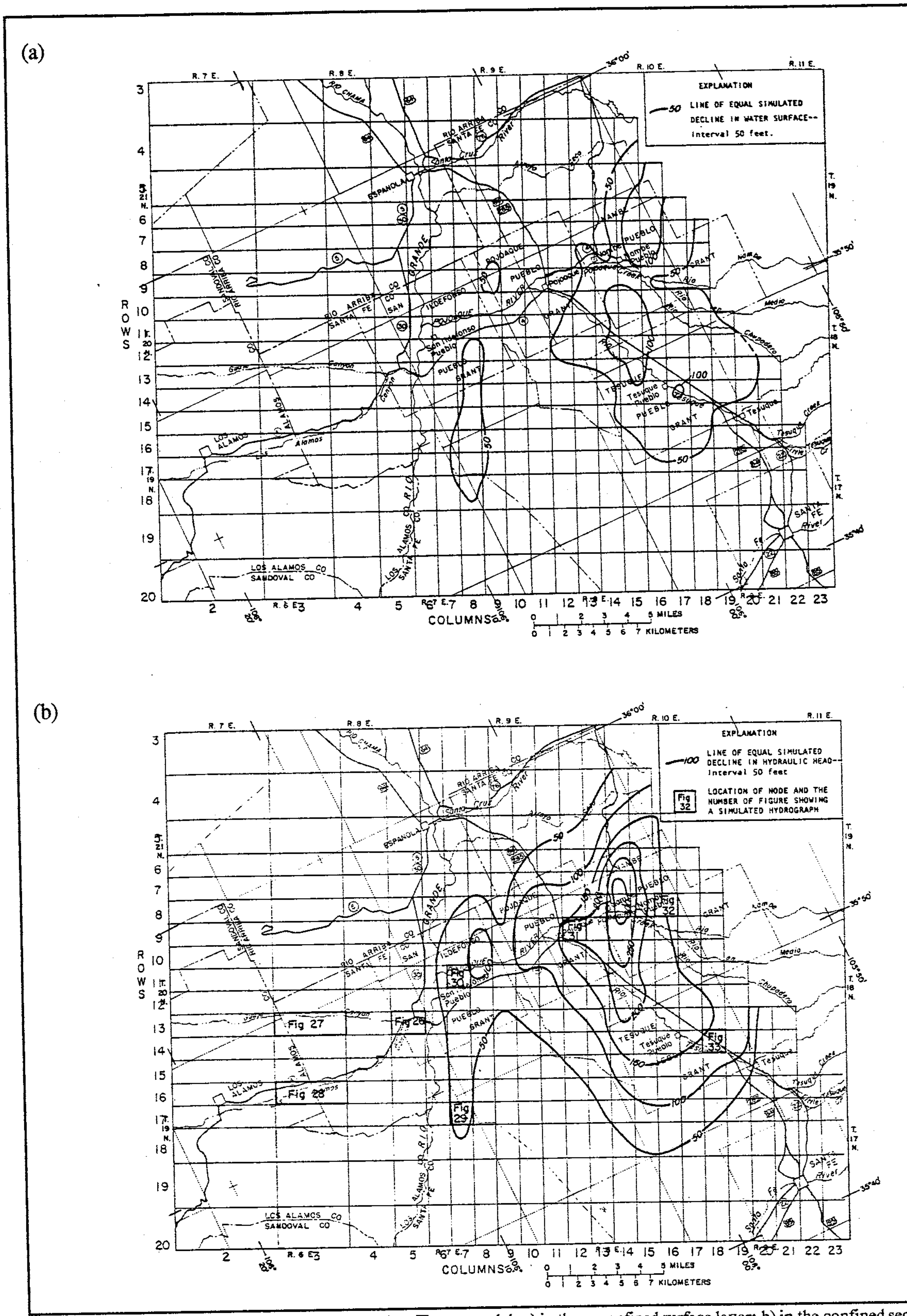


Figure 3.10: Simulated drawdown due to the BIA plan, Hearne model. a) in the unconfined surface layer; b) in the confined second layer. From Hearne (1985a), pp.46, 57.

continue withdrawals indefinitely at the specified locations".[62]

The final sections of Hearne's report give the results of a sensitivity analysis of his model. He tests the effects (both steady-state and transient where appropriate) of varying values for the four principal aquifer characteristics and also the formation thickness. These effects are presented for the same eight locations of interest discussed throughout the report: the four well fields and the four Pueblo Grant nodes. Hearne's estimate of thickness favors the smaller of published values, so he is most interested in testing the effects of greater thickness. Increased thickness increases the percentage of new supply from storage, with correspondingly less from the Pojoaque River.[63] As would be expected, his lowest plausible hydraulic conductivity increases drawdown and decreases stream capture; storage now provides 89% of the new withdrawals in the year 2030. The upper limit value for conductivity produces the opposite effect, but storage still provides 80% of the new water supply.[64] Varying the anisotropy ratio produces a "mixed response" in which the contribution from storage remains about 86% of new withdrawals, but the remaining 14% is taken from different streams. A smaller anisotropy ratio causes more water to be captured from the Rio Grande, and less from its tributaries.[65]

A steady-state simulation makes no use of specific storage or specific yield, so the effect of varying these parameters can only be tested during transient history-matching. Three of the well fields are insensitive to specific storage varied within the plausible range; the Buckman field and the Pueblo nodes show slightly less drawdown with increased storage.[66] The simulated transient response shows a mixed response to variations in specific yield. Naturally, drawdown is inversely related to the change in yield, but the source of water derived from the wells is also affected. At the lowest plausible value, storage contributes 83% of the new supply, while at the upper limit this increases to 87.5%.[67]

Hearne then uses this analysis to "translate the uncertainty with which aquifer characteristics are known into uncertainty in the predicted response".[68] He apportions the uncertainty among the five tested choices: the model is most sensitive to variations in hydraulic conductivity (40%), with aquifer thickness the next most important (30%). The other aquifer characteristics account for the rest of the uncertainty. In his conclusion Hearne repeats that "the structure and boundaries were revised until the simulated steady-state condition satisfactorily reproduced historical data". Furthermore, "it was not necessary to revise the initial estimates of aquifer characteristics to approximate the steady-state condition within the limits of the accuracy of available data". And so finally, "the simulated historical phase produced no response which forced rejection of the model based on historical data",[69] a tepid endorsement that calls to mind his initial discussion of modeling: "The state of the art of digital modeling does not permit a statement on the confidence limits bounding the projections made by the model. This still needs to be done subjectively".[70]

Hearne's quantitative sensitivity assessment is typical of standard practice in that it does not provide the maximum plausible deviation from his standard model prediction. That is, he does not provide data on the expected effect on stream capture if aquifer thickness, storage and specific storage all decrease to their minimum plausible values while conductivity and the anisotropy ratio both increase to their plausible maximums. Nor is maximum drawdown analyzed for the opposite estimations. Finally, a principal component in analogy-building is the description of structural essentials - an estimate of which is not directly addressed in the sensitivity analysis. As a result: "The confidence in the predicted response... needs to be based on a subjective appraisal of the analogy between the Tesuque aquifer system and the model".[71]

3.3 McAda and Wasiolek Model: Simple Structure and Patchwork Parameters

As it happened, those interested in further subjective evaluations of basin models would not have long to wait. The Santa Fe Metropolitan Water Board, interested in the long-term viability of their Buckman and Santa Fe field supplies, consulted with the USGS in 1982 concerning a predictive model of the region. Although covering the area of interest (see Figure 3.11), the focus of the Hearne model was further north within the Pueblo Grants, as indicated by the location of its smallest grid blocks. The Hearne model also does not consider withdrawals from the Santa Fe wellfield. In any case, it is not the policy of the USGS that any one model is definitive at a particular site,[72] and revising the Hearne model to shift its focus was not part of the original proposal. As a result, Douglas McAda of the USGS and Maryann Wasiolek of the New Mexico State Engineer Office prepared a second groundwater model of the area, designed to focus on the long-term fate of the municipal wells without regard for the BIA proposal. Their report is bound as USGS Water-Resources Investigation Report 87-4056 (1988).[73]

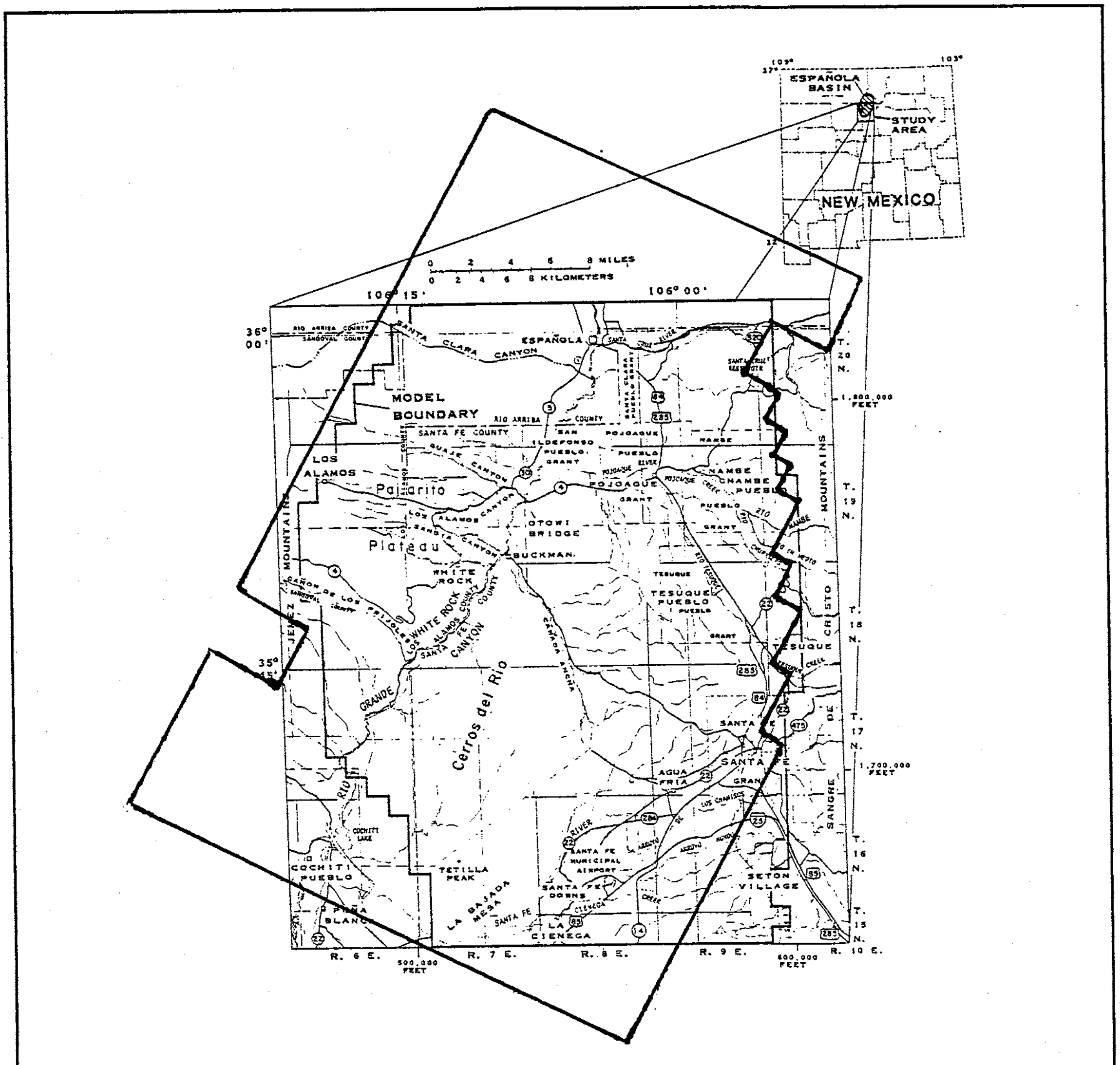


Figure 3.11: Outline of the 4-layer model. From McAda and Wasiolek (1988), p.3, with Hearne's boundaries added for comparison.

As in the case of Hearne's model, the project proposed to evaluate two specific groundwater development schemes. These were not the same "hypotheses" that Hearne considered, however. Unlike Hearne's *null hypothesis*, McAda and Wasiolek's *small water demand* scenario did not hold pumping volumes to the rates of the late 1970s; changes in demand were expected at both Los Alamos and Santa Fe. At the Los Alamos wells, maximum pumpage was figured to increase 36% from the maximum historical withdrawals. At the Santa Fe field, maximum pumpage was expected to increase just 1.3% from the maximum precedent, while at the Buckman field maximum pumpage was pegged at 25% less than the maximum recorded. For the *large water demand* projections, these same figures became: Los Alamos wells, 36% increase; Santa Fe field, 1.3% increase; Buckman field, 53% increase.[74] Our interest in this second model stems from the fact that, despite its stated objectives, it was also used by the State Engineer Office to evaluate the BIA proposal. We will therefore review the modeling rationale employed by McAda and Wasiolek, the results of their history-matching exercises, and the projected impact of the BIA plan; we will not discuss the original scenarios.

The general procedure used in modeling regional flow has already been amply illustrated in the narrative of the Hearne model; only the decisions and conclusions differed as the region was reconsidered. The published report of McAda and Wasiolek is less forthcoming in terms of explaining some of their decisions in the three principal areas of model/prototype analogy (geometry, boundaries, aquifer characteristics), perhaps because the authors did not anticipate the potential legal interest for which Hearne was prepared. They post standard disclaimers on the model's predictive accuracy, especially with respect to any particular location within the modeled area.[75] One notable difference from Hearne's introduction is that McAda and Wasiolek do *not* stress directly the need for a close structual analogy between model and prototype.

The remodelers rely, for the most part, on the same geological reports that Hearne used, and open their report with a similar survey of the relevant formations within the Espanola Basin.[76] There is little that is new. Their description of the geology gives the flavor of the modeling task as seen by McAda and Wasiolek. The natural boundaries of the regional flow system are described as the Pajarito fault zone on the west, the Cerrillos uplift on the south, and the contact with the Sangre de Cristo uplift on the east (see Figure 3.1). The natural boundaries to the north are more distant, consisting of "bedrock highs of the Picuris block and the southern end of the Brazos uplift". They note the 1963 report of Spiegel and Baldwin in which Tesuque sediments are described as "several thousand feet of pinkish-tan soft arkosic, silty sandstone and minor conglomerate and siltstone". The thickness of these sediments is "unknown". They recognize the reported "generally westward" dip of the formation and its effect on outcrop composition from east to west. The Tesuque is usually said to consist of three principal members.

The Nambe Member (coarse-grained but fining-upward conglomeratic arkosic sediment sequences) is in "both fault and depositional contact with the bedrock of the Sangre de Cristo Mountains". The Skull Ridge Member (cross-bedded sandstones and "minor but numerous" ash and mudstone beds) conformably overlies the Nambe and outcrops further west. Lastly, the Pojoaque Member (interbedded sandstones, mudstones and gravel) outcrops near the Rio Grande.

The younger Puye Formation (sand and pebbles of volcanic origin; thickness 60-700') outcrops to the west of the Rio Grande. McAda and Wasiolek are confident the Puye and Tesuque Formations are hydraulically connected; they therefore take the Puye to be part of the Tesuque aquifer system. South and west of Santa Fe, the Ancha Formation (granite-derived gravels interbedded with some silt and sand; thickness up to 300') unconformably overlies the Tesuque. The Ancha and Tesuque are apparently distinguished with difficulty, with the main differences being that the Ancha is usually unsaturated, more permeable, and less dipping. McAda and Wasiolek take it to be part of the Tesuque as well. The modeled area is shown in Figure 3.11, with the outline of Hearne's model superimposed.

The modelers then review the available data to construct a potentiometric surface map. They are careful to caution that the resulting contours are directly decipherable only by "assuming that the aquifer is horizontally isotropic, [so that] the direction of groundwater flow is perpendicular to the potentiametric contours".[77] Recharge to the system is said to occur mainly as mountain-front recharge through the fractured bedrock of the Sangre de Cristos, with a smaller contribution from stream channel losses. Geochemical records are combined with geological insights to give some further indication of the destination of water flowing through the system. As painted by Spiegel and other observers, the gross distribution of water is fairly well illustrated. It remains to construct a more detailed picture, if possible. Hearne had said earlier:

> Each cell [of the model] must be described in terms of the estimated ability of the corresponding beds of the Tesuque aquifer system to store and transmit water. This is accomplished by assuming a degree of homogeneity that almost certainly does not exist in the aquifer system itself. The assumption is necessary because the specific nature of the heterogeneity of the complex system is largely unknown.

Dipping discontinuous beds are the major controlling factors in Hearne's model. McAda and Wasiolek choose not to attempt a geometric analogy of the sort Hearne thought critical. They instead represent the Tesuque aquifer within their study area as four horizontal layers, assuming a further degree of simplicity "that almost certainly does not exist in the aquifer system itself" (see Figure 3.12). In so doing, McAda and Wasiolek demonstrate another common approach to conceptual modeling. In this alternative technique, aquifer characteristics are thrown into particular prominence because they must compensate for the relative lack of structural controls within the model (see Figure 3.13). The major benefit in this approach is the relative simplicity and speed of model construction. The major drawback is the additional burden placed on history-matching as validation, since there is even less reason to think that model geometry is akin to prototype geology. There is, however, no obvious preference between the two modeling approaches, since it could be, of course, that the two models are much more like one another, than either one is like the real system.

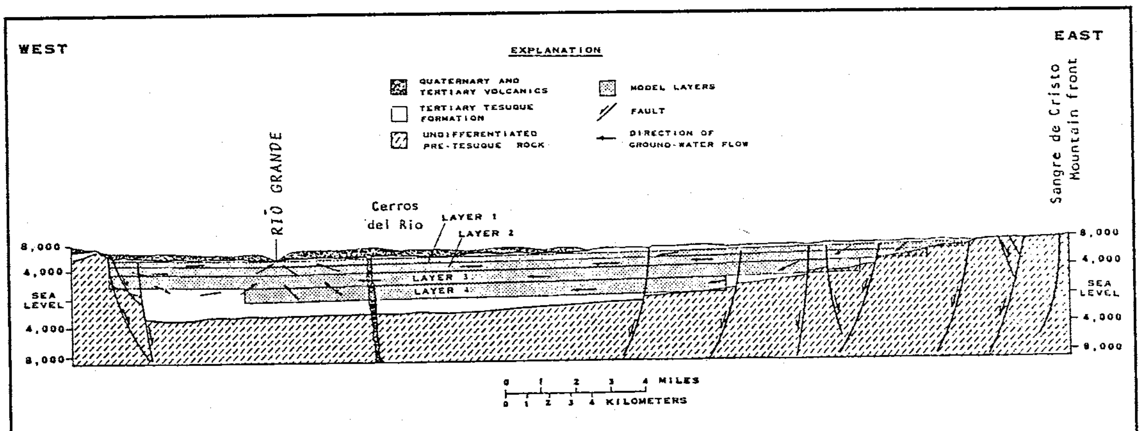
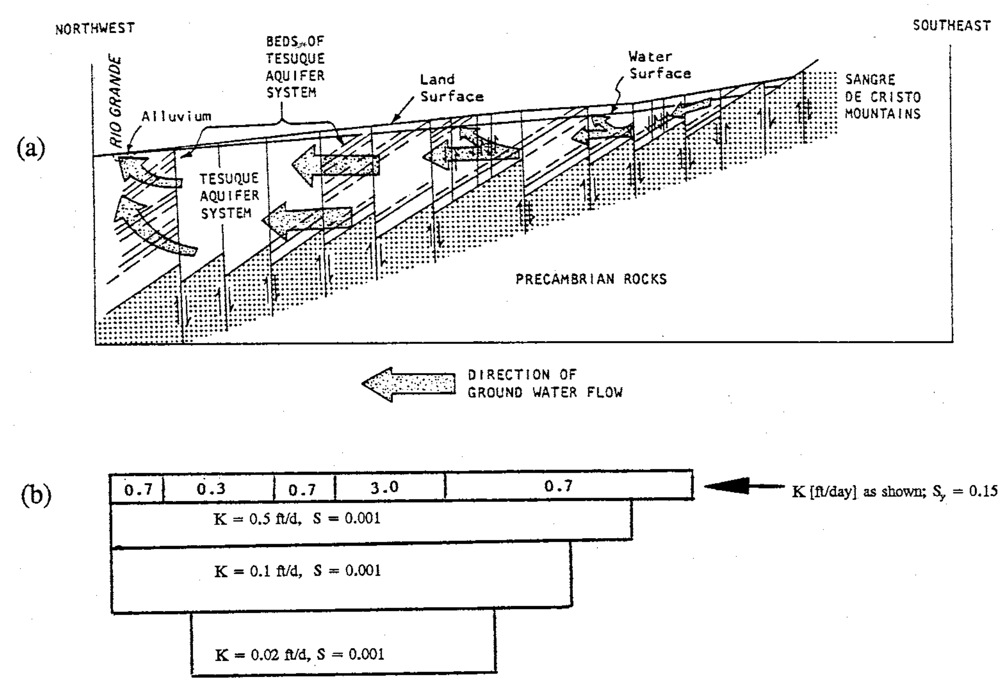


Figure 3.12: Cross-section of 4-layer model. Arrows show the direction of groundwater flow. From McAda and Wasiolek (1988), p.19.

McAda and Wasiolek caution that the four layers of the model "do not represent specific units within the Tesuque system; they are used to discretize the aquifer into three dimensions in order to simulate the vertical component of flow".[78] At the edges of the model, some subset of the upper layers are active. Mountain front recharge is thus channeled into the upper layers. All four layers are active within about 60% of the model. Each of the layers is divided into 33 rows and up to 25 columns, resulting in uniform grid block areas of $x = 1$ mile $= y$, and a total area of 712 mi^2 (see Figure 3.14). The *xy* plane is oriented within the bedding plane, i.e., horizontally: this "orientation was assumed to align with the principal components of hydraulic conductivity", despite the dip. The thickness of the uppermost, unconfined layer is dependent on the location of the water table, but has a maximum thickness of 800'. Thicknesses of the underlying confined second, third and fourth layers are 1,200', 1,800' and 1,800', respectively, giving a maximum system thickness of 5600'. Most of the wells of interest are completed within the upper layer. A critical factor in the success of their model depends on their ability to capture vertical gradient effects within such a thick layer without the benefit of dipping and/or discontinuous structural components. This procedure will be taken up in the discussion of aquifer characteristics.



Typical parameter variability defined by east-west sections A-A', B-B' through Figure 3.15a,b.

Figure 3.13: Alternate models. a) q_x is downdip; flow is influenced by anisotropy. From Hearne (1985b), p.3. b) q_x is horizontal; aquifer characteristics compensate for structural simplicity.

The governing equation for flow through this system is the same generic hydrodynamic equation as that employed by Hearne:

$$\frac{\partial}{\partial x}(K_x \frac{\partial h}{\partial x}) + \frac{\partial}{\partial y}(K_y \frac{\partial h}{\partial y}) + \frac{\partial}{\partial z}(K_z \frac{\partial h}{\partial z}) - W = S_s \frac{\partial h}{\partial t}$$

where K_x, K_y, K_z ≡ hydraulic conductivity values in the x,y,z directions, respectively;

h ≡ the hydraulic head;

S_s ≡ specific storage;

W ≡ volume recharged or discharged per unit aquifer volume per unit time [t^{-1}];

t ≡ time [t].[79]

McAda and Wasiolek follow their more general disclaimers about model results with a more specific warning. Since nodes (grid block centers) do not necessarily correspond to the geographical location of wells, the accuracy of the model at particular locations may be weak. They attribute, as did Hearne, a certain amount of the likely errors in simulation to this lack of correspondence. On the other hand, there is no effort to justify the geometry of the model in terms of the geological detail presented earlier. McAda and Wasiolek do note that representing the aquifer with more and thinner layers would have required "more specific information on the amount of pumpage for each layer, including pumpage proportioned from single wells". This information is unavailable.

Moving on to their second stage of model construction, McAda and Wasiolek consider the necessary boundary conditions. Their choices again differ from those of Hearne, and are mapped in Figure 3.14 (*cf.* Figure 3.7, p.68). Hearne put the western boundary west of Rio Grande and of constant flux; McAda and Wasiolek place the same Pajarito fault zone in a different place, and use a constant flux north of the river in their layer 1, while south of it they have a constant head. The river itself enjoyed a constant head in Hearne's model; McAda and Wasiolek utilize a head dependent flux. Except in cells through which rivers pass, Hearne's northern and eastern boundaries are designated no flow; he inputs the recharge off the Sangre de Cristos directly into the headwater cells of the rivers.[80] McAda and Wasiolek put a constant recharging flux all along the eastern border (apparently in all layers), while the northern boundary is a constant head in their upper three layers. Although both models describe the eastern boundary as being the contact with the Sangre de Cristos, the locations differ. In both models, the northern boundary does not correspond to actual physical contacts. Like Hearne, McAda and Wasiolek rely on a steady-state simulation to show that the artificial northern boundary is sufficiently distant to have negligible effects on model performance.[81] Like Hearne, McAda and Wasiolek take the southwest corner of their model to contribute a constant flux.

Modeled boundary conditions within the first (uppermost) layer also differed for stream flow. McAda and Wasiolek put head-dependent fluxes in most of the rivers. The hydraulic connection between stream and aquifer is central to the water managers' later interests, even if somewhat less critical to the modelers' stated objectives. Hearne used a method provided by C.V. Theis that corrected for dip as it equated flow through the channel bed to darcy flow through the underlying block. McAda and Wasiolek rely on a new study to estimate the leakance between stream channels and aquifer. McAda and Wasiolek calculate the conductance [L^2/t] of the channel as "the hydraulic conductivity of the riverbed multiplied by the area of the riverbed in a model cell divided by the thickness of the riverbed".[82] McAda and Wasiolek estimate riverbed conductivity to be 0.1 ft/day and the riverbed thickness to be 1 foot, and think it "feasible that the conductance may be as much as half an order of magnitude smaller or larger than these initial estimates".[83] In the steady-state calibration of the model, they varied these values slightly outside this range, resulting in a maximum conductance of 0.74 ft^2/sec in one cell in the northern reach of the Rio Grande, and a minimum value of 0.10 ft^2/sec for the Rio Pojoaque and for many reaches of its tributaries. The southeastern reaches of the Rio Tesuque are described as constant flux, as is the entirety of the Santa Fe River. In the latter case,

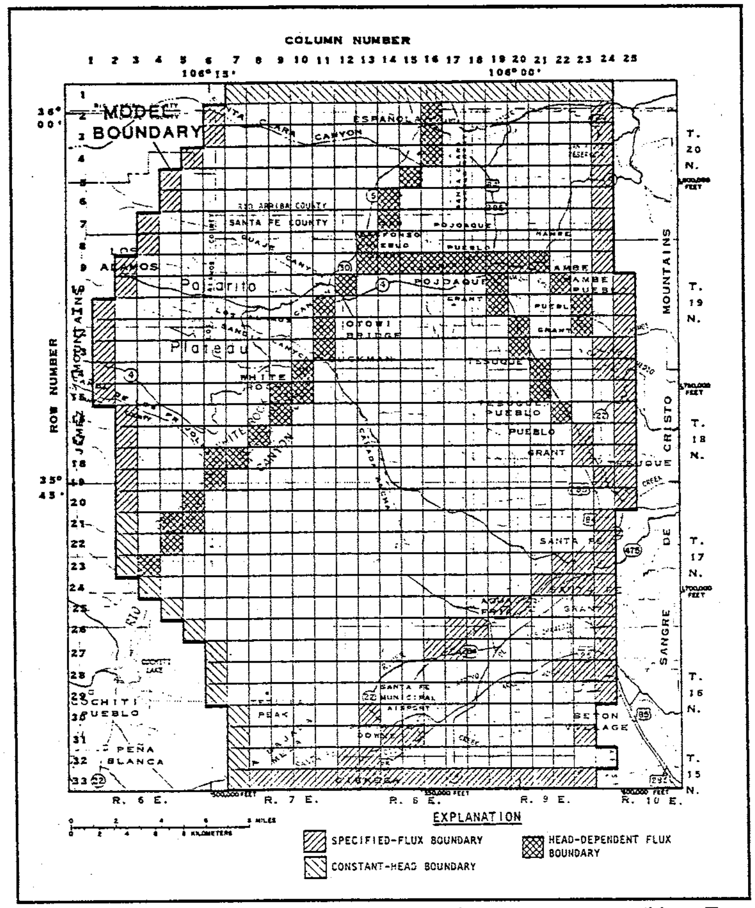


Figure 3.14: Plan view of the 4-layer model, showing the boundary conditions. From McAda and Wasiolek (1988), p.20.

"because the water table is below the level of the riverbed, a head-dependent flux boundary would not realistically represent the river". The gain to the stream has been estimated as 6.5 cfs; the modelers therefore decide to adjust "the amount and distribution of recharge throughout the simulation periods, depending upon the available flow in the river".[84]

The final area addressed concerns the aquifer characteristics. As noted above, these decisions take on particular importance because they must substitute for structural controls within the model. Several aquifer tests in the vicinity of the Los Alamos well fields suggest a range of hydraulic conductivity [ft/day] from $0.3 < K < 2$. Hearne's 1975 pump test (and review of other tests) give a plausible conductivity range further south and east of $0.3 < K < 2.8$. McAda and Wasiolek note various other tests have yielded estimates ranging from 0.2 to 20 feet per day. They review available investigations to characterize the local permeability and its variation across the basin. On this basis, they produce a detailed conductivity distribution; portions of the calibrated version are depicted in Figure 3.15a,b. The uppermost layer is the most variable for several reasons. Most of the information available characterizes material at depths of 2000' or less. Secondly, most of the producing wells in the region are completed within the uppermost layer. As a result, the most information is available for this layer, and this layer is probably the most critical to model performance. Thirdly, permeability is expected to decrease with depth, further damping the significance of detailed descriptions deep within the model.

The choices made by McAda and Wasiolek reflect this conceptualization of the system. The detailed patchwork (13 zones) of layer 1 has an average hydraulic conductivity of 1.1 ft/day, and ranges from 0.1 to 6.0 ft/day. The less variable layer 2 (4 zones) has a maximum conductivity of 0.56 ft/day, and a range of 0.1 to 2.0 ft/day. The conductivity does not vary within the two lowest layers. "On the basis of lithology", a uniform conductivity of 0.1 ft/day was assigned to layer 3, and 0.02 ft/day to layer 4. [85]

The specific yield is only needed for the uppermost unconfined layer. McAda and Wasiolek note that little site-specific information is available. Laboratory tests on material typical to the basin usually give values in a range from 0.10 to 0.20. They settle on a uniform value of 0.15 for virtually all of the basin, with only the Guaje Canyon wellfield having the smaller yield of 0.05. They note that the typical value (0.15) agrees with Hearne's estimate.[86]

Less, if possible, is known about the storage coefficient, needed for the three lower confined layers. They assume a specific storage of 1 x 10^{-6} per foot, and multiply by the model layer thicknesses to arrive at the coefficients. This results in storage coefficients ranging from 0.001 to 0.002. McAda and Wasiolek estimate the plausible range as 0.0004 to 0.005. They decide to use a uniform storage coefficient for all three confined layers of 0.001.[87]

McAda and Wasiolek rely on earlier estimates of the (vertical) anisotropy ratio, noting Koopman's (1975) estimated value of 0.04, and Hearne's estimated range of 0.001 to 0.01. Combining these two efforts into a single plausible range, McAda and

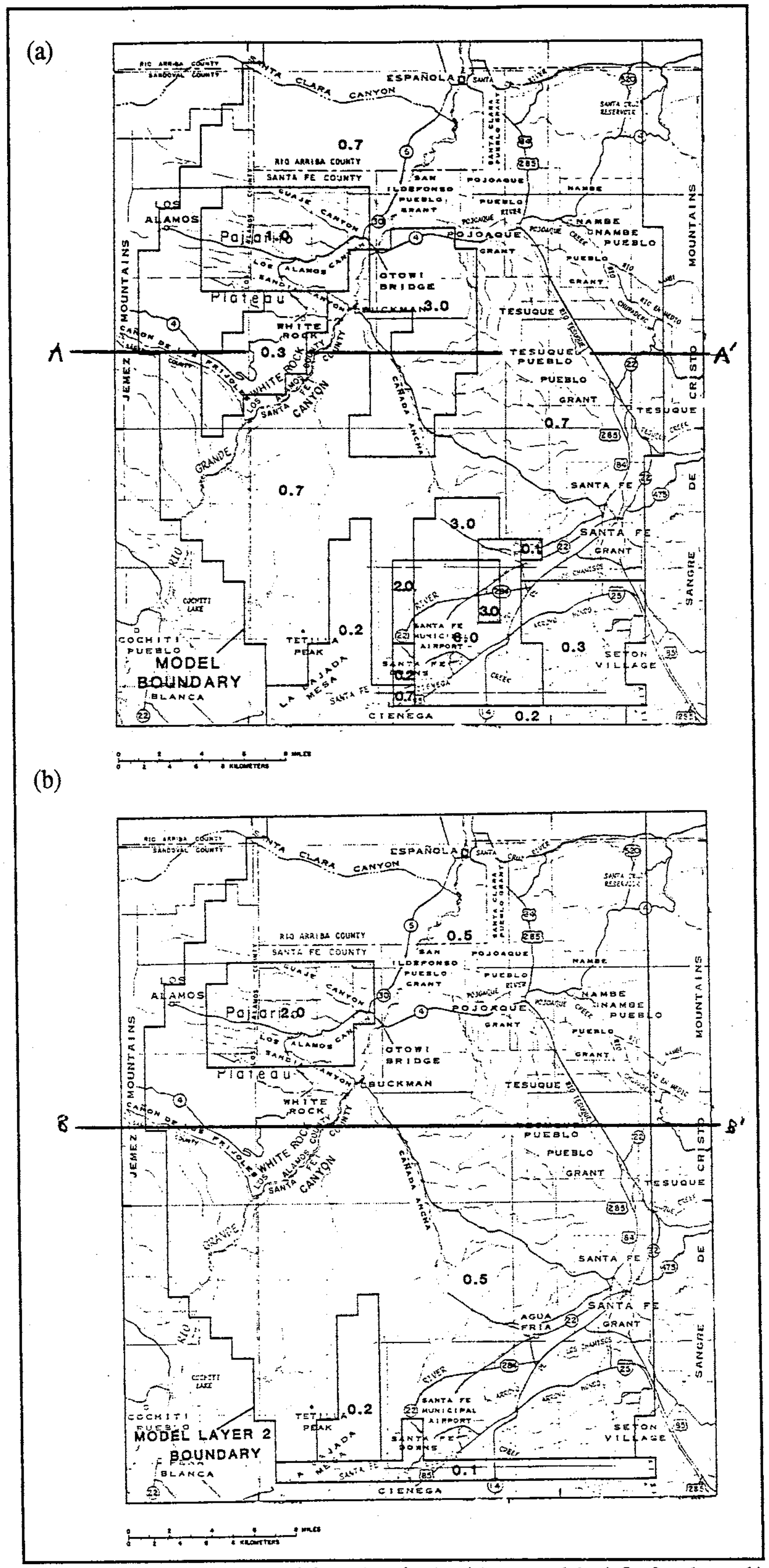


Figure 3.15: Hydraulic conductivity values in the 4-layer model. a) Surface layer; b) Second layer. From McAda and Wasiolek (1988), pp.24, 26.

Wasiolek then settle on 0.01. And finally, "because no data are available to estimate the horizontal anisotropy, no attempt was made to simulate horizontal anisotropy in the model. Therefore, it was assumed to be one".[88] The changing hydraulic conductivity across each of the upper layers will, of course, act as a large-scale horizontal anisotropy.

Estimates of mountain-front recharge and stream losses are used in preliminary simulations to arrive at constant flux values for the eastern boundary and for grid blocks through which streams pass. Reiland and Koopman give the mountain-front gain as between 0.7 and 3 cfs per mile; the mean loss rate of the upper Santa Fe River, on the other hand, "may be about 1 cfs per mile of stream".[89] McAda and Wasiolek plan to adjust such values further during model calibration. The problem of quantifying areal recharge provokes the following statement from the authors:

> It is recognized that the distribution of areal recharge is influenced by topography, land disturbance and other factors that are too numerous and whose interactions are too complex to account for, and that locally, recharge rates may vary from regional estimates. However, the amount and distribution of precipitation, permeability of the bedrock and soil cover, and evapotranspiration rate can be accounted for. Since little is known about the recharge rates, uniform values were assumed for areas of similar altitude and surface geology. Lee Wilson and Associates (1978, p.1.62) estimated that 0.28 inch per year is a low estimate for recharge to the aquifer from the area covered by the Santa Fe Group. A range of recharge rates, using this estimate as a guide, was tried and adjusted during model calibration.[90]

The modeling approach of McAda and Wasiolek thus leads to a patchwork distribution of recharge. Assigned values of areal recharge to the first layer vary (over 8 zones) from a low of 0.05 in/yr in the southwest, to 0.2 over much of the area (including throughout the Pueblo Grants), to 0.4 in/yr around Los Alamos and through the upper reaches of the Rio Grande, "which presumably are covered with more permeable alluvial sediments", to a maximum of 0.5 in/yr infiltrating the Ancha outcrops southwest of Santa Fe. Model simulations show all constant flux cells recharging the aquifer, except for a few cells in the lower reaches of the Santa Fe River and a few perimeter cells south and west of Santa Fe.[91]

The discussion of model calibration makes no mention of alternative model geometries. McAda and Wasiolek have fixed their geometry, and rely on fitted parameter values to capture the behavior of the prototype. Hearne, in contrast, has confidence in his estimates of effective aquifer characteristics, and relatively less in his geometry. Whereas Hearne reports that "the structure and boundaries were revised until the simulated steady-state condition satisfactorily reproduced historical data", McAda and Wasiolek pursue the same goal by different means: "Recharge and aquifer characteristics in the steady-state model were adjusted by a judgmental trial-and-error procedure".[92] As mentioned earlier in their report, the conductivity and thickness of stream channels are also treated flexibly by them within a factor of 5 to finalize the conductance of the riverbeds. The goal is the usual one of matching recorded heads and fluxes to their simulated counterparts as closely as possible by adjusting selected model choices within the limits of plausibility. The list of choices considered fixed or flexible simply differs in the two modeling strategies.

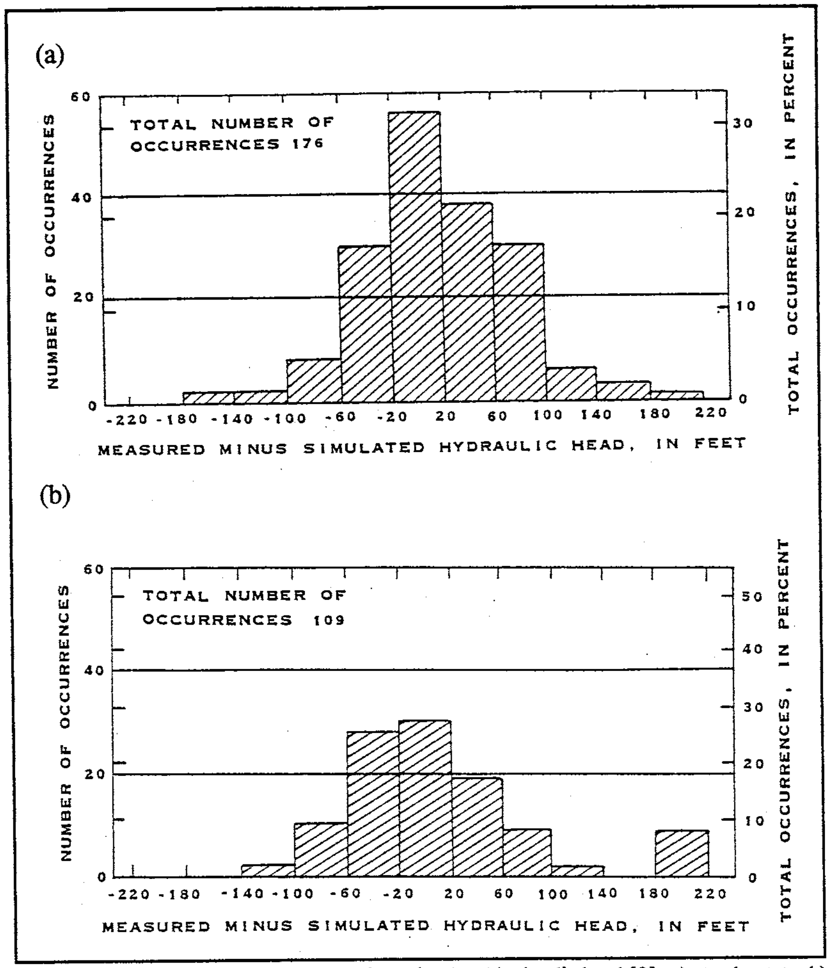


Figure 3.16: Histograms of measured minus simulated hydraulic head [ft]. a) steady-state; b) transient. From McAda and Wasiolek (1988), pp.42, 54.

The historic steady-state water levels in 176 wells are available for comparison with the simulated steady-state conditions. The mean difference between history and simulation is reported as 17.2 feet, with a standard deviation of 57.5 feet (see Figure 3.16a).[93] These errors are attributed to the same causes as mentioned by Hearne: 1) nodes not corresponding to well locations in the *xy* plane; 2) the screened interval of the well not coinciding with the center of a grid block; and 3) "the inability of the model to represent the detailed geology of the area".[94]

McAda and Wasiolek also provide a detailed steady-state water budget. This is presented in terms of 1) total sources and discharges; 2) each constant head or head-dependent flux boundary; and 3) stream flow. The overall budget shows a total of 73.8 cfs (53,500 A·ft/yr) moving through the system. The major sources are areal recharge (10.6 cfs), western mountain-front recharge (18.6 cfs), and eastern mountain-front recharge (38.5 cfs). Discharge is mainly accounted for through discharge to the Santa Fe River (6.5 cfs), discharge to the Pojoaque and its tributaries (7.3 cfs), groundwater throughflow at the southwest boundary (17.4 cfs), and discharge to the Rio Grande (39.3 cfs).[95]

The transient history-match from 1947 to 1982 again shows good agreement between simulated and measured or recorded head levels (see

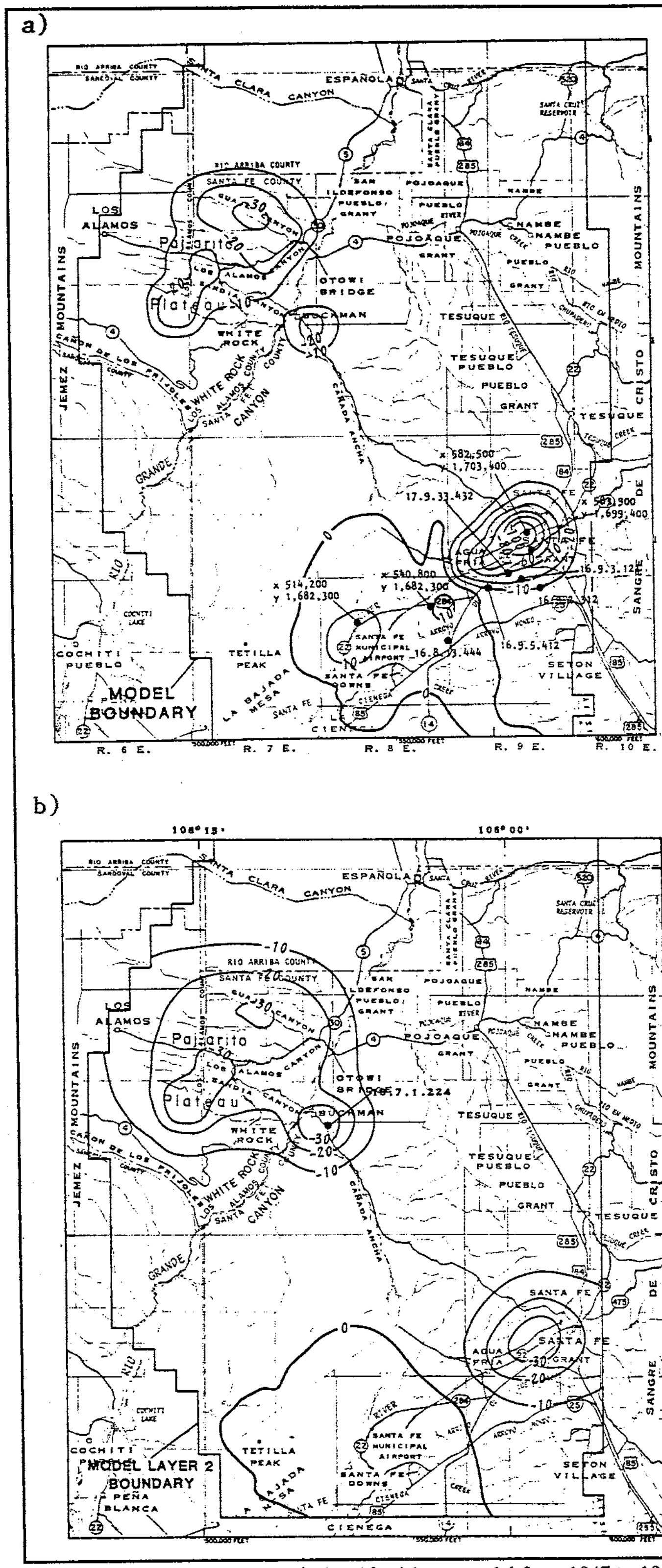


Figure 3.17: Simulated change in head in 4-layer model from 1947 to 1982. a) unconfined surface layer; b) confined second layer. From McAda and Wasiolek (1988), pp.52, 53.

Figures 3.16b, 3.17a,b). Total pumpage is 12.0 cfs; the increase from the 11.24 cfs stress assumed by Hearne reflects the inclusion of the Santa Fe well field and some other minor adjustments, such as domestic wells and sewage effluent irrigation.[96] Groundwater withdrawals for irrigation are again (as with Hearne) said to be insignificant. In response - after trial-and-error adjustments of the specific yield[97] - total flow through the system increases from 73.8 cfs to 83.4 cfs. Major changes from steady-state include: 1) 10.1 cfs from aquifer storage; 2) a decrease in eastern mountain-front and stream channel recharge from 38.5 cfs to 33.2 cfs; and 3) 4.7 cfs return flow from sewage effluent. Like Hearne, at some point McAda and Wasiolek can say: "The representation of the physical hydrogeological system in the model is substantiated by available data".[98] A sensitivity analysis of McAda and Wasiolek showed their model most sensitive to changes in hydraulic conductivity.[99]

3.5 Decisionmaking

The results of the large and small demand scenarios considered by McAda and Wasiolek (p.80) are not comparable to the results of Hearne's model, and therefore will not be presented here. Although constructed for a different purpose, the model prepared by McAda and Wasiolek was later used by the State Engineer Office to generate a second prediction of the impact of the BIA plan. By the time this was done, there were actually two versions of the Hearne code; although the differences are only numerical, not conceptual, the newer version shows even less stream impact than the original. Detailed quantitative comparisons are not possible based on material in the public domain, but representative results of the three models are presented in Table 3.4. The model prepared by McAda and Wasiolek shows much greater impact on streamflow, with correspondingly less drawdown; depending on which version of Hearne's model is used, the difference is as much as 5,000 acre·feet per year of streamflow and 120 feet of drawdown.

Possible reasons for the difference are not far to seek; for example, the combination of dipping layers and anisotropy in Hearne's model tends to cause wells to draw water from their layer, rather than horizontally from adjacent cells containing streams. The horizontal layers of McAda and Wasiolek, on

Table 3.4: Comparison of model projections of the BIA plan. Data courtesy of NM State Engineer Office.

Model	Hearne (Original)	Hearne (Modular)	McAda and Wasiolek
Predicted Stream Impact [cfs (A·ft/yr)] (captured from Pojoaque and tributaries)			
	2.45 (1774)	1.77 (1281)	8.70 (6299)
Drawdown at Selected Nodes [feet] (locations given by Row/Column/Layer; see Fig.3.2)			
8/16/8	165	165	45
11/7/17	100	100	16
9/12/12	220	220	118

the other hand, have relatively little resistance to lateral flow, allowing modeled wells to capture streamflow more readily. The problem remains, however, to determine which of the models is a better analogy to prototypal *behavior*. While other groundwater modelers might well be impressed that the difference between various competing model outputs (such as streamflow) is generally less than an order of magnitude, water managers have less reason to cheer. In this case, the very issue managers had hoped to resolve serves to highlight the divergence of the models in their predictive phases. Protracted consideration of the two models has proven inconclusive in assessing the impact of the proposed plan on fully appropriated surface water flows.

As of March, 1995, a decision has not been reached regarding the BIA proposal despite efforts to compare, reconcile or adjust divergent model ouputs. Although the two incommensurate models both history-match reasonably well, history-matching is not the only semi-quantitative test available to discriminate between models. An obvious difference between the two Tesuque aquifer models that lends itself to further investigation is the water budget. Simulated discharge to the Rio Grande provides a rough measure of the amount of water flowing through the two simulated systems. To maintain hydraulic conductivities in the plausible range, McAda and Wasiolek require 79% more water in their system as Hearne (39.3 cfs discharged to the Rio Grande versus about 22 cfs). Hydraulic conductivity varies by less than two orders of magnitude within the non-dipping, horizontally isotropic, 4-layer system; the relative lack of lateral resistance requires greater flow to maintain observed head levels. Thus it is not just the different modeling strategies nor the divergent predictions that quantitatively distinguishes one model from the other. Model discharges are not likely to be conclusive, however; river discharges, for example, are notoriously variable, and available records cannot rule out either model.

Model inputs, on the other hand, have continued to attract attention, focusing on the more time-averaged mountain-front recharge. It may be that it can be shown that one of the models depends upon implausible input parameters. Additional studies have attempted to fine-tune inputs to the plausible water budget, offering some hope that closer bounds on the water budget may constrain model output as well. Maryann Wasiolek is preparing a study of mountain front recharge (in press).[100] In 1994, Scott Anderholm of the USGS published a chloride mass balance study of natural recharge near Santa Fe. He concluded that recharge to the Santa Fe drainage is smaller than that assumed by Hearne (1985), McAda and Wasiolek (1988), and Wasiolek (in press).[101] Anderholm's estimate of overall mountain-front recharge conclusions was similar to those of several previous researchers, including Hearne, while recharge to the Arroyo Hondo south of Santa Fe again differed:

> Mountain-front recharge (mountain-stream-channel recharge plus subsurface inflow from the mountains) estimates for the Rio Tesuque drainage using the chloride-balance method are similar to those of Spiegel and Baldwin (1963), Lee Wilson and Associates (1978), and Hearne (1985). Estimates for the Arroyo Hondo drainage are larger than those estimated by Spiegel and Baldwin (1963) and smaller than those estimated by McAda and Wasiolek (1988).[102]

Especially in the presence of competing estimates, these studies are not likely to prove definitive in the "emotionally charged" atmosphere of the Pojoaque River Basin. Anderholm's work helps put the two aquifer models in perspective, but it says little more of use to decisionmakers than that Hearne is to the low side of the plausible recharge range, while McAda and Wasiolek are to the high side. We began with Jacob Bear's comments on the managerial implications of uncertainty, which he thinks should modify the objectives and expectations of the modeling exercise. By this point, we are capable of a more seasoned appreciation of these implications:

> Often the question is raised as to whether, in view of all these uncertainties, which always exist in any real-world problem, models should still be regarded as reliable tools for providing predictions

of real-world behavior - *there is no alternative!* However, the kinds of answers models should be expected to provide and the very objectives of employing models, should be broadened beyond the simple requirement that they provide the predicted response of the system to the planned excitations. Stochastic models provide probabilistic predictions rather than deterministic ones. Management must then make use of such predictions in the decision-making process. Methodologies for evaluating uncertainties will have to be developed; especially methods for evaluating the worth of data in reducing uncertainty. It then becomes a management decision whether or not to invest more in data acquisition.[103]

Common as groundwater modeling has become since Jay Lehr published his misgivings concerning the Emperor's wardrobe, not every groundwater system has attracted the routine attention of hydrologists. Among systems that have been examined at all, many are fortunate to have even a single sophisticated groundwater model. This is the situation with respect to the USGS model of the Albuquerque basin, for instance. The model updated by Kernodle, *et al.*, in the early 1990s does not face any specific competition on the same scale.[104] At the same time, it is commonplace to have multiple - and conflicting - models in legally contested situations. Under such circumstances, the motivation for model choices can be suspect, as are interpretations of model results; it is to be expected that difficulties related to the uncertainty of groundwater models will be exacerbated in a courtroom setting. Although Mary Anderson suggests "it is easy to lose sight of the limitations of a model" in such a setting, Bear emphasizes the need for reasoned responses from water managers contemplating a probabilistic model. Multiple models present different issues. Due to the obvious conflicts of interest, adversarial and underdetermined models might then rightly play a role somewhat lessened by the cynicism of "initiated" decisionmakers. Adversarial models routinely confuse issues more than they clarify them; some models are designed for no other purpose.

It is rare, conversely, to have two high-level models that are not adversarial in their origins; one such story has been told in this chapter. The United States Geological Survey is jealous of its nationwide role as impartial arbiter of water conflicts; the Survey therefore stays above the fray insofar as anyone can. Observers of the indecisiveness of USGS intervention in Pojoaque River Basin happenings may well mull Chaucer's lament: *If gold ruste, what shal iren do?* Modelers of the Tesuque aquifer have indeed, in Bacon's phrase, "fetched a wide circuit" and met with many matters, but made little progress:[105] Moreover, as Bacon said:

> the sciences to which we are accustomed have certain general positions, ...but as soon as they come to particulars... when they should produce fruit and works, then arise contentions and barking disputations, which are the end of the matter and all the issue they can yield; ...insomuch that many times not only what was asserted once is asserted still, but what was a question once is a question still...[106]

The models of Hearne and McAda/Wasiolek have since become exhibits in a lengthy legal proceeding in which "what was a question once is a question still" - not only hydrologically but legally.[107] Efforts by the State Engineer Office to assess the "competing" models have been inconclusive, impeding the resolution of water rights issues at both the state and federal levels. The original model of Hearne, whatever its faults, was clearly intended to assist in making exactly these decisions. The model produced by McAda and Wasiolek was aimed at assessing the long-term regional effects of mostly distant municipal wells; its design may be less appropriate in checking the effects of the BIA plan on stream flow. Differences in model behavior are often explained by the different purposes involved, but it is not clear in this case what difference that makes: McAda and Wasiolek modeled more or less the same area, and when their models "passed" the same qualitative validation tests, it presented an apparent opportunity to state officials looking for support in a highly complicated situation. Unfortunately, the second model has probably raised more dust than it has settled.

From the general comments of Chapter 2 and the particulars of Pojoaque a picture emerges of applied groundwater modeling as a sometimes dubiously non-unique, selective and interpretive process in which major model decisions are not readily susceptible to public review or regulatory control. Simplifying assumptions and approximations are made virtually every step of the modeling way, due to lack of information or in an effort to reduce the computational load.[108] Many factors might be fatal to the final predictive accuracy of the model, and simple ignorance and incompetence are not the least of the worries. Progress is being made in bounding and even reducing the uncertainty of model parameter inputs;

there are indications, however, that the uncertainty associated with the preferred conceptual model may remain an issue long after other doubts have been relieved.[109] The availability of multiple non-converging models clearly poses a problem for hydrologists interested in direct applications. Major decisions in model construction stem only indirectly or incompletely from any quantitative or objective measures, being most reliant on the modeler's experience, insight, intuition and resources. These conceptual model decisions are, therefore, the least capable of quantitative validation, while often having the major influence on model results.

To the extent that the conceptual model is a *qualitative surmise based on indeterminate evidence*, the end result of the modeling exercise is no less uncertain and provisional. The obvious gulf between prototype and model means that conceptualization and data extension decisions are often not amenable to validation in general and to bureaucratic review in particular. And thus the standards by which models might be judged are not ones easily codified or adjudicated. The closed, self-referential loop of expertise discussed in Chapter 2 (pp.43-44) means that relatively few are in a legitimate position to judge, and these few are notably hesitant to do so.[110] The investigator/modeler is often the only one with a semblance of the necessary "cultivated ignorance":

> Ignorance is the first requisite of the historian, ignorance which simplifies and clarifies, which selects and omits... The modern historian must cultivate this necessary ignorance for himself... He has the dual task of discovering the few significant facts... and of discarding the many insignificant facts.[111]

Powerful expectations of modeling performance have nevertheless been built into water management systems, perhaps encouraged by the overpromising of results in research proposals for the development of predictive models. Examples abound. Models, as at Pojoaque, are central exhibits in the water supply battles of the West. Models are also routinely used to set maximum contaminant levels (MCLs), by predicting what concentration at the permitted site will later result in acceptable groundwater contamination at some distant point of regulatory interest. Most remarkably, the statutory guarantee for radionuclide transport to radioactive waste repository boundaries has been set at 10,000 years.

There is a tendency to think these are purely policy decisions in which good modeling is largely a matter of fact-gathering and tool-wielding. Hydrologists themselves are not necessarily above this kind of thinking, as S.W. Lohman declares: "The role of the hydrologist is to gather and present the facts; the water manager determines who shall have how much water and from what source".[112] The weaknesses in this simplistic view have now been highlighted by example, following the more general discussion of Chapter 2. Along the way, procedural aspects of models have been clarified and two major modeling strategies have been illustrated. We have also observed the critical importance of intuition guided by experience in choosing "appropriate" and "plausible" conceptual models, and in assigning "reasonable" field values to data gaps in ways that are faithful to real systems - in short, in divining the "few significant facts" of the historian. Groundwater modeling resists reduction to a fact-gathering exercise or to a set procedure leading surely to defensible answers or implications. Which facts a hydrologist gathers and how he chooses to present them may very well determine who shall have how much water and from what source.

Detailed investigations at Pojoaque resulted in *apples and oranges* modeling results. Models are frequently indecisive in water resource questions, even when alternative conceptualizations and projections are not available to highlight this fact. These shortcomings could be only the result of temporary inadequacies related to insufficient time, funding, computer capacity and geophysical investigation. Then again, the problems may be more fundamental and related to the logical structure of hydrology and its argumentative means and goals. The history of the Tesuque aquifer models, for instance, appears to belie the usual default validation, in which model worthiness is inferred from reasonable calibration and history-matching. In this case, the usual methods clearly failed to eliminate a misleading model; within the precision required by the legal process, the two cannot both be right concerning the effect of the proposed BIA plan. To sort this out, we need to take a look at the question of validation and the logic of model testing from a more rigorous point of view. Before we are done, we might have to accept the typically unvarnished opinion of Vit Klemes: "Hydrologic misconceptions may be similar to mosquitoes; it may be necessary to drain the swamps rather than merely continue killing individuals".[113]

3.5 Notes:

1. Chaucer, G. (c.1390), *Canterbury Tales*, line 500.

2. Burroughs, E.R. (1912), *A Princess of Mars*, Ballantine Books (1983), p.15.

3. Bear, J., *et al.* (1992), "Fundamentals of Ground-water Modeling", in *EPA Ground Water Issue* (EPA/540/S-92/005), April 1992.

4. Lehr, J.H. (1980), "To Model or Not to Model - That is the Question", in *Ground Water*, **18**:2, pp.106-107.

5. Anderson, M.P. (1983), "Ground-Water Modeling - The Emperor has no Clothes", in *Ground Water*, **21**:6, pp.666-669.

6. Anderson, M.P. (1983), p.669.

7. Anderson, M.P. (1983), p.667.

8. Weinberg has written frequently on what he calls *trans-scientific* questions - technical questions that are properly asked in scientific terms, but cannot be answered by present-day science. See, *eg.*, Weinberg, G.M. (1975), *An Introduction to General Systems Thinking*, Wiley-Interscience.

9. Hearne, G.A. (1985a), *Mathematical Model of the Tesuque Aquifer System Near Pojoaque, New Mexico*, U.S. Geological Survey Water Supply Paper 2205. Pagination in notes to follow is from the 1985 W-S Paper.

10. Hearne, G.A. (1985a), p.1.

11. Hearne, G.A. (1985a), p.3.

12. Personal communication, Glenn Hearne, 31 Jan 1995. The case of interest is *New Mexico v. Aamodt*, 537 F.2d 1102 (1976).

13. Hearne, G.A. (1985a, p.1) notes: "Most of the data contained in this report have been collected by others and reported elsewhere." Reports on which he relies include:

    Borton, R.L. (1968), "General Geology and Hydrology of North-Central Santa Fe County, New Mexico, New Mexico State Engineer Office open-file report.

    Galusha, T. and Blick, J.C. (1971), "Stratigraphy of the Santa Fe Group, New Mexico", in *American Museum of Natural History Bulletin*, **144**.

    Kelly, V.C. (1952), Tectonics of the Rio Grande Depression of Central New Mexico, in *Guidebook to the Rio Grande Country*, New Mexico Geological Society, pp.93-105.

    Manley, K. (1978), "Structure and Stratigraphy of the Espanola Basin, Rio Grande Rift, New Mexico", United States Geological Survey Open-File Report 78-667.

    Purtymun, W.D. and Johansen, S. (1974), "General Geohydrology of the Pajarito Plateau", in *Guidebook to Ghost Ranch*, New Mexico Geological Society, 25th Field Conference, pp.347-349.

Spiegel, Z. and Baldwin, B. (1963), "Geology and Water Resources of the Santa Fe Area, New Mexico", Geological Survey Water-Supply Paper 1525.

Spiegel and Baldwin reference some very early reports. These nineteenth century investigations can be traced to three principal interests: mining (eg.: Blake, W.P. (1859), "Observations on the Mineral Resources of the Rocky Mountain Chain, Near Santa Fe, and the Probable Extent Southwards of the Rocky Mountain Gold Field", in *Boston Society of Natural History Proceedings*, **7**, pp.64-70); paleontology and general geology (eg., Cope, E.D. (1874), "Notes on the Santa Fe Marls and some of the contained Vertebrate Fossils", in *Philadelphia Academy of Natural Science Proceedings, Paleontology Bulletin*, **18**, pp.147-152); railroad development (1853), when railroad geologists first explored for a route. See also Hayden, F.V. (1873); Stevenson, J.J. (1881).

14. Hearne, G.A. (1985a), p.4.

15. Kelly, V.C. (1978), "Geology of Espanola Basin, New Mexico", New Mexico Bureau of Mines and Mineral Resources, Geologic Map 48.

16. Manley, K. (1978), p.10. These mudflows and volcanic ejecta are also discussed in Spiegel and Baldwin (1963).

17. See Spiegel, Z. and Baldwin, B. (1963), pp.38-50.

18. Galusha and Blick (1971), p.44, put the thickness at greater than 3700 feet in places; Kelly (1978) puts the thickness at over 9000 feet near the Rio Grande; Manley's work (1978) suggests a thickness of about 4000 feet.

19. Hearne, G.A. (1985a), p.4.

20. Hearne, G.A. (1985a), p.4.

21. Hearne, G.A. (1985a), p.5.

22. See Hearne, G.A. (1985a), p.5. The modeling code used is from Posson, D.R., Hearne, G.A., Tracy, J.V. and Frenzel, P.F. (1980), "Computer Program for Simulating Geohydrologic Systems in Three Dimensions", United States Geological Survey, Open-File Report 80-421. Finite-difference methods divide a system into a collection of grid blocks. A discretized form of the governing equations is then solved as a matrix equation for hydraulic head in the system. Finite-difference methods "compute a value for the head at a node which also is the average head for the cell that surrounds the node" (Anderson, M.P. and Woessner, W.W. (1992a), *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*, p.22). These calculations may be made on the basis of either block-centered or mesh-centered nodes. In either case, numerical solutions in general "yield values for only a pre-determined, finite number of points in the problem domain" (Wang, H.F. and Anderson, M.P (1982), *Introduction to Groundwater Modeling: Finite Difference and Finite Element Methods*, p.21). They do not provide solutions everywhere in the domain; as a result, there may not be exact solutions for specific points of interest, such as the actual location of wells.

23. Personal communication, Glenn Hearne, 31 Jan 1995.

24. Hearne, G.A. (1985a), p.5.

25. Hearne, G.A. (1985a), pp.4-5.

26. Chen Chia-Shyun (1993), videotaped classroom lecture at New Mexico Institute of Mining and Technology, Hydrology 535: Applied Hydrology, April 1993. In a letter to T.C. Chamberlain, Charles Slichter made a similar comment: "I am not at all surprised at the power and sanity with which the difficult problems are approached [in your recent book]... Whenever reading your contributions to Science I am always impressed with the conviction that the multiple scheme of approach of the naturalist seems to have over the narrower line of mathematical discussion. The distinction between the two methods is fundamental. In the naturalistic method every item is brought into the problem. Mathematical methods of necessity are characterized by the omission of many things in order to simplify conditions so that mathematical analysis can be brought to bear" (Ingraham, M.H. (1972), *Charles Sumner Slichter: The Golden Vector*, University of Wisconsin Press, p.66).

27. Hearne, G.A. (1985a), p.10.

28. Hearne, G.A. (1985a), p.4.

29. Hearne, G.A. (1985a), p.4: "Except for the ash beds, the Tesuque Formation was deposited as coalescing alluvial fans. As a result, individual clastic beds are probably not continuous over the basin. Miller and others (1963, p.50) report that '... few beds can be traced more than a mile or two'".

30. Hearne, G.A. (1985a), p.10.

31. Hearne, G.A. (1985b), "Simulation of an Aquifer Test on the Tesuque Pueblo Grant, New Mexico", United States Geological Survey Water-Supply Paper 2206, p.1: "An aquifer test was designed, executed and analyzed as a preliminary step to evaluating the effect of the irrigation development plan".

32. Hearne, G.A. (1985b), p.3. All pagination is from the 1985 Water-Supply Paper.

33. Hearne, G.A. (1985b), p.4.

34. Hearne, G.A. (1985b), p.9. The equation of interest can be derived assuming essentially horizontal flow within the beds. Hearne received a private written communication from C.V. Theis in 1974, stating:

$$\frac{\partial h/\partial z}{\partial h/\partial x} = \frac{(K_n - K_p)\ \cos A\ \sin A}{K_p \sin^2 A + K_n \cos^2 A}$$

where $K_p \equiv$ hydraulic conductivity parallel to the beds [L/t];
$K_n \equiv$ hydraulic conductivity normal to the beds [L/t];
$A \equiv$ the angle of the bedding dip (negative downdip) [/];
$\partial h/\partial x \equiv$ "horizontal" hydraulic gradient in the direction of dip [/];
$\partial h/\partial z \equiv$ "vertical" hydraulic gradient normal to the direction of dip [/].

Solving for the anisotropy ratio:

$$\frac{K_n}{K_p} = \frac{\sin A(\cos A\ \frac{\partial h}{\partial x} + \sin A\ \frac{\partial h}{\partial z})}{\cos A(\sin A\ \frac{\partial h}{\partial x} - \cos A\ \frac{\partial h}{\partial z})}$$

35. Hearne, G.A. (1985b), p.5.

36. Hearne, G.A. (1985b), p.5.

37. Hearne, G.A. (1985b), p.12.

38. Hearne, G.A. (1985b), p.11.

39. Hearne, G.A. (1985b), p.12.

40. Hearne, G.A. (1985b), p.14.

41. Hearne, G.A. (1985b), p.14.

42. Hearne, G.A. (1985b), p.22. The latter conclusion is explained: "The discontinuity of less permeable beds may improve the crossbed communication by providing a tortuous path around rather than through these beds".

43. Hearne, G.A. (1985b), p.14.

44. After Hearne, G.A. (1985a), Table 1, p.11.

45. The seepage tests are described in Hearne, G.A. (1985a), p.24. See pp.15-17 for the leakance coefficient discussion. In summary,

if $q = K_R(H-h)A = KIA_f$

where $q \equiv$ volumetric flow from stream to cell $[L^3/t]$;
$K_R \equiv$ the leakance coefficient $[t^{-1}]$;
$H \equiv$ hydraulic head in the stream;
$h \equiv$ calculated head in the cell after the last time step $[L]$;
$A \equiv$ surface area of the cell $[L^2]$;
$K \equiv$ hydraulic conductivity in the direction of flow $[L/t]$;
$I \equiv$ hydraulic gradient in the direction of flow $[/]$;
$A_f \equiv$ cross-sectional area normal to flow $[L^2]$; and
the dip of the beds is 7°.

then in due course:

$$K_R = \frac{K_x}{\Delta y/2} \cdot \frac{(\Delta x \tan 7^\circ \;\; \Delta y)}{\Delta x \;\Delta y}$$

or in this case: $K_R = 5 \times 10^{-10}$ per second, or 4.32×10^{-5} per day.

The Rio Grande seepage tests referred to gave results of a loss of 8cfs from a total flow of 336 cfs; a loss of 7cfs from 440 cfs; and a gain of 33cfs from a baseflow of 856 cfs. The net mean gain per river mile is then [(-8) + (-7) + (33)] / 3(17.2mi), or 0.34 cfs. Hearne notes that: "If the differences between these measurements are the additive effect of many factors acting at random, then the mean provides a more accurate estimate". The variability in total flow is ignored in this estimation method.

46. Hearne, G.A. (1985a), p.13. See also p.43.

47. Hearne, G.A. (1985a), p.3. *Cf*: Bredehoeft, J.D. and Konikow, L.F. (1993), "Ground-Water Models: Validate or Invalidate", *Ground Water*, 31:2, pp.178-9: "The terms validation and verification are misleading, at best. These terms should be abandoned by the groundwater community".

48. Hearne, G.A. (1985a), p.26.

49. Hearne, G.A. (1985a), p.20.

50. Hearne, G.A. (1985a), p.20. See the general comment on finite-difference methods in note 17, above.

51. Hearne, G.A. (1985a), p.20.

52. Hearne, G.A. (1985a), p.18.

53. Hearne, G.A. (1985a), p.18.

54. Hearne, G.A. (1985a), pp.22-25.

55. Hearne, G.A. (1985a), pp.24, 26.

56. Hearne, G.A. (1985a), p.20.

57. Hearne, G.A. (1985a), p.26.

58. Hearne, G.A. (1985a), p.28.

59. Hearne, G.A. (1985a), p.27.

60. Hearne, G.A. (1985a), p.31.

61. Hearne, G.A. (1985a), p.24.

62. Hearne, G.A. (1985a), p.52.

63. Hearne, G.A. (1985a), p.55.

64. Hearne, G.A. (1985a), pp.56-61.

65. Hearne, G.A. (1985a), pp.58, 61-66.

66. Hearne, G.A. (1985a), pp.66-68.

67. Hearne, G.A. (1985a), pp.68-71.

68. Hearne, G.A. (1985a), p.70.

69. Hearne, G.A. (1985a), p.72.

70. Hearne, G.A. (1985a), p.3.

71. Hearne, G.A. (1985a), p.52.

72. Personal communication, Doug McAda, 1 Feb 1995.

73. McAda, D.P. and Wasioliek, M. (1988), "Simulation of the Regional Geohydrology of the Tesuque Aquifer System Near Santa Fe, New Mexico", United States Geological Survey, Water-Resources Investigations Report 87-4056. See also, McAda, D.P. (1990), "Simulation of the Effects of Ground-Water Withdrawal from a Well Field Adjacent to the Rio Grande, Santa Fe County, New Mexico", United States Geological Survey, Water-Resources Investigations Report 89-4184.

74. McAda, D.P. and Wasiolek, M. (1988), p.57. Details of the projected demand scenarios are available in Santa Fe Metropolitan Water Board (1984), "Santa Fe Regional Water Supply System: Report on Baseline Data, Goals, Policy and Studies", Table 4.2.

75. "This model was constructed to simulate the regional geohydrologic system in the Santa Fe area and is not intended to simulate hydraulic heads at particular well sites. The simulated results can at best represent an average condition in the model cells; therefore simulated hydraulic heads may differ from those measured in wells" (p.17). Another standard disclaimer appears twice in the report: "Distributions of aquifer characteristics in the model do not constitute a unique solution" (pp.34, 48).

76. They add the recorded pumpage and fluxes since 1977, and add or appear to rely more heavily on the following:

    Koopman, F.C. (1975), "Estimated Ground-Water Flow, Volume of Water in Storage, and Potential Yield of Wells in the Pojoaque River Drainage Basin, Santa Fe County, New Mexico, United States Geological Survey Open-File Report 74-159.

    Mourant, W.A. (1980), "Hydrologic Maps and Data for Santa Fe County, New Mexico, New Mexico State Engineer Office basic data report.

    Purtyman, W.D. and Adams, H. (1980), "Geohydrology of Bandelier National Monument, New Mexico, Los Alamos Scientific Laboratory Informal Report LA-8461-MS.

    Reiland, L.J. and Koopman, F.C. (1975), "Estimated Availability of surface and Ground Water in the Pojoaque River Drainage Basin, Santa Fe County, New Mexico", United States Geological Survey Open-File Report 74-151.

    References to McAda and Wasiolek's geological summary are from pp.7-12 of their report.

77. McAda, D.P. and Wasiolek, M. (1988), p.12.

78. McAda, D.P. and Wasiolek, M. (1988), p.17.

79. McAda, D.P. and Wasiolek, M. (1988), p.16. See Hearne, G.A. (1985b), p.5.

80. Personal communication, 31 Jan 1995.

81. McAda, D.P. and Wasiolek, M. (1988), p.21.

82. McAda, D.P. and Wasiolek, M. (1988), p.22.

83. McAda, D.P. and Wasiolek, M. (1988), p.22. The conductance method is included in McDonald, M.G. and Harbaugh, A.W. (1984), "A Modular 3-D Finite-Difference Ground-Water Flow Model", United States Geological Survey Open-File Report 83-875.

84. McAda, D.P. and Wasiolek, M. (1988), p.23.

85. McAda. D.P. and Wasiolek, M. (1988), pp.23-26.

86. McAda, D.P. and Wasiolek, M. (1988), pp.25,27.

87. McAda. D.P. and Wasiolek, M. (1988), pp.27. See also p.48.

88. McAda, D.P. and Wasiolek, M. (1988), p.27.

89. McAda, D.P. and Wasiolek, M. (1988), p.30.

90. McAda, D.P. and Wasiolek, M. (1988), p.30.

91. McAda, D.P. and Wasiolek, M. (1988), pp.31-32.

92. McAda, D.P. and Wasiolek, M. (1988), p.34.

93. McAda, D.P. and Wasiolek, M. (1988), p.35.

94. McAda, D.P. and Wasiolek, M. (1988), p.35.

95. There is some disagreement over the meaning of a reference used by both models. Spiegel and Baldwin estimate a gain of 25 cfs for the Rio Grande between the Otowi and Cochiti gaging stations. The exclusion by Spiegel and Baldwin of low flow days in their analysis inclines Hearne to think they overestimated the actual typical gain. Hearne further takes 25 cfs to mean the gain over a direct distance of 20 miles, and in combination with other seepage studies concludes that the gain "averages about 0.5 cfs per mile or less and is certainly no more than about 1 cfs per mile". This is consistent with his simulated total river gain of 22.06 cfs. McAda and Wasiolek, on the other hand, treat Spiegel and Baldwin's estimate as occurring over 26 river miles, and conclude that this helps support their net simulated Rio Grande gain of 37.4 cfs. They then utilize several other studies to amend their simulated discharge to the Rio Grande to account for evapotranspiration; these calculations reduce the estimated net discharge to the river from 37.4 cfs to 30.3 cfs. Still other studies of particular reaches are used to reinforce their simulated gains in various model blocks. McAda and Wasiolek do not, however, incorporate any of these discharge adjustments into their model: "Evaporation from the river channel and transpiration by native vegetation were not simulated by the model". Their steady-state simulation water budget shows 39.3 cfs discharged to the river, less 1.9 cfs recharged to the aquifer. Their focus is on events closer to Santa Fe; reasoning that most of the ignored losses would happen within the Rio Grande channel or nearby, they evidently judged these effects to be of negligible importance to their study.

96. Spiegel and Baldwin estimated 30-50% return flow from irrigation, and claimed it was "probably near 100%... along the Santa Fe River below Agua Fria"; there is, however, nothibng in Spiegel and Baldwin's report to indicate that this infiltrated water reaches the production zones of wells, only that "at maximum flow it is completely absorbed within 1½ miles of the point of entry". McAda and Wasiolek decide that of every 1.5 A.ft applied, 0.5 to 0.75 recharges the aquifer.

97. McAda, D.P. and Wasiolek, M. (1988), p.48.

98. McAda, D.P. and Wasiolek, M. (1988), p.34.

99. McAda, D.P. and Wasiolek, M. (1988), p.37. Sensitivity was especially evidenced in changes in discharge to streams.

100. Wasiolek, M. (in press), "Subsurface recharge to the Western Side of the Sangre de Cristo Mountains Near Santa Fe, New Mexico", United States Geological Survey, Water-Resources Investigations Report 94-4072. Peter Frenzel of the USGS has also adapted the original model of McAda and Wasiolek - subdividing each of the four layers into three layers and re-calibrating the aquifer characteristics - to look specifically at the well fields around Los Alamos.

101. Anderholm, S.K. (1994), "Ground-Water Recharge Near Santa Fe, North-Central New Mexico", United States Geological Survey, Water-Resources Investigations Report 94-4078,The larger numbers in Wasiolek's forthcoming report stem in part from 1) considering the basin larger than others do; 2) the average annual rainfall to be slightly higher; and 3) finding a higher estimated chloride concentration in the mountain front drainage. Anderholm, S.K. (1994), p.66, summarizes the different findings on mountain-front recharge:

Table 3.5: Findings on mountain-front recharge north of Santa Fe. (1) ≡ Anderholm (1994); (2) ≡ Spiegel and Baldwin (1963); (3) ≡ Wasiolek (in press); (4) ≡ Reiland (1975). From Anderholm (1994), p.66.

Drainage basin (square miles)	Average annual precipitation in basin (inches)	Pm Average annual volume of precipitation intercepting basin (acre-feet)	PmCp Average annual mass of chloride intercepting basin (acre-feet milligrams per liter)	Rmf Estimated annual mountain-front recharge (acre-feet)	Rmf Estimated annual mountain-front recharge assuming runoff (acre-feet)	Measured or estimated annual streamflow (acre-feet)	Estimated chloride concentration in drainage at mountain front (milligrams per liter)
Santa Fe River (downstream point of drainage basin; latitude 35°41'12", longitude 105°38'40")							
26.95(1)	20.91(1)	30,060	8,700	2,900	2,320	5,800 (2)	1.5
18.2 (2)	26.0 (2)	25,000					
8.8 (2)	18.0 (2)	8,400					
		33,000	9,600	3,200	2,620	5,800 (2)	1.7
28.70(3)	23.62(3)	36,200	10,500	3,500	2,920	5,800 (2)	1.8
Little Tesuque Creek (downstream point of drainage basin; latitude 35°43'30", longitude 105°53'15")							
7.34(1)	21.90(1)	8,573	2,500	630		400 (4)	6.3
7.2 (2)	21.0 (2)	8,100	2,300	580		400 (4)	5.8
7.66(3)	22.96(3)	9,380	2,700	670		400 (4)	6.8
Tesuque Creek (downstream point of drainage basin; latitude 35°44'20", longitude 105°54'20")							
11.84(1)	20.77(1)	13,120	3,800	950		2,300 (4)	1.7
11.2 (2)	21.0 (2)	12,500	3,600	900		2,300 (4)	1.6
11.22(3)	24.18(3)	14,470	4,200	1,050		2,300 (4)	1.8
Rio Tesuque (sum of Little Tesuque Creek and Tesuque Creek)							
				1,580(1)	690		
				1,480(2)	590		
				1,720(3)	830		
Arroyo Hondo (downstream point of drainage basin; latitude 35°37'03", longitude 105°54'17")							
8.38(1)	19.15(1)	8,560	2,500	830		535 (2)	4.7
6.7 (2)	17.0 (2)	6,100	1,800	590		535 (2)	3.4

102. Anderholm, S.K. (1994), p.37.

103. Bear, J., *et al.* (1992), "Fundamentals of Ground-water Modeling", in *EPA Ground Water Issue* (EPA/540/S-92/005), April 1992.

104. Kernodle, J.M., McAda, D.P. and Thorn, C.R. (1995), "Simulation of Ground-Water Flow in the Albuquerque Basin, Central New Mexico, 1901-1994, With Projections to 2020", United States Geological Survey, Water-Resources Investigations Report 94-4251. In contrast to the

techniques employed by Hearne and McAda/Wasiolek in the Pojoaque River basin, it is interesting that the modelers regard poor history-matching as a guide to needed field investigations, not as a motivation for *ad hoc* parameter adjustment (see pp.104-105, *eg*.). The significance of this technique will be developed in Chapters 4 and 7.

105. Bacon, F. (1620b), *lxx*, p.48. See also *c*, p.69-70: "For experience, when it wanders in its own track, is, as I have already remarked, mere groping in the dark, and confounds men rather than instructs them. But when it shall proceed in accordance with a fixed law, in regular order, then may better things be hoped of knowledge"; see also *cxxv*, p.83.

106. Bacon, F. (1620a), preface, p.7.

107. The barest outline of the larger legal context within which the models have been evaluated can be provided here. A good source for the general background up to the early 1980s is Dumars, C.T., *et al.* (1984) *Pueblo Indian Water Rights: Struggle for a Precious Resource, University of Arizona Press*. A Pueblo Indian water right may rest on one of three contested bases. First, the right may be what Dumars, *et al.*, call *aboriginal*, meaning that the Indians have historical claims to water long pre-dating non-Indian users. Secondly, the right may be construed to rely on *treaty* provisions, in which case it is intimately related to and dates from specific tribal/non-tribal or international negotiations. Lastly, the water right may be the result of direct *Congressional* action, in which unilateral acts of the United States Congress conferred certain rights.

The aboriginal right may appear to be the simplest legal theory, since if the Indian right is aboriginal, then it is absolute, and "the priority date of the Indians is from time immemorial". Complications quickly enter with attempts to quantify such a claim, however. If an Indian water right is taken to be aboriginal, two secondary legal theories must still be adjudicated to assess the quantity of that right: 1) the water right is limited to historical uses; 2) the right is an expanding one, contingent on later events. The major case promoting the latter view is *Winters v. U.S.*, 207 US 564 (1908). The court there took a more or less aboriginal view of the water right of the Gros Venture tribe. Since the purpose of the federal government in assigning the tribe to the Fort Belknap Reservation (Montana) was to make them "a pastoral and civilized people", the reduction in their land required an expanded water use (to switch to farming in arid lands).

Pueblo water rights also arguably stem from specific treaties. These exist between Spain and Mexico, Mexico and the United States, or between the United States and a particular tribe. The legal arguments quickly become very complicated. In any case, if the Pueblo water right derives from government treaty or other action, or from putting the water to beneficial use, then it is no longer absolute but conditional. All the same, even under this reading of the law, "... in a prior appropriation state, it plainly does not hurt one's legal position to be the 'first' people" (p.16). *Winters* can also be read in places as attributing Indian water rights to treaty. 27-8: the strongest argument for a treaty right "must be sustained or must fall based on the Treaty of Guadalupe Hidalgo [1848, in which New Mexico was ceded to the U.S. by Mexico]... The Pueblos must rely on their status as specially protected citizens under Spanish or Mexican law to make the treaty water right argument", since the treaty was betw US and Mexico and makes no specific mention of Indians. In a mix of aboriginal and treaty arguments, some claim that the Treaty did not impair aboriginal rights and in fact obligates the US to honor such rights. Dumars, *et al.*, summarize the treaty argument:

> The linchpin of the treaty argument will be the extent of water rights conferred upon the Pueblos by Spanish and Mexican law. Unlike the aboriginal water rights argument - that the Pueblos' rights survived Spanish sovereignty - the treaty argument is that the water rights were created and defined by the laws of the prior sovereigns and that the US is bound to honor those laws. Conversely, if there were no rights created by Spain or

Mexico, none were preserved by the treaty (pp.28-29).

The final possible source of a Pueblo water right is acts of Congress. Priority dates here range from the Indian Trade and Intercourse Act of 1851 to the establishment of the Middle Rio Grande Conservancy District in 1928. The latter law establishing the Conservancy District gave "paramount water rights" to the six tribes living within the district "that could not be lost by nonuse or abandonment" (p.43). The 1851 law has been argued to recognize Indian rights prior to the Treaty of Guadalupe Hidalgo, which in effect asserts absolute but limited aboriginal rights. The Treaty specifically limits Pueblo water use to the maximum they had used under Mexican sovereignty. Another claim based on this law asserts a priority date the same as the law's enactment, treating the law as the equivalent of establishing a reservation. Such a claim would, if upheld, likely lead to a right that is conditional but limited only by the *Practicably Irrigable Acreage (PIA)* standard of *Arizona v. California* 373 US 546 (1963). Under this standard, a reservation water right is quantified by present *and future* water required to irrigate all land "susceptible to irrigation". There is no basis for PIA under Mexican law (Dumars, pp.28-29). Counterarguments note that if the 1851 law was the equivalent of establishing a reservation, then Congress would not have bothered with the granting of Pueblo patents in 1858.

The relevance of other Indian water rights cases to the Pueblos is vastly complicated by several facts. First, the Pueblos were established neither by assignment to reservation nor by treaty; the Pueblos existed for centuries before successive rule under Spanish, Mexican and United States law. Second, the Pueblos never exercised "complete dominion" over their waters, unlike the Gros Venture (for example) whose land bordered the non-navigable Milk River. Third, Indians enjoy an at least quasi-sovereign status for certain purposes. Fourth, specific Congressional action on the Pueblos (in the form of patents) came only after significant non-Indian water use. Fifth, there are precedents in Spanish and Mexican law for both the principal or prior appropriation and for some relaxation of this right in dry periods (Dumars, pp.29-31).

As a result, priority date is not so simple as the date a reservation was established or a treaty signed between non-Indian sovereigns; and compensation is not so easily set as when historical use and control is compromised. The limited sovereignty of the tribes is another pivotal issue: "The Pueblos will forcefully argue that control over water was just such an inherent [sovereign] power" (p.16). In terms of priority, the patent date of the Pueblos is sometimes advanced as equivalent to the date a reservation was established. In the *Winters* case, the reservation was established 1888, and the contested non-Indian diversions occurred in 1898. But in New Mexico, "Congress took no action vis-a-vis the Pueblos until after 1851, long after the water was diverted by non-Indians" (Dumars, p.16). Furthermore, even if the Treaty of Guadalupe Hidalgo preserved Indian rights granted under Spanish or Mexican rule, no one imagines the Treaty means that either Spanish or Mexican law still prevails throughout the ceded territory. Current New Mexico water law does share some of the attributes of Spanish law, including the doctrine of prior appropriation. Spanish adjudication records also show precedents for senior users to yield a portion to others in hard times. Under Mexican water law, legal right was only one consideration in distribution of water; "the expanded needs of a user were considered in the context of the needs of others" (Dumars, pp.29-31). Reliance on the law of previous sovereigns precludes a rigid quantification standard that ignores the needs of other users (Dumars, p.41). Not only is there a significant difference in the nature and quantification of a Pueblo water right under the different legal readings, but the status of non-Indian users during this period is also affected. It has sometimes been claimed that the BIA proposal of the early 1970s was less a serious farming plan than an attempt to quantify the PIA of the Pueblos. An irrigation plan is an essential component within a water right claim that relies in any way on the concept of practicably irrigable acreage. Clearing their right to the water would open many options to the successful Pueblos, only one of which is to implement the BIA proposal.

108. Tsang, C. (1991), "The Modeling Process and Model Validation", in *Ground Water*, **29**:6, p.827: "These uncertainties may arise not only from data uncertainties, but also from every step of the modeling process..."

109. Bredehoeft, J.D. and Konikow, L.F. (1993), p.178: "However, these confidence limits do not bound errors arising from the selection of a wrong conceptual model..." Anderson and Woessner (1992b, p.167) point out: "model validation takes place during the final steps in the modeling process when the accuracy of the model is tested. As such, the success of model validation depends on satisfactory completion of all the other steps in the modeling process".

110. Yeats said: "The best lack all conviction, while the worst are full of passionate intensity".

111. Carr, E.H. (1961), pp.13-4. The repeated references to this well-known collection of essays are intended to reinforce the parallels between modeling in the physical sciences and what are more easily accepted as preliminary and speculative efforts in the humanities and social sciences.

112. Lohman, S.W. (1979), United States Geological Survey Professional Paper 708, in *Ground-Water Hydraulics*, p.62.

113. Klemes, V. (1986), "Dilettantism in Hydrology: Transition or Destiny?", in *Water Resources Research*, **22**:9, p.177s.

4

Remodeling: Logical Issues in Applied Modeling

> No one has claimed that the predictability of seasonal reproductive cycles (as expressed by Tennyson in *Locksley Hall*: "In the spring a young man's fancy lightly turns to thoughts of love") makes biology a predictive science.
>
> - Stephen G. Brush [1]

4.0 Introduction

Our laundry list of motivations for a study of hydrologic methodology included the usefulness of distinguishing temporary from more resistant inadequacies in hydrologic science and policy. As a practical matter, standard procedures have not proven decisive in the evaluation of the USGS models at Pojoaque, a fact that may be taken to reinforce the general reservations of many modelers that were surveyed in Chapter 2. Taking their cue from regulatory activity, most practitioners nevertheless see the primary goal of the modeling exercise as a prediction, often with legal implications. Hassanizadeh and Carrera, for example, declare: "A model is useful only if it can be used to make predictions for new conditions and future situations".[2] Bear reminds us (p.53) that is often no practical alternative to an attempt at predictive modeling. The apparently reasonable expectation that model predictions guiding important decisions should conform to reality remains embedded in both agency regulations and common sense. The general idea of validation is not easily jettisoned.

The critical issue at this point is whether problems in the validation of groundwater theories are only technological and therefore largely temporary, or more resistant due to strictly logical impasses.

Hydrologists have occasionally ventured a quasi-philosophical explanation of their views on validation; these discussions have rarely been optimistic. For example, despite the fact that some regulatory language clearly links validation and *a priori* proof of predictive accuracy, modelers Konikow and Bredehoeft have written on occasion that validation is simply impossible, and have leaned on the work of Karl Popper to buttress their claim.[3] Popper is of course only one of many approaches to the question of validation, but since Konikow and Bredehoeft cite him as an authority we will begin with an overview of his philosophy as it might pertain to a taxonomy of hydrologic models. As a minimum, we should gain a useful foothold in the logic of validation.

If validation is permanently out of reach, we then want to ask: If so, so what? Many implications follow. It becomes reasonable to wonder whether hydrology is a legitimately predictive science - common sense should question the utility of predictive modeling if there is no convincing method to demonstrate a reasonable equivalence between the behaviors of model and prototype. Many managerial repercussions are then unavoidable. Moreover, accepting categorically the idea that it is impossible to validate hydrologic models - or, more generally, scientific theories - also appears to depict progress in the past as a series of anomalies, and progress in the future as a myth. We don't believe either of these things are so. The later chapters of this study will attempt to append an account of progress to the apparent uncertainty of hydrologic science. We will also see in due course that denying any basis for validation leads to the initially puzzling idea that applied groundwater models might be best treated as process-oriented research tools. The link to crucial tests and demonstrations - and to Bacon's *Experiments of Light* - will then become apparent.

For our immediate purposes, we can gain valuable perspective by carefully describing the methodological pairs of validation versus falsification, and crucial demonstrations versus crucial tests. As we will see, Popper rejects validation on logical grounds and highlights the idea of falsifying faulty theories through crucial tests (rather than validating good ones through crucial demonstrations). The appeal to testable predictions therefore plays a pivotal role in his philosophy of science. The details (and criticisms) of this argument can help clarify the logical status of prediction in hydrology as we explore the methodological implications of various positions in the validation debate. A brief analysis of Popper's philosophy of science can provide some orienting context and useful terminology.

4.1 Karl Popper: Sticking Hume's Fork[4] Into Bacon

Karl Popper is one of the few philosophers of science enjoying much name recognition among scientists (Thomas Kuhn being another),[5] but the analysis of Konikow and Bredehoeft betrays no intimacy with the subtleties of either Kuhn's or Popper's program.[6] The hydrologists condense the relevant part of Popper's philosophy to the dictum that "as scientists we can never validate a hypothesis, only invalidate it", and therefore claim that

> site-specific ground-water models are elements of applied earth-science - in effect, an agglomeration of multiple hydrogeologic theories. As such, they are subject to improvement via invalidation, but cannot be proven valid. Validation cannot add to the fund of knowledge.[7]

Konikow and Bredehoeft also like the paraphrase of Popper offerred by physicist Stephen Hawking: "A good theory... makes a number of predictions that could in principle be disproved or falsified by observation".[8] Although these are accurate enough paraphrases of Popper's conclusions, the question of scientific proof can be more effectively approached by beginning at the beginning. Accordingly, we first quickly survey Popper's views on discovery, observation, induction, demarcation between science and pseudo-science, and only then justification.

The source of scientific statements is an intriguing puzzle in the philosophy of science. Popper is at pains to distinguish discovery from justification. He doesn't care where ideas come from - he only insists that the scientific method begins when someone suggests an idea that is formulated in a testable form. Events prior to such a formulation are explicitly outside the bounds of his philosophy. *The Logic of Scientific Discovery* [*Logik der Forschung*], one of the most argued books in the philosophy of science, may also take the prize for the most misleading title, since Popper deals in only the most cursory way with the source of scientific theories (discovery); he emphasizes instead how one might recognize proper theories and, to a degree, test them (demarcation and falsification):

> The initial stage, the act of conceiving or inventing a theory, seems to me to neither call for logical analysis nor to be susceptible of it. The question of how it happens that a new idea occurs to a man - whether it is a musical theme, a dramatic conflict, or a scientific theory - may be of great interest to empirical psychology; but it is irrelevant to the logical analysis of scientific knowledge... Accordingly I shall distinguish sharply between the process of conceiving a new idea, and the methods and results of examining it logically.[9]

The source of these ideas or theories is left completely open by Popper:

> I am inclined to think that scientific discovery is impossible without faith in ideas which are of a purely speculative kind, and sometimes even quite hazy; a faith which is completely unwarranted from the point of view of science, and which, to that extent, is 'metaphysical'. Yet, having issued all these warnings, I still take it to be the first task of the logic of knowledge to put forward a *concept of empirical science* in order to... draw a clear line of demarcation between science and metaphysical ideas - even though these ideas may have furthered the advance of science throughout its history.[10]

Popper rejects simple observation in terms reminiscent of Bacon. Bacon, we recall, shuns aimless observation as "stupid and blind", and argues for research guided by whatever understanding is available and motivated by an explicit quest for organizing principles. With Bacon, Popper notes that the amorphous cacophany of simple observations cannot be resolved into coherent, not to say testable, statements:

> Thus the real situation is quite different from the one visualized by the naive empiricist, or the believer in inductive logic. He thinks that we begin by collecting and arranging our experiences, and so ascend the ladder of science... But if I am ordered: 'Record what you are now experiencing' I shall hardly know how to obey this ambiguous order. Am I to report that I am writing; that I hear a bell ringing; a newsboy shouting; a loudspeaker droning; or am I to report, perhaps, that these noises irritate me? And even if the order could be obeyed: however rich a collection of statements might be assembled in this way, it could never add up to a *science*. A science needs points of view, and theoretical problems.[11]

Bacon and Popper thus quickly part company, since the former was not inclined to recognize theoretical problems, and certainly did not let them frame his point of view. Moreover, Bacon's apparently confident realism is unacceptable to Popper: "Anyone who envisages a system of absolutely certain, irrevocably true statements as the end and purpose of science will certainly reject the proposals I shall make here".[12]

They disagree again over the role of induction from experience. Bacon's cumulative theory of knowledge expects a continuous induction to lead to ever more general theories and laws of nature; thus induction is for Bacon an implicit method of proof. Popper, on the other hand, views discovery as a matter for "empirical psychology", not for the logic of science. The source of ideas is irrelevant to his means of judging their soundness, and he inveighs against induction as a method of proof at every turn. While metaphysical and poorly defined influences (including loose inductions from experience) might be essential in the development of theories, Popper does not consider this part of the scientific method, and hence, accumulating experience does not constitute proof of any sort. He disagrees with both Bacon and the more traditional Positivists (such as Carnap[13]) over the role inductive logic might play in building up true theories. Popper believes the scientific method is synonymous with the evaluation of theories and characterized by a complete dependence on deductive logic.[14]

As Konikow and Bredehoeft point out, he goes further and claims there is no basis whatsoever on which empirical theories might be validated.[15] Any phenomenon can be explained in more than one way; his recognition that theories do not converge to a final explanation makes Popper an anti-realist with respect to theories, since ruling out validation means any theory is *always* subject to replacement:

> Now in my view there is no such thing as induction. Thus inference to theories, from singular statements which are 'verified by experience' (whatever that may mean), is logically inadmissable. Theories are, therefore, *never* empirically verifiable.[16]

Since Popper's methodology defines the ultimate sources of theories as foreign material, his prescription begins, not with discovery, but with a logic of theory evaluation. Theory evaluation is in turn closely linked to his means of demarcation, *i.e.*, the procedure or criteria for separating scientific from unscientific claims. In his autobiography he describes his experiences as a young student in Vienna in the 1920s; these included encounters with Marxism, psychology and a lecture by Albert Einstein. He was interested in specifying just how Einstein differed from what Popper was already convinced were merely pseudo-scientific types like Freud, Adler and Marx; he also considered evolutionary Darwinism unscientific for a time.[17] Referring to the famous light-bending experiments during the solar eclipse of 1919, Popper quotes admiringly Einstein's statement that "if the redshift of spectral lines due to the gravitational potential should not exist, then the general theory of relativity will be untenable".[18] Marxists and psychologists, on the other hand, could always fall back on excuses like third world imperialism or psychological repression to explain away the predictive failures of their theories. Thus "the Marxists... and the psychoanalysts of all schools were able to interpret any conceivable event as a verification of their theories".[19] The distinction between these two research attitudes led Popper to his famous "falsifiability criterion" for the demarcation of scientific statements from non-scientific ones. The scientific ones are scientific because they can be presented along with a clearly stipulated procedure that explains what it would take to show that they are false. They are, in a word, falsifiable. Non-scientific claims are not attended by the same sort of at least theoretically decisive falsification tests.[20]

Having rejected the inductive logic of the more traditional Positivists, such as Carnap, Popper is left without a means of justification. He therefore develops a deductive logic of falsification, in which deductions from the theory of interest yield testable and potentially falsifiable statements. The central importance of demarcation is easily seen, inasmuch as only scientific statements can generate such potentially falsifiable expectations. The devil, as usual, is in the details. As it happened, measurements taken by the highly admired experimentalist Walter Kaufmann in 1905-6 appeared to refute the new theory, but Einstein was unmoved by the results. He ignored the widely discussed experimental efforts for a few years, and then shrugged off the results without even pointing out the obvious, namely, that there may have been some experimental sloppiness that distorted the measurements. Einstein believed instead that his theory was one of first principles; as such, it neither required experimental evidence, nor was it susceptible to experimental falsification. It was not until 1915 that a consensus began to emerge that Kaufmann's original experiments had been flawed by an insufficiently high vacuum, producing what until then had certainly appeared to be falsifying observations.[21] And still Popper claims that:

> Einstein was looking for crucial experiments whose agreement with his predictions would by no

> means establish his theory; while a disagreement, as he was the first to stress, would show his theory to be untenable. This, I felt, was the true scientific attitude. It was utterly different from the dogmatic attitude which constantly claimed to find "verifications" for its favourite theories. Thus I arrived, by the end of 1919, at the conclusion that the scientific attitude was the critical attitude, which did not look for verifications but for crucial tests; tests which could *refute* the theory tested, though they could never establish it.[22]

The idea is that experimental results inconsistent with a proposed theory should eliminate - falsify - the evidently faulty theory. On the other hand, positive results will corroborate or confirm a theory to a degree warranted by the severity of the test. Popper warns, however, that results in accord with predictions should not increase one's confidence in the theory to that same degree: "corroborated" should not be taken to mean "probable"; these "not invalidated" theories merely live on to face new challenges, as clarified in the following grumbling about Carnap :

> I introduced the terms "corroboration" and especially "degree of corroboration" in my book because I wanted a *neutral* term to describe the degree to which a hypothesis has stood up to severe tests, and thus "proved its mettle". By "neutral" I mean a term not prejudging the issue whether, by standing up to severe tests, the hypothesis becomes "more probable"...
>
> Carnap translated my term "degree of corroboration"... as "degree of confirmation"... I did not like this term, because of its associations ("make firm", "establish firmly", "put beyond doubt"; "prove"; "verify"...)
>
> ..."Degree of confirmation" was soon used - by Carnap himself - as a synonym of "probability". I have therefore abandoned it in favour of "degree of corroboration".[23]

If falsification is achievable, unlike validation, the proper scientific method should emphasize attempts to disprove scientific explanations. Thus progress is said to consist in the weeding out of initially plausible but now demonstrably flawed theories. However, the history of science is replete with events - on all sorts of scales - that mimic Einstein's reaction to seemingly falsifying data.[24] Moreover, apparently falsified theories are sometimes later resurrected with impressive results. It thus appears falsification can sometimes be a judgment as elusive and provisional as validation. Accounting for this state of affairs requires a closer look at the programmatic heart of Popper's writings.

4.2 Falsifiable Predictions and Crucial Experiments

"Never trust a survivor", my father used to warn me, with Vartan Mamigonian in mind, "until you find out what he did to stay alive".

- Kurt Vonnegut, *Bluebeard*[25]

Popper elaborates on his demarcation criterion; his objection is really to the idea that "there are statements in science which we have, resignedly, to accept as true merely because it does not seem possible, for logical reasons, to test them". Popper does not, however, "demand that every scientific statement must *have in fact been tested* before it is accepted. I only demand that every such statement must be *capable* of being tested". And thus: "*It must be possible for an empirical scientific system to be refuted by experience*".[26]

But Popper's train of thought is soon stopping at another station; and when hydrologists imply that falsification is a fairly straightforward procedure (at least relative to validation),[27] they not only misrepresent experience; they also misread Popper [italics added]:

> In point of fact, no conclusive disproof of a theory can ever be produced; for it is always possible to say that the experimental results are not reliable, or that the discrepancies which are asserted to exist between the experimental results and the theory are only apparent and that they will disappear with the advance of our understanding...*If you insist on strict proof (or strict disproof) in the empirical sciences, you will never benefit from experience, and never learn from it how wrong you are.*[28]

Popper is acknowledging the common practice of blaming or introducing auxiliary hypotheses to salvage the main hypothesis by "immunizing" it to negative test results.[29] The basic model is thus preserved and not open to falsification. Some philosophers have claimed that the core of a theory is *rarely or never tested*, with the brunt of the experimental evidence falling always only on secondary conditions or assumptions that can be readily rearranged.[30] This is a bit of over-reaction, but it is true that Popper is soon reduced to talk of "degrees of testability" and "degrees of falsifiability",[31] and embarks on an extended discussion of probability in science.[32]

In this light, major complications emerge for Popper's main themes of demarcation and falsification. If falsification can sometimes be evaded, people who are obviously scientists will sometimes be hard to distinguish from their pseudo-brethren, since Popper's only distinction between science and pseudo-science is that statements in the former domain can be - at least in principle - tested. New sticking points occur as Popper implies theories may be "accepted" without being tested (p.116), while expecting that the progress of science requires "faith in ideas of a purely speculative kind" (p.112). The sociologists of science Mulkay and Gilbert point out that these practical difficulties (whether categorical or occasional) threaten to reduce his pristine logical system to the usually unarticulated and seemingly *ad hoc* preference system in common use.[33] Degrees of invalidation are not epistemologically distinct from degrees of corroboration:

> Both Popperians and non-Popperians have to weigh positive as well as negative evidence in order to come to some 'reasonable' conclusion; and neither has any definitive criterion with which to establish the adequacy of their conclusion.[34]

Popper's familiarity and sympathy with science are apparent throughout his books. Historical inaccuracies, however - like his version of Einstein's method - highlight the fact that he, in common with the more traditional Positivists, is essentially *prescriptive* in his philosophy, not *descriptive*,[35] as Kuhn and the historical school at least attempt to be. Kuhn, for example, recognizes the role of experimental uncertainty in the evaluation of theories: "The operations and measurements that a scientist undertakes in the laboratory are not 'the given' of experience but rather 'the collected with difficulty'".[36] Popper, on the other hand, declares that "it never happens that old experiments one day yield new results",[37] a statement that is only strictly true if we assume the experiment was previously done "correctly"; *ie.*: with due attention to all relevant prerequisites, variables and results.[38]

All philosophy of science is discussed against the background of the obvious general success of science; as noted earlier, Popper also views science as characteristically progressive. Hilary Putnam has pointed out, however: "It is a remarkable fact about Popper's book, *The Logic of Scientific Discovery*, that it contains but a half-dozen references to the *application* of scientific theories and laws".[39] Here is further warning that applied hydrologists should be careful. Putnam is unhappy with the logical blandness of Popper's tests; he argues that a sort of inductive intuition must play a role in the logic of the scientific method; he claims Popper cannot account for progress if he is content only to say that some theories have not yet been invalidated. For example, Popper does seem to obscure rather than illuminate how scientists gain confidence in theories; when he rules out induction but says that theories can be accepted without being tested, on what basis might this be accomplished?[40]

Having accused Popper of excessive abstraction from scientific practice, Putnam naturally wants to restore some balance. His comments speak directly to the critical role played by professional judgment in the construction and "validation" of hydrologic models; they are worth noting at length:

> When a scientist accepts a law, he is recommending to other men that they rely on it - rely on it, often, in practical contexts. Only by wrenching science altogether out of the context in which it really arises - the context of men trying to change and control the world - can Popper even put forward his peculiar view on induction. Ideas are not *just* ideas; they are guides to action. Our notions of "knowledge", probability", "certainty", etc., are all linked to and frequently used in contexts in which action is at issue: may I confidently rely upon a certain idea? Shall I rely upon it tentatively, with a certain caution? Is it necessary to check on it?
>
> If "this law is highly corroborated", "this law is scientifically accepted", and like locutions merely meant "this law has withstood severe tests" - and there were no suggestion at all that a law which has withstood severe tests is likely to withstand further tests, such as the tests involved in an application or attempted application, then Popper would be right; but then science would be an entirely unimportant activity. It would be practically unimportant, because scientists would never tell us that any law or theory is safe to rely on for practical purposes; and it would be unimportant for the purpose of understanding, since in Popper's view, scientists never tell us that any law or theory is true or even probable. Knowing that certain "conjectures" (according to Popper all scientific laws are "provisional conjectures") have not yet been refuted is *not understanding anything*.[41]

In summary, there are several objections to Popper's version of scientific logic that hydrologists will want to remember. Among these, the inevitable blurring of the falsifiability test (and therefore the line of demarcation) stand out, notwithstanding Popper's attempts to answer his critics.[42] His philosophy also appears to gloss several practical problems; as condensed in the casual reports of scientists, these oversights often become complete. Or, as Mulkay and Gilbert would have it, despite Popper's claim to be providing a practical guide for working scientists, his method breaks down at the very point where guidance is needed.[43] Not only can theories be immunized against falsification and at least temporarily salvaged by adjusting secondary conditions, etc., events also show that the role of novel predictions is not so decisive within the scientific community as Popper and his admirers seem to take it to be.[44] And lastly, there is some difficulty in accounting for the obvious progress of science, if Popper's account is taken as descriptive of the actual conditions under which scientists work (as he occasionally claims). It is also not clear how adopting Popper's prescriptions would accelerate the rate of progress. If validation is impossible - and more, if judgments as to the probability and workability of "not invalidated" theories are discouraged - then the successes of science become an unexplainable series of anomalies.[45]

Here we are interested to puzzle out the difficulties faced by hydrologists in pursuit of their goals. Accounts of how science works persist in part because science in fact does work. Logical accounts, such as that of Popper, may be weak in places, but the described process itself obviously is not altogether so. And thus we are directly interested in concisely describing the conditions that lead to the occasional successes. Of non-verifiability,[46] we really want to ask: If so, so what? What implications are there here for a predictive exercise like directly applied groundwater modeling? Before we consider ramifications internal to the science, we should pause to reconsider a few external factors.

4.3 An Aside on the Motivations for This Study

The initial motivation for this study of methodology dwelled on the extra-scientific pressures under which hydrologists must often operate. In this light, there are good reasons 1) to avoid what Henry Bauer cautioned was an unfortunate "reliance on Karl Popper for defining what science ought to be";[47] and 2) to be careful that assertions about the non-verifiability of hydrologic models are not delivered without appending either an account of progress or a strategy for dealing with no progress. Certainty did not suddenly come into short supply. Did absolute certainty ever play a role in science? What role did a willingness to make mistakes, to make affirmative inferences, in short to take calculated risks play in scientific discovery and engineering progress? Can there be scientific progress under conditions of demanding virtual certainty? Discrediting the idea of validation may change but not eliminate the practical expectations under which hydrologists labor.

A recent paper by Oreskes, *et al.*, is a good example of an effort to debunk overly sanguine claims for modeling by pointing out the uncertainty of the results; it fails, however, to give any idea how progress either has been made or can be made in the future. They rely on dictionary definitions to assert that: "To say that a model is verified is to say that its truth has been demonstrated, which implies its reliability as a basis for decision-making".[48] They then claim on logical grounds that models can never be verified, or, somewhat more weakly in their terminology, validated.[49] A reasonable negative inference that might then be drawn is that since models can't be verified, their truth cannot be demonstrated, and they cannot provide any basis for decision-making. This is not plainly stated, although the inference is available to the reader opposed to projects involving hydrologic modeling predictions. The motivation for this analysis by Oreskes, *et al.*, is not altogether clear, but the politicization of technical issues is often accompanied by a similar kind of rhetorical slide; hydrologists would do well to be able to muster an articulate reply. In this case, observers might be more interested in Putnam's *practice of science*, and less impressed with what Clark Glymour might characterize as "a paradigm of misplaced rigor".[50] If so, they will focus on what the analysts of logic would call negating the converse: to demonstrate the reliability of a model for decision-making does not require showing that it is true in every particular.

When Oreskes and her co-authors take up the actual IAEA definition of validation given above (p.42), which calls for a "good representation of the actual processes occurring in a real system", they claim this is impossible because "the establishment that a model accurately represents [these processes] is not even a theoretical possibility". This statement is only strictly true if the IAEA takes "good" to mean "exact". Familiarity with the rhetoric of validation as it relates to investigatory methodology will strengthen the contributions of hydrologists to debates over both.[51]

Logic divorced from practice is not likely to lead to pertinent conclusions regarding an applied science like hydrology. Putnam has similarly criticized Popper for making no substantial distinction between knowledge and conjecture, and claims: "Popper can maintain his extreme skepticism only because of his extreme tendency to regard theory as an end in itself".[52] Objecting to this philosophical wrenching of science from its practical base, Putnam concludes:

> Practice is primary... The primary importance of ideas is that they guide practice... We judge the correctness of our ideas by applying them and seeing if they succeed; in general, and in the long run, correct ideas lead to success, and ideas lead to failures where and insofar as they are incorrect. Failure to see the importance of practice leads directly to failure to see the importance of success.[53]

Putnam's views align with Jacob Bear's insistence that "there is no alternative" to the use of predictive models (p.53). Hydrologic models are often, for better or worse, practical "guides to action"; the evaluation of causal theories and models cannot be divorced from an ultimately practical motivation. Similarly, rejection of a strict validation should not preclude the practical notion of assessing what confidence is associated with particular ideas. The "importance of success" will figure heavily in our later exploration of process-oriented research in hydrology and its relation to predictive models.

The Waste Isolation Pilot Plant (WIPP) near Carlsbad, NM is a U.S. Department of Energy investigation that may lead to the permanent disposal of low-level radioactive waste some 2,000 feet below the surface in semi-plastic salt formations. Yucca Mountain in Nevada is the sole location being considered for the permanent disposal of high-level radwaste. These and similar projects have brought the issue of validation of hydrologic predictions to the forefront of policy debates. On the one hand, public safety demands a critical review of the basis for important decisions. Whatever the merits of the disposal projects, if hydrologists cannot defend what they claim to know, they will lend only an uncertain weight to the resolution of these debates. On the other hand, typical opposition logic insists on virtual certainty, then shows that certainty is either operationally lacking or programatically impossible. The result, if the arguments are treated as credible, is a very effective obstructionism. The distance between such partisans and the operational reality of science calls to mind a comment from the appropriately anti-realist Berkeley: They "have first raised a dust, and then complain, we cannot see".[54] Certainly Shrader-Frechette has made ample use of the non-verifiability of groundwater models as part of her argument against geological disposal of nuclear waste.[55] Similar issues are commonly played out on a lesser scale involving the siting and design of landfills, industrial operations, and the cleanup of the same.

On the other hand, some writers appear to be attempting to invigorate the feeble autonomy of hydrologic science, and to discredit what they consider unrealistic standards imposed by regulation. Certainly Konikow and Bredehoeft imply that the obsession with validation is detracting and distracting from scientific accomplishment. Others may feel the regulatory emphasis on certainty is a bureaucratic obstacle to critical engineering projects. If so, emphasizing non-verifiability without accounting for progress may be self-defeating. If by their own admission hydrologists can provide neither certainty nor a rational plan for progress, they appear to sanction overt politicization of decision-making with respect to technical projects. If directly applied models are so unfounded in practice, so unjustified in their conclusions, and so unpromising for the future, outside interests may reasonably conclude that what is needed is not less but more extra-scientific control. If hydrologists do not follow the prescriptions of philosophers, for example, some will argue they should. Under the circumstances, there will be those who reject the assertion of Konikow and Bredehoeft that "society's actions will be based upon our professional judgments".[56] Hydrologists cannot hope to satisfy a prescriptive program that does not accurately depict the standards and capabilities of mainstream scientific research.

Thus even while Konikow, Bredehoeft and (maybe) Oreskes seem to argue for some slack in the validation of models, their rhetorical strategies undermine their bids (if that is what they are) for a reassertion of scientific autonomy in hydrology. There is no question that from one perspective, non-verifiability appears to devalue the opinions of the so-called experts. It is equally certain that hydrologists, like geologists, have been far too generous in letting others write their philosophy for them. Of greatest importance to our project are the methodological implications of this discussion; we are inclined to think Popper was onto something when he announced: "I shall try to show that *the non-verifiability of theories is methodologically important.*"[57] And therefore we ask: what are the possible non-trivial methodological consequences of accepting non-verifiability? What connection would such a concession have to an account of progress?

4.4 Toward a Philosophy of Hydrologic Models

Projects of practical (strategic) import are often pursued by constructing hydrologic models that figure in causal arguments. As we have seen, the overall effort is properly termed a causal theory, since specified rules, conditions and forces produce subsequent behavior or observed conditions. As noted in Chapter 1, the model itself (as a geometric structure within which certain processes and properties apply) serves an intermediate function, allowing the derivation of expected behavior from given preconceptions. Since these expectations frequently have implications for public policy, the strength of the analogy between model and prototype is often the subject of technical and legal disputes. Much-maligned regulations or courtrooms attempt to apply either codified or more *ad hoc* standards for the assessment of this analogy.

Konikow and Bredehoeft give us two ideas from Popper. The first is that theories and their included models cannot be validated with absolute certainty. Vague references to Popper are not, of course, the only route to this conclusion. Oreskes, Shrader-Frechette and Belitz take a more detailed approach, and make several arguments against certainty in groundwater models.[58] Their principal claim is that hydrologic models are open systems, and therefore incapable of verification (note that by this they mean what we have usually referred to as "validation" in deference to regulatory language). For the rest, they mostly dwell on the "semantic ambiguities" that Konikow and Bredehoeft warned us about. Argumentative differences aside, Oreskes, *et al.*, reach a conclusion similar to that of Konikow and Bredehoeft, namely that models in the earth sciences cannot be shown to be strictly true. Little time was spent on this idea above since it is a trivial result in the philosophy of science. It is little more than common sense, given the use of idealization and approximation.[59]

A more compelling explanation for the systematic failure to validate applied models appeals to scientists' own understanding of what they do. For example, the subsurface is sometimes investigated using electro-magnetic methods in which the electrical resistivity or conductivity gives some idea of subsurface materials through which a current of known strength is passing. The standard governing equations for these methods of analysis are all derived assuming the presence of flat, homogeneous layers.

In practice, any number of layers may be inferred in a model of the system, but the indicated values for resistivity or conductivity will only be strictly accurate insofar as the subsurface actually consists of flat, well-defined, homogeneous layers. In the event this condition is not met in the field (it usually is not), error enters into the results. Scientists operating with an implicitly instrumentalist view are not unnerved by this situation, and standard research almost invariably includes 1) delineating the domain within which the results from various imperfect investigatory tools are "reasonably good"; and 2) searching for ways to more reliably interpret incomplete and possibly inaccurate data. It is not necessary to believe that the theory is perfect in order to gain useful insight, as is illustrated by the routine use of geophysical methods in, say, oil and gas exploration.

Theoretical derivations in hydrology are likewise only strictly applicable (*true*) for physical systems in which the assumptions are strictly correct.[60] As we have seen, idealizations, approximations and guesses often figure prominently in models because basic field investigations and related experimental work are incomplete. Derived expectations therefore warrant predictive confidence only insofar as actual and modeled conditions are equivalent. Models are therefore the target of certain analyses (such as that by Oreskes, *et al.*) that conclude models are frequently (or always) doomed to failure for the same reasons that prompt their use - *i.e.*, the need for simplification and idealization.

Certain conclusions by Oreskes, *et al.*, are supported by our study. These include that models are often of limited practical usefulness, and may be best used to clarify system analysis. Their characterization of groundwater models as "open systems" is incorrect, however, and potentially important. They allege that the model remains open due to uncertainty about the structure of the real system, the proper input parameters, and the processes operating within the system.[61] As a result, they conclude that validation is "not even a theoretical possibility". They refer, of course, to the validation of the applied model. Despite these objections, the model itself is actually closed, if numerical aberrations are avoided. Simulated flow and transport within a numerical model *can* meet the standard conditions imposed on a closed mathematical construct: all operations take place among exactly defined terms according to fixed rules.

Far from being dismissed as open and unverifiable, therefore, numerically validated[62] models are best thought of as *perfectly controlled experiments*. Properly posed analytical and numerical experiments do not leak unless they are supposed to, whether simulating a single pore or an aquifer. The model is simply a geometrically-defined construct within which certain properties hold and certain processes operate; these experimental conditions have consequences that can be quantified precisely because - within the model - mass, momentum and energy are conserved and continuously audited. The geometry, processes, fluxes, initial and boundary conditions, and parameter values can be absolutely known; they are known because they are the arbitrary choice of the modeler.

The model thus constitutes a mathematical formalism of the sort that - as Oreskes, *et al.*, point out - is capable of complete theoretical validation. This does *not* mean the model results are unique and precisely repeatable, any more than is the distribution of gas molecules released within a closed isothermal box. The closed model is capable, however, of a similarly constrained response. Given the conditions and laws listed, repeated runs can establish a range of possible model responses; this range is capable of validation, in that known causes produce a limited range of model response.[63] Given sufficient simulations, the mean model response becomes invariant and indicates the most probable consequences of the specified conditions.[64] This is more than a numerical code validation; it is a demonstration of causal relations, given the conditions and rules within the model.

Uncertainty about the linkage between model and prototype thus does not result in an "open" model, although it may raise serious doubts about the *applicability* of the model to the real system, and therefore about its validation in the regulatory sense. Naturally, theoretical validation of a controlled experiment does not immediately address the correspondence of model results to the eventual state of real physical systems; the latter are rarely so neatly describable as are controlled experiments, nor are their conditions and rules so readily known. The validation of causal relations within the model is not meant to suggest that the model is as yet a valid representation of any particular physical system. It is, however, important to clarify that *there is no theoretical barrier to constructing closed model systems that incorporate the complexities of hydrologic systems*, however daunting the practical obstacles may be. Oreskes, *et al.*, by contrast, do not leave their readers much hope that groundwater models can ever overcome their logical difficulties. They say: "The burden is on the modeler to demonstrate the degree of correspondence between the model and the material world it seeks to represent".[65] By their analysis, this would seem to be a grim prospect.

In this vein, the dialogues of Galileo offer some exceptionally instructive commentary. Ernan McMullin has written extensively on the nature of Galileo's science;[66] among the episodes McMullin considers is the following. Galileo, in the guise of Salviati, has produced a geometric physical argument (now discredited, as it happens) for objects to remain anchored to the spinning earth. His interlocutor[67] Simplicio airs the doubts of the Scholastics (Aristotelians), who insist that the practice of mathematical formalisms and idealizations leads to *necessarily* false conclusions:

> mathematicians may prove well enough in theory that a sphere touches a plane at a single point, a proposition similar to the one at hand; but when it comes to matter, things happen otherwise. What I mean about these angles of contact and ratios is that they all go by the board for material and sensible things.[68]

The roles of abstraction, simplification and idealization are clearly being questioned, at least when practical results are wanted involving imperfect objects and situations. Simplicio has serious doubts whether the idealized calculation can actually be relevant to events in the non-ideal world. Galileo, however, answers that the "geometrical philosopher"

> when he wants to recognize in the concrete the effects which he has proved in the abstract, must allow for the impediments of matter, and if he is able to do so, I assure you that things are in no less agreement than are arithmetical computations. The errors lie, then, not in the abstractness or concreteness, not in geometry or physics as such, but in a calculator who does not know how to keep proper accounts.[69]

Hydrologic modelers are equally concerned to demonstrate, if possible, that "proper accounts" have been kept. As indicated, this should not be an insurmountable problem in the assessment of cause and effect within the arbitrary closed system of the causal model. Problems arise as predictive models require reassurance that the conditions specified in the model are approximately *true*. The remainder of this thesis is devoted to an account of how hydrologists "allow for the impediments of matter" in both the construction and evaluation of models. Earlier, in Chapter 2, we surveyed the misgivings of leading hydrologic modelers. At this point we find ourselves in better position to consider the logical role of history-matching in model validation.

4.5 The Logic of Model Testing: History-Matching

Since *a priori* model assessments are needed for practical decisions, standard modeling procedure must evaluate the proposed analogy. Applied groundwater models do not often lead to cause-and-effect conclusions that can be readily checked against our experience. The equivalence of model and prototype behaviors is not normally evaluated directly; instead, a reasonable match is sought between model results and certain field information, as described in Chapter 2 and observed in Chapter 3. The model is typically used to predict historical data at equilibrium and/or under transient conditions. Although the applied model is causal in its argumentative structure, there are often significant gaps in site characterization due to inadequate data. As these are filled and smoothed, empirical fitting parameters and secondary geometry may be adjusted to achieve a better match of model output to the historical data. A reasonably close match is often taken to justify inferring that model structure and conditions adequately capture the essential features of the real system. This inference in turn warrants confidently using the model to predict the future condition of the real system.

Many arguments have already been advanced to discredit the idea of strict validation of applied groundwater models. Primary among these is the availability of multiple, mutually inconsistent models, each of which may satisfy the history-matching tests (as recounted in the history of the Tesuque models). The indirect and qualitative validation scheme recapped above cannot eliminate the objection that many alternative model designs might result in decent history-matches, but respond differently when projected into the future.

Konikow and Bredehoeft advance a second Popperian idea as a possible methodological implication of non-verifiability: progress must stem from attempts to invalidate hydrologic models.[70] Oreskes, *et al.*, also conclude by stating: "Models are most useful when they are used to challenge existing formulations, rather than to validate or verify them".[71] What does this mean? Can crucial tests be designed to reveal when a theory is not "keeping proper accounts"? Are groundwater models actually testable, either before implementation or after?

Earlier the conceptual model was characterized as an hypothesis or theory of a particular system, a description that agrees with that of Konikow and Bredehoeft.[72] Only the most obviously incompetent primary theory can be eliminated by test immediately. Results sometimes indicate that something is obviously wrong with model construction. Outside these and similar rather narrow instances, models are only subject to the semi-qualitative appraisals of experts, in the manner described in Chapters 2 and 3, including sensitivity analyses to determine the most important parameters. Certain structural options might be eliminated by comparative model studies. For example, the significance of a leaking aquitard might be investigated by building a 3-layer model and turning on and shutting off the aquitard and lower aquifer; if model results are insensitive to leakance at plausible rates, the model can be justifiably simplified to a single layer. The effects of radically different model structures (as at Pojoaque) are much more difficult to assess and complicate an assessment of the source of model error.[73]

Poor history-matching performance is unlikely to falsify a model in any case, in the sense of forcing the modeler to abandon the model altogether. Standard history-matching does not directly test grand conceptual theories; modelers typically make qualitative assessments of the minutiae of reported data and model output. At some point, as we have observed, they usually decide that: "A more precise adjustment of the model is not justified by available data" (Hearne); or, somewhat more positively: "The representation of the physical hydrogeological system in the model is substantiated by available data" (McAda and Wasiolek). A National Research Council committee observed (italics added):

> Since rigorous statistical validation tests are generally not appropriate in ground water applications, model validation is typically an ad hoc exercise that does not have a firm scientific foundation. Instead, model parameters are adjusted until a "reasonable" fit is obtained and the result is presented as a "validated model". *Modelers practically never declare their models to be "invalidated"*, primarily because ground water models nearly always have enough adjustable parameters to fit a limited set of field observations.[74]

Thus the auxiliary hypotheses (specific flow conditions, properties, minor geometric decisions, boundary conditions, and other "adjustable parameters") are recognized as non-fixed details; within reason, these can be cut and fit as needed, thereby "immunizing" the main body of the conceptual model against invalidation.[75] For example, when Glenn Hearne oriented the strike of his Tesuque aquifer model due north, flow in the system was to the north, not to the west and south as it was known to be.[76] Hearne did not jettison his entire model, nor any of its essential choices; he simply adjusted the strike to another plausible orientation that gave better results. Immunization is in fact just a secondary calibration using new information.

While it is unobjectionable to say that faulty theories should be eliminated if possible, it is another thing to say that this is the proper focus of scientific investigations. The complications arise from the fact that the conceptual model is really a set of bundled hypotheses of varying impact. Poor history-matching could conceivably result from minor defects, suggesting that it is not necessarily efficient to discard models on the basis of poor history-matching. Immunization and the natural resistance of modelers normally defeats falsification. As a result, some writers have urged a more thorough archiving of predictive models and routine testing of their accuracy by re-visiting modeled systems after some time to catalog and contrast modeled and measured responses.

4.6 The Logic of Model Testing: Post-Audits

From the technician's viewpoint, excitation of a system generally provides greater understanding than a static analysis. Hence pump tests and the like provide a basis for better model construction, by generating more relevant history (and, of course, spatially-averaged parameter values). And surely the most revealing excitation of the system in the evaluation of predictive models would be to continue or introduce the contemplated stresses themselves. For example, the best way to evaluate the Pojoaque River Basin model predictions regarding an expanded irrigation system would be to implement that very system and monitor drawdown and streamflow later as part of a post-audit.

A more extended effort to determine if an applied model keeps "proper accounts" thus naturally leads to an emphasis on the post-audit. Anderson and Woessner, in fact, *define* "model validation" on this basis: "Model validation, as defined here, tests whether the model can predict the future. This type of validation test has been called a predictive validation [by C.F. Tsang] or a postaudit".[77] This is clearly a much more relaxed view of validation than we have been entertaining of late, but it illustrates Putnam's grievance with Popper: "not invalidated does not mean we understand anything!" Although post-audits obviously cannot help make the original decision, they can still yield valuable hindsight perspectives on modeling choices.

Modelers clearly feel they learn and gain in understanding through exercises like the post-audit; not only might a post-audit highlight the source(s) of model failure, but even confirmed or corroborated models would represent real gains, despite their epistemological status as merely "not invalidated". Validation - in the sense used by Anderson and Woessner - can "add to the fund of knowledge". Putnam recognizes the commonplace and commonsensical inclination to think that corroborating instances make theories more probable in the utilitarian sense that one is more inclined to base opinions and recommendations on them. Many post-audits under different conditions and at different sites might allow modelers to improve their models, both generically and at specific locations. The search is for a pattern to both success and failure that might eventually lead to better short-term decisions.

This hunt is bound to be at least partly qualitative, since a good match between model predictions and real system response does not actually constitute a validation of the model in the stricter sense that the term has been used here. While it is surely valuable to know if model predictions were reasonably accurate, the model could always be "right" for the wrong reasons,[78] and therefore of uncertain value for future applications. At the same time, a poor match is again unlikely to dismiss a model as useless. Similar immunizing strategies can be employed as in the history-matching stages of model construction. For example, in the preliminary discussion of directly applied models, we noted de Marsily's argument that the basic structure of his model is excellent; he blames its poor predictive capability on auxiliary hypotheses about long-term climatic forcing functions. The core of his model is thus preserved and not open to falsification. It is partly as an effort to eliminate this sort of confusion - in which predictive models that err badly can still be described as excellent - that we consider predictive models to include their auxiliary hypotheses.[79] It may be in de Marsily's case that in fact the source of error is limited to the vagaries of weather prediction; if so, he is right to imply that nothing is to be gained from a wholesale rejection of his model as "invalid". The problem remains, therefore, of identifying the source(s) of model failure, since a fairly limited set of models is open to significantly challenging direct tests.

The post-audit is actually an extended history-matching exercise, that still does not address directly the model/prototype analogy; it is thus no more conclusive in either validation or falsification than the original history-match (though accumulating evidence may be qualitatively convincing). Not even de Marsily's demonstration of a third successful history-matching gives us any reason to think his model is a useful predictive tool, unless we now believe he can predict the weather for the *next* ten years. Despite its occasional persuasiveness, the post-audit is not yet a definitive assessment; the soundness of the modeling procedure and reasoning cannot be empirically demonstrated even after the fact.

Anderson and Woessner have strongly promoted the use of post-audits, even as they concede both the "difficulties of carrying out a successful validation and the low probability of success".[80] The purpose of post-audits is less to validate models than to highlight the source(s) of error, if possible. Their idea is not just to salvage models but to improve them. By characterizing post-audit success as improbable, however, Anderson and Woessner come very close to saying that *hydrology is not yet a predictive science*, certainly not yet a reliably predictive science. From this view it becomes apparent that, at least for now, the fundamental goals of scientific and legal/regulatory investigations may not often be easily reconciled. Hydrologists may make predictions as a regular part of their practice, but the question has now been clearly raised as to whether there is usually an adequate foundation for these claims. Neither the usefulness of accurate prediction, nor regulatory insistence, nor hydrologists' willingness to try, makes hydrology a credible predictive science; credibility (like validation in C.F. Tsang's phrase) derives from a thorough understanding of the hydrologic system of interest. Here we find the core of the methodological implications of non-verifiability. Glenn Hearne sounds the critical challenge when he says: "Confidence in the predicted response [is] based on a subjective appraisal of the analogy between the... aquifer system and the model".

Now that it is apparent that model appraisals are qualitative and to some extent uncertain - due to both practical and logical obstacles - how do hydrologists gain confidence that they have grasped the essence of a physical system? The challenge for both hydrologists and those interested in their opinions is to eventually meaningfully salvage the concept of validation for practical applications. Extending the question a little further, what is needed to improve the model/prototype analogy? The central importance of an account of progress and the link between modeling practice and supporting experiments are the focus of our remaining sections.

4.7 From Vulcan to Minerva: Methodological Implications of Non-Verifiability

Francis Bacon makes it perfectly clear that the practical utility of science is what makes it important. He is equally emphatic, however, that a "premature tarrying" over practical affairs can be self-defeating in the quest for scientific progress. With respect to hydrology, there is no question either that strongly predictive models will be needed in the future, nor that variously imperfect models will be used in the meantime, despite all that has been said concerning the uncertainty of the results. Jacob Bear (pp.53, 93) has pointed out the managerial implications of uncertainty, which should modify the objectives and expectations of present-day modeling exercises. He emphatically points out that *there is no alternative* to regarding the best available model as a reasonable representation of real systems of interest. The story of alternative Tesuque aquifer models has complicated this operational realism.

Nevertheless, since predictive models will and must be used, *a priori* model validation remains an essential concept and practice, however it might be modified in some circumstances by the recognition that a reasonable engineering accuracy or dependability is what is wanted, rather than some more absolute guarantee. Recognition of the validation dilemma does not eliminate the need to make the best choices possible, though it might, as Bear suggests, modify the alternatives considered. Water managers need to consider to what extent their plans rely inappropriately on non-verifiable models. Applied modelers might apply themselves to a similar delineation of the valid questions they are prepared to answer, rather than accepting directly the challenge of any and all external applications.[81] From either direction, the applied science is clearly experiencing its full share of growing pains. Given the current state of affairs, we find ourselves in the position that Stillman Drake describes as "erecting a beacon to denote the presence of a shoal that we cannot remove".[82]

Pressing practical goals that prohibit casual dismissal of the notion of validation also demand an account of progress and some idea of how it can be achieved. The relationship of models, experiment and theory may well be reconsidered under the circumstances. In Chapter 2, a methodological debate among surface water hydrologists provided an introduction to the use of causal models. Vit Klemes bemoaned the more or less medieval state of affairs and championed the development of causal models in surface water hydrology, insisting:

> The unifying framework for all these *ad hoc* models can hardly be anything else than a physically consistent model of the catchment mechanisms, i.e., a causal theory of the hydrologic cycle... This line of inquiry seems to be the most promising way out of the unenviable present situation that has been aptly compared (Dooge, 1978) to "a riot of growth reflecting a variety of scale, colour and type and... a cacophany of noise... confronting... a traveler lost in the jungle.[83]

Bacon argues that *our steps must be guided by a clue*. Klemes and other advocates of causal models argue that this clue must be a "causal theory of the hydrologic cycle". By this Klemes clearly means much more than the bundled hypotheses that constitute the usual applied hydrologic model. We have considered the difficulties inherent in history-matching as a means for validating such models. Referring to the impressive possibilities for causal methods, Klemes pointed out:

> The potential of causal-hydrologic models lie in their ability to derive the behavior of a hydrologic process for a given set of states of nature (physical variables) from the dynamic mechanisms of the process, without recourse to model calibration by empirical fitting. Consequently, causal models offer a possibility to predict the behavior of a hydrologic process under conditions that did not exist during the process-recorded history, i.e., under conditions for which an empirical model cannot be constructed... It is not a question of prediction accuracy for known conditions but one of model credibility in unknown conditions. It is the difference between blind extrapolation and sound judgment.[84]

The availability of sound causal models would de-emphasize or even bypass history-matching as a means of model evaluation. If the essentials of a system were more directly understood, Klemes says, system response to any given conditions could be predicted, without recourse to calibration and indirect means of "verification". Or, as Bacon has it, "For the form of a nature is such that, given the form, the nature infallibly follows".[85] Such, in any event, is the rationale for causal models.

Just a few years later, as we have seen, Klemes was grumbling about the new developments in terms that to date give a general summary of the situation in directly applied groundwater modeling, as well:

> The so-called hydrological conceptual models have been conceived in the spirit of good science, trying to improve on the black box models by introducing hydrological mechanisms and processes into the modelling of rainfall-runoff relations. The problem was that most of these mechanisms were not well understood; nor were their interactions, either in general or in specific conditions of different river basins. Thus the resulting models are concoctions of a few facts and many artificial constructions, in which the individual processes are represented by postulated rather than by established mechanisms and interactions and by assumptions arrived at by simplistic reasoning rather than serious hydrologic research. In effect they are, for the most part, just complex assemblages of black or, at best, dark grey boxes.[86]

As we have seen, the uncertainty in associating a groundwater model with any particular real system generally forces a reliance on *ad hoc* history-matching "fixes" that share some of attributes of empirical models. Indirect "validation" shares in the opacity of black box models; knob-twiddling calibrations can adjust model output to better fit historical records without a detailed understanding of the relation between model design and behavior.[87] Automated parameter optimization (a numerical technique leading quickly to optimal history-matches) does not strengthen a weak logic of model testing; as Henry Thoreau said, "All our inventions are but improved means to an unimproved end". In a better world, empirical investigations would more revealingly aid in the building of causal models, with less emphasis on indirectly "verifying" them. Directly controlling the model/prototype uncertainty is essential if hydrology is ever to be a legitimately predictive science.

Although groundwater models are typically causal in design, their predictive potential is not being fulfilled due to continued weakness in the "theory of the hydrologic cycle", both in general and as explanations of particular systems. No one imagines it to be a brief project to advance hydrologic models from "concoctions of a few facts and many artificial constructions" to detailed and faithful analogies of real systems. A local theory of the hydrologic cycle is precisely the conceptual model outlined in Chapter 2 in combination with certain general laws. The defects in present-day applied causal models are, of course, in large part symptoms of locally inadequate hydrogeological understanding. The work needed to advance this understanding can take the form of engineering technology development, bench-scale laboratory investigations, formal mathematical derivations, or what we might call research-oriented or process-oriented applied models. Most of these areas are pursued by those less directly tied to practical applications. The current trend to de-emphasize fundamental research appears particularly foolish under the circumstances.

In some quarters, at least, non-verifiability therefore results in a renewed emphasis on research. Chapters 5, 6 and 7 will explore the relationship between, on the one hand, process-oriented experiment, models and theory, and, on the other hand, progress toward compelling if uncertain accounts. The implications of non-verifiability also extend to what would appear to be routinely applied models. Despite their involvement in applied modeling - and despite the prominence of validation questions in regulatory language - several leading modelers have expressed another view on the best purpose of their craft. They tend to dismiss the current standards as simply unrealistic, and in the process propose an intriguing answer to the question (p.42): what good are models if their predictions cannot be validated? These writers appear uniform in their opinion that it would be more appropriate to view models as exploratory and expository tools - *research tools* - rather than as factual evidence for predictive purposes. For example, Anderson and Woessner emphasize proper documentation, rather than *a priori* confidence:

> The issue of validation is mainly a regulatory one, not a scientific one... Because our understanding of a system will always be incomplete a model can never be proven valid from a scientific standpoint. Hence regulators must be content with some degree of partial validation... Given the difficulties of carrying out a successful validation and the low probability of success, it seems wise to seek an

> alternative to validation as a regulatory objective... The regulatory focus should shift from demands for validation to demands for good modeling protocol, including providing a complete description of model design, a thorough assessment of model calibration, and an uncertainty analysis. Existing protocols for validation... should be replaced by protocols for performing and documenting the entire modeling process.[88]

This is much plainer talk than the sometimes open-ended conclusions of Konikow, Bredehoeft and Oreskes. Anderson and Woessner are quicker to label the regulatory requirements as bogus and unscientific.[89] It is apparent from their comments that they feel current regulatory and remediation procedures place an unreasonable burden on what should be viewed as preliminary models. In the context of a post-audit report, Konikow comments that at present "the predictive accuracy of these models does not necessarily represent their primary value", again suggesting that a greater benefit may be to highlight research needs.[90] Konikow has elaborated elsewhere:

> If models cannot be validated, why are they useful? Models provide a tool for critical analysis. They are a means to organize our thinking, test ideas for their reasonableness, and indicate which are the sensitive parameters. They point the way for further investigation. They help formulate critical experiments with which to test hypotheses... They serve to sharpen our professional judgement.[91]

Among such experts, versed in both hydrogeology and the strengths and weaknesses of the modeling approach, the primary benefit and even the purpose of applied modeling is to clarify one's *thinking about systems* by showing the implications of various conceptual models.[92] Occasionally one might identify defects in the models, leading to improved iterations; model feedback and post-audits are used to improve the models, and hence, understanding. Straightforward predictions based on models, already suggested to be futile due to their inherent uncertainty, here become a goal of secondary importance. Anderson explains:

> It is my philosophy that applying a model is an exercise in thinking about how a system works. Automating a modeling exercise to the extent that the model can be used by someone lacking the necessary background in ground-water hydrology destroys the essence of modeling. It is the thought

> process needed when applying a model that should lead to a decision, not necessarily and certainly not exclusively the answers generated by the model itself.[93]

Despite the resistance of decision-oriented water managers, the use of multiple models without inflexible preference, in the manner proposed by Jacob Bear (p.33), recalls the classic geoscience method of multiple working hypotheses, promoted by T.C. Chamberlin.[94] Modeling lends itself to the playing of these *What if?* sorts of games, whose results can sometimes be checked against data or experience. The distinction between process-oriented and applied is not strictly or even necessarily based on *scale*; it is really more a question of *purpose*.

Though such models may be linked to specific systems and even to practical decision-making, an effort may be made to advance a more general understanding, both of real systems and processes and of the possibilities and pitfalls of model construction. Such efforts implicitly recognize the status of models as closed causal experiments. Models, validated in the unapplied sense of Section 4.4, allow the investigation of the effects and interactions of different geometries, processes and parameter value distributions. Slightly further removed from direct applications, enhanced computational efficiency and capacity can eliminate some of the enforced simplicity that causes models to lose validity. Models with one foot on either side of the process/applied demarcation might thus proceed on several distinct levels: 1) code building, possibly for new and more inclusive models; 2) code validation, or efforts to certifiably eliminate numerical aberrations; 3) model-driven deductions from the theoretical *What if?* games, abstracted from practical concerns; and lastly, 4) attempts to match idealized model structures to specific sites for predictive purposes.

All such research postures recognize that models provide at best a very conditional "proof": *if* the conceptual model is faithful to the real system (including available data and any interpolated and/or extrapolated parameters), then the results can directly address performance criteria for the model. Since model results tend to reflect the extent of the modeler's understanding, more direct insight into systems of interest is also needed. Numerical codes at least somewhat abstracted from practice can be developed to incorporate new information, but additional research that is needed to constrain model design. These constraints are in the form of general processes and site-specific information.

To some degree the shortcomings in applied hydrology are technological, or at least the solutions are. More sophisticated subsurface insight at the usually large scale of interest depends in part on improvements in geophysical methods. More efficient characterization methods that are both broader and more inclusive are needed to improve both parameter estimation and scaling of geologic detail. It is obvious that improved data collection can begin to close the gap between model and prototype. The development of "larger and wider" theory, as Bacon put it, is also clearly dependent on additional investigations into processes.

Process-oriented experimenters contribute to the theory of the hydrologic cycle (at any scale) by advancing insight into as yet oversimplified, ignored or assumed factors. These factors include not only distinct physical, chemical and biological processes, but also relational issues, such as how to scale up from bench-scale experiment or measurement to the field scale. "Process-oriented applied modelers", on the other hand, contribute by pursuing better representations of critical mechanisms and their interactions within applied models.

Improved geophysical measurement, process representation and numerical sophistication may, in combination, lead eventually to improved predictive models. All of this research is intended to narrow the gulf between model and prototype; any success in this direction will decrease the burden on the indirect model tests described above by confining model choices. As important conceptual choices in the model map more closely onto those of the real system, both history-matching and post-audit validation become more routinely confirming. As Klemes initially hoped, increasing the linkage between model design and system reality will force model response to more likely resemble that of the real system; confidence in predicted real system response would then share to a larger degree in the absolute validation of the predicted response. Here we see the importance of clarifying the status of the model itself as closed and both numerically and causally verifiable.

The remainder of this study is devoted to process experiment, and the means by which hydrologists gain confidence in causal theories that cannot be, strictly speaking, validated. The critical need for more decisive and reliable geophysical exploration tools is beyond the scope of this study; the details of numerical simulation are also of only passing further interest. This study of methodology in hydrology began as an effort to relate four things: observation, experiment, theory and models. Thus far models have figured in the application of causal theories to particular groundwater problems. Within process research, a more complicated give-and-take utilizes models as an intermediary between experiment and theory.

The occasional difficulties with experiments as crucial tests that rule out theories have been explained above. While Popper thinks progress consists of weeding out faulty theories, we are convinced with Putnam that an equally important role is played by crucial demonstrations that advance compelling theories. Numerically validated causal models that are applied to specific sites are often fed by the results of controlled laboratory experiments. These typically bench-scale (or smaller) efforts routinely employ idealized conditions to isolated specific factors in either cause or effect. Of such experiments, Bacon says:

> Therefore a separation and solution of bodies must be effected, not by fire indeed, but by reasoning and true induction, with experiments to aid; and by a comparison with other bodies, and a reduction to simple natures and their forms, which meet and mix in the compound. In a word, we must pass from Vulcan to Minerva, if we intend to bring to light the true textures and configurations of bodies...[95]

Hence we consider at last the experimental pursuit of a stronger model/prototype linkage through what Bacon called, in a happy phrase, *Experiments of Light*. We turn next to crucial demonstrations in hydrology, first as performed in an historical tale of the discovery and subsequent analyses of Darcy's law for macroscopic flow (Chapter 5), and then in a recent bench-scale investigation of solute mixing behavior at fracture junctions (Chapter 6). Of particular interest are the complementary roles played by Baconian empiricism and causal modeling methodologies in the pursuit of scientific explanation and model constraint. In the end, the shifting roles of models will emerge as the key to a coherent and compelling scientific procedure and evaluation. Before we attempt a synthesis of the prospects for progress in hydrology, we need to take a close look at process-oriented research in a laboratory setting.

4.8 Notes:

1. Brush, S.G. (1989), "Prediction and Theory Evaluation: The Case of Light Bending", in *Science*, **246**, 1 Dec 1989, p.1125.

2. Hassanizadeh, S.M. and Carrera, J. (1992), "Editorial", in *Advances in Water Resources*, **15**:1, pp.1-3.

3. Konikow, L.F. and Bredehoeft, J.D. (1992), "Groundwater Models Cannot Be Validated", in *Advances in Water Resources*, **15**:1, pp.75-83; and Bredehoeft, J.D. and Konikow, L.F. (1993), "Ground-Water Models: Validate or Invalidate?", in *Ground Water*, **31**:2, pp.178-9.

4. Flew, A. (1979), *A Dictionary of Philosophy*, St. Marin's Press, p.156, defines *Hume's Fork* as: "The increasingly popular nickname for an aggressive employment of Humes's fundamental distinction between propositions stating or purporting to state only 'the relations of ideas' and propositions stating or purporting to state 'matters of fact and real existence' [*Inquiry concerning Human Understanding*, IV:i]... It is its aggressive employment in Hume that makes his version the classical anticipation of logical positivism's challenge to choose between, on the one hand, analytic, a priori and necessary and, on the other hand, synthetic, a posteriori and contingent." Ian Hacking (1983), *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*, Oxford University Press, points out more plainly: "Popper agreed with David Hume, who, in 1739, urged that we have at most a psychological propensity to generalize from experience. That gives no reason or basis for our inductive generalizations" (p.4).

5. Casual acquaintance with the philosophy of Popper appears to resonate with many scientists. Statements like "scientific knowledge... is common-sense knowledge writ large, as it were"; or: "Theories are nets cast to catch what we call 'the world': to rationalize, to explain, and to master it. We endeavor to make the mesh ever finer and finer"; are unobjectionable. Popper's overview of his method and his obvious enthusiasm for scientific objectives make him a sympathetic figure. As a result, scientists often heap accolades on *Logic*. On the dust jacket of my copy, Sir Hermann Bondi and C.W. Kilmister rave:

    > Popper speaks as a working scientist to the working scientist in a language that time and again comes straight out of one's heart... Popper faces on every turn the problems against which the scientist runs up, and solves them... in full accordance with the procedure of science.

    Peter Medawar, from whom we took our initial orientation on method, is also a big fan: "I think Popper is incomparably the greatest philosopher of science that has ever been". Before we ask whether Medawar has forgotten his own spoof, and whether he is reading much more into his emeritus-tinted methodology than is reasonable, we will need to take a closer look at the details of Popper's philosophy.

6. Konikow, L.F. and Bredehoeft, J.D. (1992), p.77: "There are two principal schools of philosophical thought on this issue [validation]. One school, called positivism, holds that, '... theories are confirmed or refuted on the basis of critical experiments designed to verify the consequences of the theories'. One of the principal exponents of positivism is Thomas Kuhn. A second school, espoused by Karl Popper, argues that 'as scientists we can never validate a hypothesis, only invalidate it'." Curiously, Konikow and Bredehoeft credit Matalas, *et al.*, (1982), "Prediction in Water Management", in National Academy Press, *Studies in Geophysics: Scientific Basis of Water-Resource Management* as the main source for these comments. This article correctly states that: "More recently Kuhn (1962), Feyerabend (1963) and others have strongly criticized positivism, suggesting that science is a social enterprise, that the human process cannot be disassociated from scientific inquiry" (p.122).

7. Konikow, L.F. and Bredehoeft, J.D. (1992a), pp.77-78.

8. Konikow, L.F. and Bredehoeft, J.D. (1992a), p.78, in reference to Hawking, S.W. (1988), *A Brief History of Time: From the Big Bang to Black Holes*, Bantam Books.

9. Popper, K.R. (1959), *The Logic of Scientific Discovery*, Routledge (1992), p.31.

10. Popper, K.R. (1959), p. 38-9. Medawar has a great deal more to say on this topic, much of which is contained in Medawar, P.B. (1969), *Induction and Intuition in Scientific Thought*, American Philosophical Society. Medawar dislikes calling discovery induction, preferring to label it imagination; and if by induction we mean only the piling on of facts with no hint of an ordering or linking scheme, then clearly he is right. Hacking, I. (1983), *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*, p.248, prefers to distinguish three stages of activity: speculation; articulation/calculation; and experiment.

11. Popper, K.R. (1959), p.106. In the context of paradigms in science, Kuhn, T.S. (1962), *The Structure of Scientific Revolutions*, University of Chicago Press, 2nd edition (1970), p.113, comments: "What a man sees depends both upon what he looks at and also upon what his previous visual-conceptual experience has taught him to see. In the absence of such training there can only be, in William James' phrase, 'a bloomin' buzzin' confusion'".

12. Popper, K.R. (1959), p.37.

13. Rudolph Carnap (1891-1970) was one of the leaders of logical positivism, in which "the meaning of a proposition consists in the method of its verification" (Flew, A. (1979), p.214). In particular, metaphysical statements were rejected as meaningless, since confirming experience was unavailable. Carnap's philosophy was closely tied to linguistic theory that "showed how highly theoretical statements, that do not apparently describe immediate experience, are reduced by definitions to ones that do" (Flew, p.56), a progression that in Popper's view leads to testable hypotheses. From the other direction, this "confirmation theory" was intended to "delimit what kind of proposition or theory is confirmable (that is, capable of gaining support from experience), this being a necessary task of any kind of postivism" (Flew, p.71). This last is the "inductive" link that Popper denied.

14. Popper, K.R. (1959), p.52-3: "We may consider and compare two different systems of methodological rules; one with, and one without, a principle of induction. And we may then examine whether such a principle, once introduced, can be applied without giving rise to inconsistencies; whether it helps us; and whether we really need it. It is this type of inquiry which leads me to dispense with the principle of induction: not because such a principle is as a matter of fact never used in science, but because I think that it is not needed; that it does not help us; and that it even gives rise to inconsistencies".

15. Note that in Popper's terminology, *verification* occupies the same role as *validation* in ours.

16. Popper, K.R. (1959), p.40. He also says: "Nothing is easier than to construct any number of theoretical systems which are compatible with any given system of accepted basic statements [observations and necessary conditions]" (p.266). The standard view is summarized by Resnick, D.B. (1992), "Convergent Realism and Approximate Truth", in Hull, D., *et al.*, (eds.), *PSA 92*, Proceedings of the Philosophy of Science Assn, 1992, p.422: "Theories... have an infinite number of observational consequences; and since we can never check all of these observational consequences, we can never know that a theory is true; it could always be refuted by an observation".

17. See Popper, K.R. (1974), in Schilpp, P.A. (1974), *The Philosophy of Karl Popper*, Open Court, pp.133-143, in which Popper describes Darwinian selection as at best a "metaphysical research programme", not a falsifiable scientific theory.

18. Popper, K.R. (1974), "Autobiography", in Schilpp, P.A. (ed.) (1974), *The Philosophy of Karl Popper*, Open Court, p.29.

19. Popper, K.R. (1974), p.32.

20. Popper, K.R. (1974), pp.31-33. Despite his radically different account of science, Kuhn attributes some of his own later inspiration to a similar demarcation. He credits some of the development of his thinking and writing on the *Structure of Scientific Revolutions* (1962) to an opportunity to spend the year 1958-9 in a community of social scientists: "Particularly, I was struck by the number and extent of the overt disagreements between social scientists about the nature of legitimate scientific problems and methods... Somehow, the practice of astronomy, physics, chemistry, or biology normally fails to evoke the controversies over fundamentals that today often seen endemic among, say, psychologists or sociologists. Attempting to discover the source of that difference led me to recognize the role in scientific research of what I have since called 'paradigms'... Once that piece of my puzzle fell into place, a draft of this essay emerged rapidly" (p.*viii*).

21. See Hon, G. (in press), "Is the Identification of Experimental Error Contextually Dependent? The Case of Kaufmann's Experiment and Its Varied Reception", in Buchwald, J. (ed.), *Scientific Practice: Theories and Stories of Doing Physics*, University of Chicago Press, pp.170-223. Another famous test of Einstein's theories occurred a few years later when a solar eclipse allowed the observation of bending light rays. Stephen Brush has recounted the history of the light-bending experiment, its results and Einstein's reaction [Brush, S.G. (1989)]. Finally, Brush notes: "Popper seems not to have noticed that Einstein stuck to his theory even though the [redshift] prediction was never satisfactorily confirmed during his lifetime", a confirmation attributed to the Pound-Rebka experiment of 1960" (p.1125).

22. Popper, K.R. (1974), p.28-29.

23. Popper, K.R. (1959), pp.251-252, note *1.

24. Allchin, D. (1992), "How Do You Falsify a Question?: Crucial Tests versus Crucial Demonstrations", in Hull, D., *et al.* (eds.), *PSA 1992*, pp.74-88, recaps the story of the Ox-Phos (oxidative phosphorylation leading to the production of ATP) controversy in bioenergetics during the 1960s and 1970s: "Among philosophers, the notion of simple or 'naive' falsification has been heavily criticized and largely abandoned. Among scientists, however, the concept may function as a heuristic or rhetorical device... Peter Mitchell, for example, selfconsciously framed his novel chemiosmotic theory according to Popperian principles; others, as well, appealed to the scientific authority of falsification. Still others argued, without explicit philosophical reference, that single experiments were decisive against the opposing theory. But it is also clear that the researchers often did not follow their own advice or adhere rigidly to their own rhetorical claims. Elements of Mitchell's original hypothesis, for example, failed repeatedly in the early development of the theory to match actual observations... and the discrepancies certainly did not escape the notice of critics. Yet Mitchell persisted in his broader, more general program... At one point, in fact, he was 'feeling more and more that we... could succeed in falsifying the chemiosmotic hypothesis'... On the verge of crisis (perhaps), Mitchell dramatically revised the theory and introduced an arguably *ad hoc* concept... In brief, and perhaps to no philosopher's surprise - this episode exhibited no falsifications in large-scale intertheoretic debate" (p.75). Peter Mitchell later received the 1978 Nobel Prize in Chemistry for his chemiosmotic theory.

25. Vonnegut, K. (1987), *Bluebeard*, Delacorte Press, p.27.

26. Popper, K.R. (1959), p.48; all italics in original.

27. See Bardsley, E. (1993), "A Moribund Science?", in *Journal of Hydrology (New Zealand)*, **31**:1, pp.1-4: "Without getting too philosophical, a valid scientific alternative would be to seek to reject as many hypotheses as possible, on the basis that advances are made by refutation rather than verification" (p.3).

28. Popper, K.R. (1959), p.50.

29. Popper, K.R. (1974), p.32. Thus on the laboratory scale one might blame insufficiently controlled or sensitive apparatus, while on the planetary scale, one might blame slight inaccuracies in prediction on the gravitational pull of as yet undiscovered bodies. See the story of Universal Gravitation (UG) in Putnam, H. (1969), "The 'Corroboration' of Theories", in Schilpp, P.A. (ed.) (1974), *The Philosophy of Karl Popper*, Open Court, pp.221-240; also in Hacking, I. (ed.) (1981), *Scientific Revolutions*, Oxford University Press, pp.60-79. Popper attempts to deal with these objections in several places; see, *eg.*, Popper, K.R. (1959), pp.81-86; or Popper, K.R. (1974), pp.32-3, where he says "Immunization is *always* possible, but the evasion would usually be dishonest". He illustrates this idea first with the re-interpretation of Marxism in light of the Russian Revolution of 1917. Considering astronomical theories, Popper says a "prima facie falsification may be evaded" by the introduction of two sorts of auxiliary hypotheses: 1) those that are themselves testable. "We may regard this... as a *growth* in our knowledge"; or 2) those that are "merely evasive immunizing moves", and cheat as they "decrease the [informative] content".

30. Duhem, P. (1954), *The Aim and Structure of Physical Theory*, Princeton; Lakatos, I. and Musgrave, A. (eds.) (1970), *Criticism and the Growth of Knowledge*, Cambridge University Press. The denial of objective theory evaluation methods leads to more extreme views: "... others such as Feyerabend, advocate epistemological anarchy, under which there are no established criteria to separate rubbish from sensible theory" (Flew, A. (1979), p.71).

31. Popper, K.R. (1959), pp.112-135.

32. Popper, K.R. (1959), pp.146-214; pp.251-281.

33. Mulkay, M. and Gilbert, G.N. (1981), "Putting Philosophy to Work: Karl Popper's Influence on Scientific Practice", in *Philosophy of the Social Sciences*, **11**, pp.389-407: "Consequently, in relation to scientific practice, one can talk only of positive and negative results, and not of proof or disproof. Negative results, that is, results which seem inconsistent with a given hypothesis, may incline a scientist to abandon a hypothesis, but they will never require him to abandon it, on Popper's own admission... Thus the utter simplicity and clarity of Popper's logical point are lost as soon as he begins to take cognizance of some of the complexities of scientific practice and as soon as he makes the transition from his ideal scientific actor to real scientists engaged in research. The mundane Popperian scientist, then, seems to be no better off than his non-Popperian colleague" (p.391).

34. Mulkay, M. and Gilbert, G.N. (1981), p.391. Hacking, I. (1983), *Representing and Intervening: Introductory Topics in the Philosophy of Science*, Cambridge University Press, comments further: "According to Popper, we may say that an hypothesis that has passed many tests is 'corroborated'. But this does not mean that it is well supported by the evidence we have acquired. It means only that this hypothesis has stayed afloat in the choppy seas of critical testing. Carnap, on the other hand, tried to produce a theory of confirmation, analyzing the way in which evidence makes hypotheses more probable. Popperians jeer at Carnapians because they have provided no viable

theory of confirmation. Carnapians in revenge say that Popper's talk of corroboration is either empty or is a concealed way of discussing confirmation" (p.4).

35. See, *eg.*, Brush, S.G. (1989), p.1128, note 20: "Popper says he rejects a 'naturalistic' approach to the theory of the scientific method, that is, a 'study of the actual behavior of scientists' (*L.Sc.D.*, p.52; *Realism and the Aim of Science*, p.*xxv*), but forgets this disclaimer when he wants to give the impression that he is describing how science really works, not just prescribing an idealized method"; an example of the latter stance is found in Popper (1959, p.50): "Thus I shall try to establish the rules, or if you will the norms, by which the scientist is guided when he is engaged in research or in discovery, in the sense here understood".

36. Kuhn, T.S. (1962), *The Structure of Scientific Revolutions*, University of Chicago Press, 2nd edition (1970), p.126. Kuhn also comments that "data are not unequivocally stable", as scientists reconsider interpretations of them (p.121); and in connection with "paradigm shifts" that "previously completed work on normal science projects would now have to be done again because earlier scientists had failed to recognize and control a relevant variable" (p.58).

37. Popper, K.R. (1959), p.252.

38. Barber, B. (1961), "Resistance by Scientists to Scientific Discovery", in *Science*, **134**, 1 Sept 1961, pp.596-602. Barber gives many examples from past and present, involving major and minor discoveries, to illustrate his point that "because of their substantive conceptions and theories, scientists sometimes miss discoveries that are literally right before their eyes" (p.598). See also Stewart, J.A. (1986), "Drifting Continents and Colliding Interests: A Quantitative Application of the Interests Perspective", in *Social Studies of Science*, **16**, pp.261-279. Stewart attributes resistance to continental drift theory from 1907 to 1950 in part to "intellectual investments in terms of previous publications" on the subject (p.271).

39. Putnam, H. (1969), p.61.

40. The answer is not one to which most would leap at first blush. Scientists are instructed to "accept" the most improbable theory, on the grounds that it is most testable. It is at times like this that the gulf between prescription and practice is most evident. At other times, Popper backs away a little from this prescription, and employs a Darwinian metaphor as he says: "How and why do we accept one theory in preference to others? The preference is certainly not due to anything like an experimental justification of the statements composing the theory; it is not due to a logical reduction of the theory to experience. We choose the theory which best holds its own in competition with other theories; the one which, by natural selection, proves itself the fittest to survive. This will be the one which not only has hitherto stood up to the severest tests, but the one which is also testable in the most rigorous way" (p.108).

41. Putnam, H. (1969), p.62. In this regard, we again should recall Bear's comment (pp.53, 93) that there is no alternative to using predictive models, despite the associated uncertainty.

42. Popper, K.R. (1959), pp.265-276.

43. Mulkay, M. and Gilbert, G.N. (1981), p.389-392.

44. See also, Putnam, H. (1969), p.64: "I claim: in a great many important cases, scientific theories do not imply predictions at all". He supports this claim with an account of Newton's theory of universal gravitation (UG): "Popper's doctrine gives a correct account of neither the nature of the scientific theory nor of the practice of the scientific community in this case... Scientists did not try to falsify UG because they could not try to falsify it". Putnam contends that it is only the

combination of a theory - "a set of laws" - with auxiliary statements (AS) that gives rise to predictions. It is therefore most often the case that the AS - "highly risky suppositions" - are most at test in observations. The more obvious uncertainty of the AS tends to insulate the core theory from falsification, while encouraging the revising and replacement of the AS.

45. Lauden, L. (1984), *Science and Values: The Aims of Science and Their Role in Scientific Debate*, University of California Press, pp. 16-17, attributes the same shortcoming to writers like Feyerabend and Mitroff, who "have laid out elaborate machinery for explaining how disagreement could arise and persist (e.g., from incommensurability or underdetermination). But, as I have already hinted, these writers [he includes Kuhn here] are ill-equipped to explain how agreement ever congeals". On pp.4-5, Lauden has more to say on the "perplexing broad agreement in science", where he notes the comment of "the well-known philosopher of science, N.R. Campbell, [who] puts it quite bluntly: 'Science is the study of those judgments concerning which universal agreement can be reached'. Speaking for the sociologists, John Ziman concurs: '[Consensus] is the basic principle upon which science is founded. It is not a subsidiary consequence of the "scientific method". It is the Scientific Method itself.'"

46. As mentioned above (Chapter 2, note 22), we will use *non-verifiability* to mean the situation in which a theory is incapable of validation. The term non-verifiability has a rich history in the philosophy of science (for example, in Popper's writing), but contortions like non-validatability have been avoided by all.

47. Bauer, H.H., "Darwin on Trial [review]", in *Journal of Scientific Exploration*, **6**:2. In this review of Phillip Johnson's *Darwin on Trial*, a challenge to evolutionary theories, Bauer quotes and comments: "Johnson thinks in terms of 'the scientific method of inquiry, as articulated by Popper' - apparently unaware that Popper has few if any followers left. 'Karl Popper provides the indispensable starting point for understanding the difference between science and pseudo-science' - not according to most philosophers of science, he doesn't... 'If Darwinists wanted to adopt Popper's standards for scientific inquiry...' - but there are no mainstream scientists anywhere who want to do that".

48. Oreskes, N., Shrader-Frechette, K. and Belitz, K. (1994), "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences", in *Science*, **263**, 4 Feb 1994, pp.641-646.

49. This paper defines "verification" to be mean the literal truth of a model has been demonstrated. This is therefore a far stronger claim than a "validated" model, which they say only "does not contain known or detectable flaws and is internally consistent".

50. Glymour, C. (1980), *Theory and Evidence*, Princeton University Press, p.102.

51. Stephen Brush gives an example of the curious uses to which the validation debate can be put by outsiders. In 1982, creationists used Popper's attack on evolutionary biology as non-scientific (later recanted) to introduce legislation calling for equal time for creationism in Maryland schools, claiming that since "evolution-science like creation-science cannot be... logically falsified", it does not deserve preference in the classroom. Again, more than a passing familiarity with the terms of these debates is needed for effective participation (and, possibly, rebuttal).

52. Putnam, H. (1974), p.62.

53. Putnam, H. (1974), p.78.

54. Berkeley, G. (1710), *A Treatise Concerning the Principles of Human Knowledge*, introduction:§3.

55. Shrader-Frechette, K.S. (1993), *Burying Uncertainty: Risk and the Case Against Geological Disposal of Nuclear Waste*, University of California Press. By way of contrast, this co-author of Oreskes, *et al.*, (1994) shows a curious kindness to the much more woebegone models of the ecologists. In Shrader-Frechette, K.S. and McCoy, E. (1994), "Applied Ecology and the Logic of Case Studies", *Philosophy of Science*, **61**, pp.228-249, the authors grant the reasonableness of ecologists' efforts to "make sense" in many ecological prescriptions, despite the attacks on these studies (by others) for using "general theories that were untested 'in the real world', for employing inadequate 'rigor', and for drawing conclusions unlikely to be supported by other reasonable persons". In brief, they were attacked as Shrader-Frechette has attacked geological disposal of nuclear waste. In defense of Spotted Owl studies, however, Shrader-Frechette says: "Unfortunately, in investigating particular cases, no simple logic, such as hypothesis-deduction, is applicable. Instead, one must follow a method, a set of procedures and rules of thumb, that help one confront the facts of a particular situation and then look for a way to make sense of them through a set of informal inferences ('logic')... As a consequence of uncertainties about the relevant variables, researchers using the method of case studies have often been forced to use a 'logic' of informal causal, inductive, retroductive, or consequentialist inferences in order to 'make sense' of a particular example or situation". She concludes this explanation by noting parallels to the method Bernstein and Woodward used to evaluate the causes of the Watergate coverup.

Shrader-Frechette (1993), p.51, chastises hydrogeologists at Yucca Mountain for using versions of Darcy's law in their models, saying "This law, however, is an empirical, causal, and mathematical idealization", and brushes off suggestion that errors may be minor: "The combined effect of numerous value judgments and small errors might be great". Similarly, she says: "They are able, for example, merely to say that there is a 'high level of probability' that groundwater travel time to the water table will exceed ten thousand years. In other words, the degree of uncertainty regarding groundwater travel time is very great... Yet... some people appear to believe that Yucca Mountain will be predictably safe or in compliance with government regulations requiring a groundwater travel time greater than one thousand years" (p.165). Although apparently willing to make allowances for the ecologists' best efforts, Shrader-Frechette is astonished that "the DOE has affirmed that 'on the basis of the geologic record, no dissolution [of subsurface rock] is expected during the first 10,000 years or repository closure, or thereafter', and rejoins: "How could one affirm that will *never* be rock dissolution at Yucca Mountain?" (p.48). What are speared as "subjective" and indefensible "value judgments" that "exemplify bad science" (p.72) at Yucca Mountain become "informal logic" statements that "make sense" in the forests of Oregon. It is not hard to see that a sensitivity to indefensible value judgments really *is* key to these kinds of conversations.

56. Bredehoeft, J.D. and Konikow, L.F. (1993), p.179.

57. Popper, K.R. (1959), p.253.

58. Oreskes, N., Shrader-Frechette, K.S. and Belitz, K. (1994), "Verification, Validation, and Confirmation of Numerical Models in the Earth Science", in *Science*, **263**, 4 Feb 1994, pp.641-646.

59. Writing in 1962, Thomas Kuhn noted that "Few philosophers of science still seek absolute criteria for the verification of scientific theories. Noting that no theory can ever be exposed to all possible relevant tests, they ask not whether a theory has been verified but rather about its probability in the light of the evidence that actually exists" (p.145). The climate for verification has only deteriorated since Kuhn wrote.

60. This is the essence of Nancy Cartwright's argument in *How the Laws of Physics Lie* (1983), a topic that will be considered further in Chapter 6.

61. Oreskes, N., et al., (1994), p.641: "Numerical models may contain closed mathematical components that may be verifiable... However, the models that use these components are never closed systems. One reason they are never closed is that models require input parameters that are incompletely known. For example, hydrogeochemical models require distributed parameters such as hydraulic conductivity, porosity, storage coefficient, and dispersivity, which are always characterized by incomplete data sets. Geochemical models require thermodynamic and kinetic data that are incompletely or only approximately known. Incompleteness is also introduced when continuum theory is used to represent natural systems. Continuum mechanics necessarily entails a loss of information at the scale lower than the averaging scale... Finer scale structure and process are lost from consideration, a loss that is inherent in the continuum approach... Another problem arises from the scaling-up of non-additive properties... Another reason hydrological and geochemical models are never closed systems is that the observation and measurement of both independent and dependent variables are laden with inferences and assumptions. For example, a common assumption in many geochemical models of water-rock interaction is that observable mineral assemblages achieve equilibrium with a modeled fluid phase...[and] kinetic effects are assumed to be negligable". Note that all of this is of concern in the *application* of the model, as discussed in Chapter 2; it is completely irrelevant to assessing the *openness* of the model.

62. As discussed earlier, p.38, code validations probe the precision of numerical algorithms. We will refer to such tested codes as *numerically validated.*

63. Oreskes, N., et al., (1994), p.641: "Purely formal structures are verifiable because they can be proved by purely symbolic manipulations, and the meaning of these symbols is fixed and not contigent on empirically based input parameters". Once again, the meaning and value of these parameters within the model is not contigent upon estimates from the field. These values are arbitrarily chosen by the modeler; in short, they are fixed, as they must be in such a formal statement.

64. Note that this is very different from using Monte Carlo simulations to establish a mean property value, such as hydraulic conductivity, and then assuming this to be the actual value. Physical systems are unique, and the assumption that property values correspond to mean values is questionable. The probability of system responses, on the other hand, is a more meaningful concept, especially if we recognize that continuum assumptions apply; the interest is not in tagging molecules, generally, but predicting bulk distributions, etc.

65. Oreskes, N., *et al.*, (1994), p.644.

66. McMullin, E. (1978b), "The Conception of Science in Galileo's Work", in Butts, R.E. and Pitt, J. (eds.) (1978), *New Perspectives on Galileo*, D. Reidel Publishing Co., pp.209-257; McMullin, E. (1985), "Galilean Idealization", in *Studies in the History of the Philosophy of Science*, **16**:3, Pergamon Press Ltd., pp.247-273.

67. This is McMullin's (1985) perfect characterization; an "interlocutor" is "the man in the middle of the line in a minstrel show who questions the end men", Webster's 7th New Collegiate Dictionary.

68. Galileo, G. (1638), *Two New Sciences*, Drake, S. (trans., ed.), Wall and Thompson, Toronto, 2nd edition (1989), p.*xiii*.

69. An excellent discussion of Galileo's method of idealization and geometric proof can be found in McMullin, E. (1985), "Galilean Idealization", in *Studies in the History of the Philosophy of Science*, **16**:3, pp.247-273. Of particular interest for our purposes later on is McMullin's idea that equations become 'physicalized' over time, in so far as they become more accurate, incorporating both formal and material idealizations. McMullin (1985) elaborates on Galileo's response to

Simplicio: "Salviati's response to Simplicio is to point to an ambiguity in his objection. If he means that matter is such that when a sphere is realized in it, it may touch a plane at more than one point, this is demonstrably false. If on the other hand, he means that perfect spheres are never, in fact, realized in Nature, or as Simplicio puts it, that 'a metallic sphere being placed upon a plane, its own weight would press down so that the plane would yield somewhat', then this may well be true. But from this it does not follow that *if* such a sphere *were* to be realized in Nature, it would not have the properties that geometry demands of it. Matter cannot *alter* those properties; it merely makes them difficult to reproduce exactly" (p.250). The mathematization of the physical sciences is accepted as a commonplace and an inevitability today; the beginnings of hydrology as a quantitative science stem in large part from the adoption of differential equations and solutions developed for the flow of heat and electricity. Bacon underappreciated the power of mathematical insight, even in his own day. Stillman Drake comments: "Galileo never wrote an equation in his life, whereas we tend to think of physics mainly in terms of physical equations. The physical constants that loom so large in our thinking simply cancel out in proportionalities of the kind that Galileo used exclusively in his physics" (p.*xxx*). In commenting on Galileo's discovery of the pendulum law and the law of fall, Drake notes that it "exemplifies what can be learned by doing things step by step, in ratios and proportionalities without equations or any arbitrary constants depjendent upon the units adopted" (p.*xxix*). Nevertheless, Galileo's method was clearly more reliant on geometry than the norm in his time. Kuhn, T.S. (1962) also discusses Galileo's work in context, noting: "Galileo was not raised completely as an Aristotelian" (pp.118-125).

70. Konikow, L.F. and Bredehoeft, J.D (1992a), pp.81-82.

71. Oreskes, N., *et al.* (1994), p.644.

72. Konikow, L.F. and Bredehoeft, J.D. (1992a), p.75.

73. The situation is analogous to that described by Hacking, I. (1983), p.5: "Whenever we find two philosophers who line up exactly opposite on a series of half a dozen points, we know that in fact they agree about almost everything. They share an image of science, an image rejected by Kuhn. If two people genuinely disagreed about great issues, they would not find enough common ground to dispute specifics one by one".

74. National Academy Press (1990a), *Ground Water Models: Scientific and Regulatory Applications*, p.231.

75. The reader may wonder, at this point, if Popper would consider hydrologic models to be unscientific due to the difficulties in testing their predictions. Although we have left the discussion of demarcation far behind, it can be noted in passing that Popper (1959) was sensitive to the charge that difficulties in falsification might cast *most* scientific theories as "pseudo-scientific". He comments: "If we wish to avoid the Positivist's mistake of eliminating, by our choice of criterion of demarcation, the theoretical systems of natural science, then we must choose a criterion which allows us to admit to the domain of empirical science even statements which cannot be verified" (p.40). Popper expects a sort of Darwinian combat to lead to progress: "According to my proposal, what characterizes the empirical method is its manner of exposing to falsification, in every conceivable way, the system to be tested. Its aim is not to save the lives of untenable systems but, on the contrary, to select the one that is by comparison the fittest, by exposing them all to the fiercest struggle for survival" (p.42). His many critics are not convinced, in part because this winnowing struggle will leave us with the truth only if the true theory happens to be among the combatants, something that is by no means assured.

76. Personal communication, 31 Jan 1995.

77. Anderson, M.P. and Woessner, W.W. (1992b), p.168.

78. Oreskes, *et al.*, (1994), p.643: "To claim that a proposition (or model) is verified because empirical data match a predicted outcome is to commit the fallacy of affirming the consequent. If a model fails to reproduce observed data, then we know that the model is faulty in some way, but the reverse is never the case"; Klemes, V. (1982), "Empirical and Causal Models in Hydrology", in National Academy Press (1982), *Studies in Geophysics: Scientific Basis of Water-Resource Management*, pp.100-102, discusses why hydrologic models work (or seem to) on occasion, and enjoy a reputation for effectiveness in various places. His comments may be contrasted to the optimism prevalent among surface water modelers mentioned very early in this report (p.19). Klemes lists his reasons in order of decreasing frequency of occurrance (discussion deleted): "1) Model is empirical and works well as an interpolation formula; 2) Model works well because it portrays only a small and relatively well-understood segment (or component) of the hydrologic cycle; 3) Model is essentially untestable and its good performance is a matter of faith; 4) Results obtained by a good economic decision model reflect favorably on the quality of a hydrologic model embedded within it; 5) Model is largely irrelevant to results obtained with its aid; 6) Empirical model has a form that, without conscious effort of the modeler, happens to describe some essential aspect of the physical mechanism of the system; 7) Model's good reputation is based on superficial appearances; 8) Model works for the wrong reasons; 9) Model is deemed good by default; and 10) Model that does not work is not publicized".

79. Including the auxiliary hypotheses (AH) as part of the model is not meant to blur useful distinctions between the component parts of a model. For instance, Anderson, M.P. and Woessner, W.W. (1992b), "The Role of the Postaudit in Model Validation", in *Advances in Water Resources*, **15**, p.169, notes that model failure is generally due to either faulty conceptual models or to inaccurate values for future stresses (recharge, pumping, contaminant loading): "The more serious problem... is the first of these. It is unfortunate but true that there will always be errors in the conceptual model... While it is true that uncertainties involved in estimating future stresses are often large, these are less serious because the model could be rerun using accurate values for the stresses once they become known. Then if the conceptual model and the calibrated parameter values accurately represent system behavior, the validation would be successful".

80. Anderson, M.P. and Woessner, W.W. (1992b), p.172. See also Anderson, M.P. and Woessner, W.W. (1992a), pp.275-285, where the authors discuss the details of model documentation; and pp.286-294, where they review the four hydrologic post-audits reported in the hydrologic literature. They attribute the neglect of hundreds of other predictive models to the use of models in "crisis mode rather than a management mode" (p.293). Some of these studies are also reported in Konikow, L.F. (1986), "Predictive Accuracy of a Ground-Water Model - Lessons From a Postaudit", in *Ground Water*, **24**:2, pp.173-184; Konikow, L.F. and Person, M. (1985), "Assessment of Long-Term Salinity Changes in an Irrigated Stream-Aquifer System", in *Water Resources Research*, **21**:11, pp.1611-1624; Konikow, L.F. and Swain, L.A. (1990), "Assessment of Predictive Accuracy of a Model of Artificial Recharge Effects in the Upper Coachella Valley, California", in Simpson, E.S. and Sharp, J.M. (eds.), *Selected Papers on Hydrogeology*, 28th International Geological Congress, Washington D.C., 9-19 July 1989; Goode, D.J. and Konikow, L.F. (1990), "Reevaluation of Large-Scale Dispersivities for a Waste Chloride Plume: Effects of Transient Flow", in Kovar, K. (ed.), *ModelCARE 90: Calibration and Reliability in Groundwater Modeling*, The Hague, Netherlands, September 1990, IAHS Publication 195, pp.417-426.

81. By way of comparison, Glen, W. (1982), *The Road to Jaramillo*, Stanford University Press, pp.26, 46, 67, recounts the approach of the early investigators in Potassium-Argon dating of rocks. They refused for a time to even attempt to date rocks outside certain types and probable ages, until they had confidence in what amounted to a solid baseline for their evolving techniques and apparatus. Some similar discrimination might be appropriate on the part of hydrologists as to what is

"legitimate" to model.

82. Johnson, A.B. (1836), *Treatise on Language*, noted in Galileo, G. (1638), *Two New Sciences*, Wall and Thompson (1989), in Stillman Drake's introduction, p.*xxxv*.

83. Klemes, V. (1982), p.103.

84. Klemes, V. (1982), p.102. In the same volume, see also Dunne, T. (1982), "Models of Runoff Processes and Their Significance", pp.17-29. Dunne contrasts the 3-parameter empirical Horton infiltration equation to physically based expressions, beginning with that of Green and Ampt in 1911 (p.22). Similarly, "the depth, velocity, and discharge of Horton overland flow are calculated on the basis of equations that describe flow in a shallow channel. These equations involve statements of the conservation of mass and of momentum (the 'shallow-water equations') and a relationship, such as the Darcy-Weisbach equation, between the velocity and depth and slope of the water surface. These last three variables are related by an empirical coefficient... that ranges over two orders of magnitude or more along a hillside" (p.23). Also see the discussion of empirical versus causal models in Grayson, R.B., et al., (1994), "Physically Based Hydrologic Modeling, 1 and 2", in *Water Resources Research*, **26**:10, pp.2639-2666. The authors note the typical use of empirical fitting rather than physically based models for surface water; the former hope for reasonable correlation of observed and predicted events, while the latter are causal in sense of attempting to wire in PDEs of the actual driving forces.

85. Bacon, F. (1620b), *iv*, p.90.

86. Klemes, V. (1988b), "A Hydrological Perspective", in *Journal of Hydrology*, **100**, p.12. Konikow, L.F. (1986), "Predictive Accuracy of a Ground-Water Model - Lessons From a Postaudit", in *Ground Water*, **24**:2, pp.173-184, agrees: "Thus, although one advantage of deterministic models is that they represent processes and thus have cause-and-effect relationships built into them, careful attention must be paid to the accuracy with which future 'causes' (stresses) can be predicted (or estimated), because that can be the major source of error in the predictions of future 'effects' (system responses)" (p.183).

87. Bredehoeft, J.D. and Konikow, L.F. (1993b), "Reply to Comment", in *Advances in Water Resources*, **15**, pp.371-372: "An historical data set can also be matched and predictions made using a *black box* model. Such models, while they may be useful for decision making, usually yield little increased physical or chemical understanding of the ground-water system. On the other hand, structure imitating models often lead to increased understanding" (p.372). See Chapter 3, note 103.

88. Anderson, M.P. and Woessner, W.W. (1992b), p.172.

89. This is a very different reaction than that of Shrader-Frechette, who uses the anticipated inability of modelers at Yucca Mountain to meet regulatory requirements as another argument against the meaningfulness of their conclusions (see, *eg*., pp.52-3; 161-165). It also differs from the conculsion of Oreskes, Shrader-Frechette and Belitz (1994): "The burden is on the modeler to demonstrate the degree of correspondence between the model and the material world it seeks to represent..." (p.644).

90. Konikow, L.F. (1986), "Predictive Accuracy of a Ground-Water Model - Lessons From a Post-Audit", in *Ground Water*, **24**:2, March/April 1986, pp.173-184. Konikow continues: "Rather, they provide a means to quantitatively assess and assure the consistency within and between (1) concepts of the governing processes, and (2) data describing the relevant coefficients. In this manner, a model helps the investigators improve their understanding of the factors controlling ground-water flow" (p.183).

91. Konikow, L.F. and Bredehoeft, J.D.: "Groundwater Models Cannot Be Validated", pp.82-83. Compare Carr, E.H. (1961), pp.34-5: "As any working historian knows, if he stops to reflect what he is doing as he thinks and writes, the historian is engaged in a continuous process of moulding his facts to his interpretation and his interpretation to his facts. It is impossible to assign primacy to one over the other."

92. Oreskes, *et al.*, (1994), p.644, also note this use of models.

93. Anderson, Mary P.: "Ground-water Modeling - the Emperor Has No Clothes", in *Ground Water*, Vol.21, No.6, 1983, pp.666-669.

94. Chamberlin, T.C.: "The Method of Multiple Working Hypotheses", in *Science* (old series), **15**, 92 (1890), reprinted in *Science*, **148**, 7 May 1965, pp.754-759; and another paper of the same name but more focused on geoscience, 1892, reprinted in *Journal of Geology*, **39**, 1931, pp.155-165.

95. Bacon, F. (1620b), *vii*, p.93. Vulcan, Roman god of fire and metalworking; Minerva, goddess of wisdom.

Part Two:

Experiments of Light

We now come to ask whether and how far science can help us out of the obscure wood wherein philosophy has lost its way [concerning causation].

- Max Planck, *Where is Science Going?*

5

Ants, Bees and Spiders:

Darcy's Law and Structural Explanation[1]

The habit of precipitate explanation leads rapidly
on to the development of tentative theories.

- T.C. Chamberlin [2]

5.0 Process-Oriented Modeling Methodology

As hydrologists *pass from Vulcan to Minerva* - from practice to study - experimentation naturally emerges as a more important consideration. The emphasis on simple measurement that figures so prominently in applied groundwater modeling is correspondingly muted. Part Two of this study adds an inductive experimental perspective to the deduction from theory that dominated throughout Part One. A major goal is the elucidation of the nuances of experiment within hydrology, as both an independent undertaking and also in support of theoretical developments. Sometimes experiments are of broad interest and applicability; this is the case in our present project, an account of the discovery and historical treatments of Darcy's law. Other research programs are more highly specialized and of less general value; such is the case, for example, for research into fractured systems, to which Darcy's law of flow through porous media does not apply. In Chapter 6 we will examine contemporary issues involved in solute transport through fracture junctions. On all scales, a clearer picture will be developed of the mutually supporting roles of theory, model and experiment within an immature science. The need for this perspective stems in part from the unbalanced emphasis on theory within philosophy of science; for our purposes, philosophers dwell overly much on mature sciences. Models occupy the pivotal and mediating role in our account of process-oriented hydrology. Our principal interest remains how compelling accounts are constructed in the absence of strict validation. Process-oriented hydrology is evidently progressive, since understanding does increase; not only its results but also its methodology may therefore be of interest to applied modelers. We begin with an inter-disciplinary perspective.

According to a recent argument, models serve two essential functions in the argumentative structure (method) of solid state physics and chemistry.[3] On the one hand, models represent arguments from specific data to more general prescriptions; they are inductively constructed in the usual generalizing-upward concatenation of scientific experience. This is certainly not the "childishly simple enumeration" that Bacon rejected, but takes full advantage of his "proper rejections and exclusions; and then, after a sufficient number of negatives, comes to a conclusion on the affirmative instances".[4] On the other hand, models figure in deductive exercises that argue from fundamental laws to the proposed empirical relationships, *i.e.*, from the most general to the more specific. These inverted exercises are efforts to verify that our experience, experimental results and empirical conclusions "make sense", insofar as suggested relationships resting only on observation can be shown to be special cases of fundamental laws. These do not constitute validations of empirical theories, for reasons to be shown, but they can be qualitatively compelling nonetheless.

In the first case, inductive models aim at the description of a phenomenon and hence, for this purpose, models are the culmination of phenomenological description at some specified scale. The typical random variation in data requires that an empirical model be an idealized representation of the structures and processes thought to be involved in the behavior experimentally monitored and measured. The purpose is definitely not to account for each datum point, since:

> data provide evidence for the phenomena that theory explains. The full complexity of experimental data is never actually explained by theory. This is because data are idiosyncratic with respect to experimental equipment, materials and conditions - all factors left unaddressed by theoretical formalism. Furthermore, attempts to replicate experimental procedures result in data with values that vary over some range; the complexity of the causal factors in any interesting experiment make it

> impossible to explain the exact data resulting from any specific trial... Phenomena are different from data and have a stability that can be expressed through a variety of different types of data.[5]

Thus rather than accounting for *every* detail of Galileo's "impediments of matter", theorists take a somewhat more distant view, emphasizing phenomena rather than data. The construction of such descriptive and potentially explanatory phenomenological theories is not necessarily a simple business, however, as Per-Olov Lowdin, a leading quantum chemist, wrote in the first issue of the *International Journal of Quantum Chemistry* (1967):

> It is obvious that the use of experimental data... may provide valuable guidance for the development of the theory, but it is more seldom realized how *difficult* it is to utilize experimental information properly. The construction of meaningful semi-empirical theories is therefore one of the most important future goals for applied quantum chemistry, but in the meantime, the existing theories will have to live side-by-side.[6]

The success (predictive or otherwise) of theories built around experimental models is related to the strength of the analogy between prototype and model; on the other hand, the relevance or generality of models and theories depends on the nature and breadth of the experience from which they are derived. Lowdin suggests that decisive (crucial) events may be difficult to contrive.

Doubt about the strength of the prototype/model analogy combines with concern for theoretical consistency to motivate the second, deductive, use for models. For quantum physicists and chemists, the fundamental law of interest is the Schrödinger equation.[7] It is an important theoretical goal to show that their experimentally generated phenomenological laws can be approximately derived from an application of the Schrödinger equation to an appropriate model. Such a model will again necessarily result in an idealized and simplified mathematical description of the relevant system, a requirement that constrains both model choices and epistemological conclusions. Nevertheless, carefully chosen causal models are the means through which the Schrödinger equation gives a theoretical unity to what would otherwise be a disparate set of empirical laws and descriptions, each with limited scope. Failure in this endeavor might indicate that proposed theories developed from empirical observations are literally unfounded. In the same introductory issue of the *International Journal of Quantum Chemistry*, honorary editor W. Heitler commented on his approximate calculation in 1928 of the hydrogen bond length:

> The purpose was to understand the phenomena of chemistry and to reduce them to the laws of the newly created atomic physics. Numerical accuracy is less important for this aim than an overall understanding of the rules...[8]

Here the reductionist approach displays the process-oriented theorist's traditional indifference to "numerical accuracy". This is wholly appropriate inasmuch as the discussion turns on the development of higher order phenomenological laws.

In addition to serving as a sort of *post facto* reality-check, deductive arguments from fundamental principles serve an *a priori* purpose: deductions from the implications of first principles can guide research by suggesting model structures that might be experimentally testable. Thus besides arguing from selected data to general formulations, or from fundamental laws to experimentally supported correlations, investigators also proceed by arguing from general principles to expected behavior and needed experiments; this last reflects the sequence of events in our investigation of solute mixing at fracture junctions, to be considered in Chapter 6. J.C. Slater made a strong statement in support of this orientation, in direct contrast to Lowdin's empiricist prescription, again in the inaugural issue of the *International Journal of Quantum Chemistry*:

> ... the direction which I believe is most important, in the solid-state and molecular theory of the next few years, is an extension of the fundamental approach directly from fundamental principles, rather than a proliferation of semiempirical methods.[9]

The two functions of models (inductive and deductive) are not incompatible, but they do reflect two distinct orientations. Models may be either a summary of experience or figure in the derivation of expected behavior from theory. Debate over the general primacy of experiment or theory is as old as the scientific method itself. Francis Bacon commented on a less partisan stance in 1620:

> The men of the experiment are like the ant; they only collect and use; the reasoners resemble spiders, who make cobwebs of their own substance. But the bee takes a middle course; it gathers material from the flowers of the garden and the field, but transforms and digests it by a power of its own. Not unlike this is the true business of philosophy: for it neither relies solely or chiefly on the powers of the mind, not does it take the matter which it gathers from natural history and mechanical experiments and lay it up in the memory whole, as it finds it; but lays it up in the understanding altered and digested. Therefore from a closer and purer league between these two faculties, the experimental and the rational (such as never yet been made) much may be hoped.[10]

In his rebellion against the dominant Scholastic deduction of his day, Bacon unnecessarily deferred establishment of the most general laws (the conservation laws, for example) to the final stages of scientific understanding. He also underestimated the power of mathematical analysis. As a result, despite his conciliatory remarks, he tended to underrate the possibilities of causal arguments, even in concert with experimental evidence.

As a philosophical term, *scientific explanation* is most often taken to mean the explanation or justification of theories by reference to more fundamental rules. Fundamental laws are invoked to "explain" the apparent success of phenomenological laws through approximate derivations of the latter laws from the former. This procedure when successful demonstrates that no fundamental laws have been violated in the empirical statement; it also authorizes a measure of confidence in the proposed empirical relations by showing them to be special cases of more basic laws. At the same time, phenomenological laws certainly seek to summarize accumulating data, and on that basis are sometimes taken, rather crudely, to "explain" trends in the data. Explanation is thus a potentially ambiguous term, and some confusion can result from the two major functions of models - functions that are related but inverted.

All of the modeling operations described are common exercises in what has been called *internal* or *process-oriented* science, in eventual support (it is hoped) of *applied* science. As we turn our attention to the details of hydrologic process-oriented modeling, we will avoid the view that empirical constructions explain data. Rather, data will be simply viewed as evidence (in which explanatory causes are not addressed) for empirical relationships (which are the target of causal explanation).[11] As indicated earlier (pp.16-17) process-oriented theorists are more interested in *functions* that summarize *phenomena* than in *parameters* that merely reflect the details of case-specific *data*. Various details will be needed in the application of any discovered empirical laws, but they tend to be downplayed in establishing the general governing relationships. That the internal pursuits of theoreticians emphasize understanding generic processes through detailed examination of modeled relationships between possibly contributing factors is amply illustrated in Heitler's comment above (p.156).

Armed with these preliminary concepts, distinctions and cautions, the present paper illustrates the explanatory functions of hydrologic process models with reference to a specific phenomenological relationship. Darcy's law not only signaled the beginnings of hydrology as a quantitative science, it remains the centerpiece of applied hydrology today. In the remaining sections of this paper, the "dual functions of modeling as the culmination of phenomenology and the commencement of explanation"[12] will be taken up in turn. After an account of the origins and subsequent history of the law, a detailed analysis follows of its theoretical status in hydrology. A taxonomy of how models function with respect to Darcy's law offers insight into both the argumentative structure and the explanatory and descriptive goals of hydrology, particularly in the absence of a quantitative validation protocol. The different functions of models must be clearly distinguished to make sense of the resulting hierarchy of methods. In the specific case of Darcy's law, the theoretical derivations have all occurred after the experimental evidence was reported. There will, therefore, be little further discussion in this paper of derivations as a guide to research; the reader is again referred to the account in Chapter 6 of solute mixing at fracture junctions, in which this aspect of the argument is more fully discussed. The pivotal position of models in hydrologic research will now be illustrated, beginning with the historical discovery and subsequent empirical development of Darcy's law, and concluding with a discussion of theoretical derivations of the law from fundamental principles. Like Popper, we distinguish the methodologies of discovery and proof; unlike Popper, we find a conflation of the two not only possible but essential to the construction of compelling accounts.

5.1 Darcy's Law

In 1856 Henry-Philibert-Gaspard Darcy published a lengthy report on the upgrading of the public water system in the French city of Dijon.[13] Darcy designed an aquaduct and piping system to deliver water to Dijon from a spring some seven miles away. This work was mainly performed from 1835-1840, and had nothing to do with groundwater flow. In an appendix to his comprehensive report, Darcy included a description of much later experimental work on flow of water through sand filters, performed in the winter of 1855-1856. Characterizing it as "extremely rare", M. King Hubbert included Darcy's appendix (in the original French) in the 1969 edition of Hubbert's *Theory of Groundwater Motion*; portions of the appendix have since been translated into English by Allan Freeze. Darcy's data analysis resulted in an empirical relationship that has since come to be known as Darcy's law, and that in various forms remains basic to much hydrologic analysis. Although Darcy's law is a general law of flow through porous media, and hence of interest and use in many fields of science and engineering, our discussion will be confined to its use and status in hydrology.[14]

Darcy included a diagram and a description of his apparatus (see Figure 5.2). The sand constituting the filter was contained in a vertical tube said by Darcy to be 0.35 meters in diameter and 2.5 meters in length (although the diagram labels the length as 3.5 meters). Darcy performed a sieve analysis on his sand and estimated the porosity at 38%. Care was taken to minimize entrained air in the sand column, by first filling the column with water and then pouring and packing the sand, "so that the voids of the soil mass would contain no air". The height of sand could be varied above a screen and grillwork located 0.2 meters above the bottom of the column. Water entered the sand column from an adjacent hospital through a pipe near the top of the column and exited through a faucet mounted in the wall of the chamber below the grillwork supporting the sand. Both entry and output rates could be regulated, and both top and bottom compartments were vented to the air, "which is essential for the operation of the apparatus". To record pressures, two mercury manometers were installed in the chambers above and below the sand. For the purposes of these measurements, the bottom of the sand was taken as the datum plane with elevation zero.

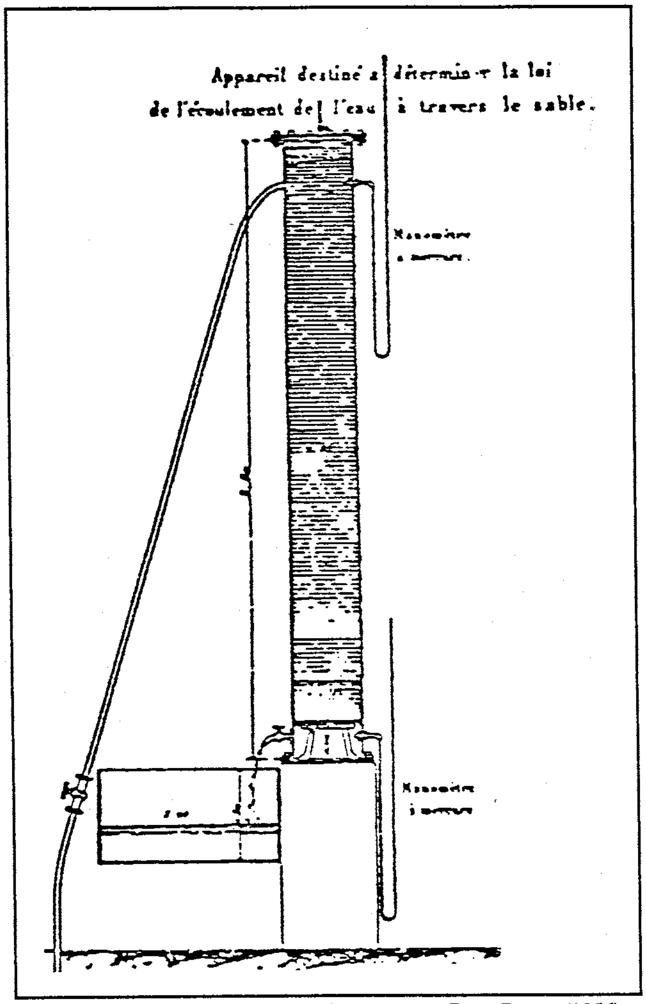


Figure 5.2: Darcy's illustration of his apparatus. From Darcy (1856).

In modern terminology, total hydraulic head is the sum of elevation head (the vertical distance between datum plane and the point of interest) and pressure head (expressed as the height of fluid registered by a manometer installed at the point of interest). Although Darcy did not explicitly explain all of his computational conventions, his usage was consistent given the specific circumstances of his apparatus. In particular, since Darcy chose the bottom of the sand as his datum plane, the elevation head at the bottom of the filter (sand) was zero. The bottom of the column was also open to the atmosphere; thus, using gage pressure conventions, the pressure head at the lower manometer position was zero. As a result, Darcy's calculations of hydraulic head at the upper end of his column amounted to adding the length of his sand column to the height of fluid (water equivalent) in the upper manometer arm.

Darcy initially tabulated four series of data for sand columns that were 0.58, 1.14, 1.71 and 1.70 meters high respectively. Water-hammer in the hospital plumbing forced him to use a mean value for the level of mercury in the upper manometer arm. He noted that "under weak heads, the almost total quiescence of the mercury in the manometer permitted measurement to the nearest millimeter". However, "when operated under strong pressures, the inflow tap was almost entirely open, and then the manometer, in spite of the diaphragm with which it was furnished, exhibited" strong random oscillations. Under high flow rate conditions, therefore, it was only "possible to measure the mean height of the mercury to the nearest 5 millimeters". In all cases, the upper manometer value was reported as the "mean pressure" for each experiment after he converted his mercury pressure readings to equivalent heights of water. The length of his reported measurement error bar then becomes 0.0262 meters of water in the "quiescent" intervals, increasing to 0.131 meters of water at maximum flow.[15]

Darcy performed as few as three and as many as ten different measurements of flow rates for his four heights and types of sand; in each case he gradually increased the height of mercury in the upper manometer arm (his mean pressure) by adjusting his inflow and outflow faucets. His records show each experiment took between 10 and 31 minutes. He was interested in saturated steady flow conditions; for each experiment, he used two initial observations to check that "the flow had become essentially uniform" before continuing. The sand was identical throughout each series, with the reported height of sand being "measured only at the end of each series of experiments after the passage of water had suitably packed it", a procedure that could result in a very slight shortening of his column over the course of a series. In his first four sets of measurements, the lower end of the column was open to atmospheric pressure. Darcy observed that, for any given elevation head, the outflow volume always increased with the pressure head. He evidently plotted the results, as he made this assertion on the basis of both "the table of experiments as well as their [unprovided] graphical representation".[16]

He then averaged the ratios of volumetric flow rate to hydraulic head (Darcy's *charge*) for each set of measurements, obtaining four proportionality constants, one for each series of experiments (hence, one for each sand). Darcy attributed the variation among the "constants" to differences in grain size and purity between the sands in different columns, noting that "the sand employed has not been consistently homogenized".[17] He also claimed without argument that the data showed that the flow rates varied in inverse proportion to length of sand column. This conclusion was not obvious since the data provided did not include multiple measurements at fixed heads for different column lengths; however, it is substantiated by comparison of his data for differing column lengths with roughly equal mean pressure values.[18] Darcy later performed a similar set of experiments differing mainly in that the pressure at the bottom of the column was adjusted widely above or below atmospheric pressure. He was satisfied that his earlier conclusions held in these cases as well, though it is not known how he lowered the pressures to the negative values reported; most likely a vacuum was applied to the lower column chamber.[19]

One point that stands out in this analysis is the clarity of the distinction between data and phenomena. As Jim Hofmann notes in his study of solid state physics and chemistry: "Data provide evidence for the phenomena that theory explains... Phenomena are different from data and have a stability that can be expressed through a variety of different types of data".[20] Darcy did describe the texture of his sand samples in some detail. However, aware of the unique nature of his apparatus and the variable effect of heterogeneities, he attributed little importance to the specific magnitudes of individual data readings. Rather, he emphasized two aspects of the general phenomenon he claimed his data supported - the proportionality of flow rate to total hydraulic head, and the inverse proportionality of flow rate to column length, *for identical sands*: "Il parait donc que, pour un sable de même nature, on peut admettre que le volume débité est proportionnel à la charge et en raison inverse de l'épaisseur de la couche traversée".[21]

Darcy assumed that flow is proportional to the cross-sectional area, and assembled his conclusions in the following equation:

$$q = k(s/e)(h + e \pm h^*)$$

where:

$q \equiv$ rate of water flow (volume per time);

$k \equiv$ a coefficient dependent on the "permeability" of the sand;

$s \equiv$ cross sectional area of the sand filter;

$e \equiv$ length of sand filter;

$h \equiv$ reading of the upper manometer arm, in equivalent height of water;

$h^* \equiv$ reading of the lower manometer arm, also in equivalent height of water.

At this point, Darcy made use of his datum plane convention, the direct implication of which is that $h^* = 0$. Only under this convention, Darcy's law reduces to:

$q = k\ (s/e)\ (h + e)$, for steady saturated flow.[22]

In modern format and notation, using a particular sign convention, Darcy's law is usually rendered as:

$$Q = -KA\ dh/dl$$

where:

$Q \equiv$ rate of water flow (volume per time);

$K \equiv$ hydraulic conductivity;

$A \equiv$ column cross sectional area;

$dh/dl \equiv$ hydraulic gradient, that is, the change in head over the length of interest.

The law is often transformed by dividing through by the cross-sectional area and is then restated as:

$$q = Q/A = -K\ dh/dl$$

where q (specific discharge) now has the dimensions of a velocity, and is sometimes referred to as the Darcy, or superficial, velocity; and K, the hydraulic conductivity, gives a measure of the ease with which flow occurs (necessarily in similar units of length per time). Modern usage thus formally distinguishes between hydraulic conductivity and permeability, with the former being a composite measure of the effects of matrix and fluid on flow, while the latter is a property of the solid matrix only. In practice, however, "permeability" is often used casually to denote "hydraulic conductivity".

Perhaps due to the ambiguous nature of some of Darcy's terminology and conventions, there was some initial confusion over the content of his law. Darcy's total head was essentially the sum of column length and water equivalent pressure head, and all experiments within a given series used the same column length. As a result, it is not surprising that Darcy's law might be taken by others to be a simple proportionality between flow rate and pressure head alone, neglecting elevation head. As late as the 1930s, M. King Hubbert was dismayed to find that widely used and respected texts in hydrology at least appeared to state Darcy's law in just this way. Beginning in 1940, Hubbert wrote several influential essays that eventually brought about a new consensus concerning the fact that hydraulic head functions as the potential in the law, and that total head is the sum of elevation head and pressure head.[23]

Although a full discussion of the historical development of Darcy's law is beyond the scope of this paper, one aspect that bears directly on present concerns is the recognition of both lower and upper bounds for the dependable use of the law's stated relationships. Briefly, some investigators consider a lower limit below which there is no flow in certain low permeability materials, positing the existence of a minimum threshold hydraulic gradient to motivate flow; this lower limit can affect the hydraulics of clays. The upper limit is of more general significance; the law has been found, once again experimentally, to be inappropriate when the flow regime is not both laminar and dominated by viscous forces. In laminar flow the velocity profile is smooth and fluid particles move along fixed streamlines with little mixing. The determination of laminar flow usually relies on calculating the magnitude of the Reynolds' number, a dimensionless ratio of inertial forces to viscous forces. At low Reynolds' numbers, viscous forces dominate, and Darcy's law is valid. There follows a transition zone in which inertial forces become more important and flowlines become distorted; Darcy's law cannot be accurately applied to the nonlinear laminar flow in this zone. Finally, turbulent flow is both nonlinear and non-laminar, and deviations from Darcy's law can become very large.[24]

Table 5.6: Calculated seepage velocities and Reynold's numbers for Darcy's First Series.

Experiment # (series:run)	Flow rate [L/min]	Q/nA [cm/s]	N_{Re}
1:1	3.60	0.1642	0.82
1:2	7.65	0.3487	1.75
1:3	12.00	0.5471	2.63
1:4	14.28	0.6511	3.26
1:5	15.20	0.6929	3.47
1:6	21.80	0.9937	4.97
1:7	23.41	1.067	5.34
1:8	24.50	1.117	5.58
1:9	27.80	1.267	6.34
1:10	29.40	2.632	6.71

Flow predictions based upon Darcy's law in the presence of large conduits in matrix material such as karstic limestones or highly fractured crystalline rocks can lead to large errors. Flow in such cases cannot be described adequately by a linear relationship such as Darcy's law. Darcy himself was not oblivious to the possibility of limits within which his law applied. In his discussion of Darcy's law, Hubbert says: "Darcy also noted, significantly, that the relationship was no longer valid for fluid velocities greater than 10-11 cm/sec",[25] a flow rate that in Darcy's column results in a Reynolds' number of about 50, a value well outside the strictly laminar range but still far below the turbulent range (> 300). Hubbert does not give the source of this comment. On closer inspection, however, this upper limit is actually for a similar empirical law in pipe flow.[26] Although perhaps thus pre-conditioned to look for such limits on the validity of his law for flow through porous media, Darcy found no evidence of limits within his experimental setup. Rough Reynolds' numbers have been calculated in Table 5.6 for Darcy's first series of experiments, assuming a mean grain diameter of 0.5mm, $\rho_{water} = 1000$ kg/m^3, and $\mu_{water} = 1$ cP $= 0.001$ kg/m·sec. The resulting Reynolds' numbers range from about 0.8 to about 6.7. In his 1940 paper, Hubbert puts the Reynolds' number where inertial forces become effective at 4; current practice puts it at about 1. [27]

In many aquifer materials, the assumption of laminar water flow may not cause significant loss of accuracy. Groundwater generally moves slowly. As noted in the discussion of applied groundwater models, hydrologists typically incorporate versions of Darcy's law into the governing equations for spatial variation of hydraulic head. These equations with accompanying initial and boundary conditions then become mathematical models applicable to a wide variety of specific sites. Throughout this procedure, Darcy's law is taken as an experimentally verified relationship, though the careful investigator will always bear in mind the empirical foundation and limited scope of the law. The Darcy velocity is a macroscopic parameter influenced by quite possibly unmeasurable microscopic factors.

One practical difficulty is that virtually any matrix considered, even on the laboratory scale, will be heterogeneous to some extent. Relationships have been developed for estimating the composite hydraulic conductivity for laminar flow through heterogeneous systems. However, field-scale investigations can fail to detect strata of significantly different conductivities. The presence of such heterogeneities does not theoretically prohibit the use of Darcy's law, but it lends uncertainty to generalizations on a large scale. It is for this reason that statistical models employing probabilistic parameter distributions are sometimes used to model the spatial variability of hydraulic conductivity. Of course, Darcy's own samples were heterogeneous to varying degrees. In this context, and with reference to Hubbert's work, it can be seen that Darcy's law remains tied to its empirical roots, and that modern questions on the details of its proper scope and application can be settled only with the help of a similarly experimental approach.[28] Some of these details are taken up in Chapter 6, in which the laws of flow through fractured systems are considered.

Whatever the initial confusion over exactly what the law entailed, there seems never to have been any doubt concerning the empirical basis of Darcy's law. In this respect it is somewhat surprising that in a recent discussion of Darcy's law, Kristin Shrader-Frechette repeatedly refers to it as either a "theoretical" or "fundamental" law.[29] She emphasizes the "idealized" nature of the law by claiming that it is "experimentally" verified only by applying the Bernoulli equation to an idealized model. She concludes that "this fundamental or theoretical 'verification' is highly idealized, since actual flow velocity is a function of the microstructure of the medium through which the water is flowing".[30] Similarly, she claims that:

> apart from the falsity of Darcy's Law on all three levels (micro, molecular, macro) what is significant is that 'corrections' to it do not come from the theory built into the law itself, but from phenomenological or observational factors not deducible from the theoretical or fundamental law.[31]

There are several misleading aspects to these comments. First, Darcy's law is a phenomenological law rather than a fundamental (universal) law. Such an assessment is fully borne out by a careful reading of Darcy's own account; suspicion that he may have relied on other empirical laws for pattern and guidance hardly makes his own law any less empirical. The experimental discovery of the law, as carried out by Darcy himself, is a matter of quantitative measurements with specific coarse sands and shows no reliance on fundamental laws or theory.[32] Hydrologists themselves routinely recognize this fact; for example, to quote an authoritative textbook by Freeze and Cherry, "Darcy's law is an empirical law. It rests only on experimental evidence".[33] Furthermore, intimations of limited accuracy and scope were recorded by Darcy in 1856, and have since become apparent to hydrologists; fundamental laws are not not so readily restricted. Although hydrologists do distinguish velocities on several different orders of magnitude, Darcy himself was interested only in macroscopic discharge and its proportionality to certain directly measurable properties of his filters. Apparent guesses from existing empirical laws hardly constitute causal derivations from fundamental laws. In fact, Darcy's work exemplifies a thoroughly Baconian approach, in that it is not the result of mere unguided observation nor was Darcy reliant in its *discovery* upon logical deductions from more fundamental laws. Any insights he gleaned from Fourier's or Ohm's work only illustrates the central role of *expectations* in Bacon's idea of thoughtful experiment.[34]

Darcy's law fully satisfies the requisite criteria to be considered a phenomenological law; it thus is a potential candidate for *structural explanation*, in which its subsequent derivation from fundamental laws gives a further measure of confidence in its validity. The difficulty and importance of delineating the valid application of Darcy's law at field scale were major issues in Shrader-Frechette's analysis of the Maxey Flats radioactive waste dump in Kentucky.[35] Nevertheless, it appears that in her subsequent more specific discussion of Darcy's law she overlooked its empirical basis; she then mistakenly interpreted as an "experimental verification" of a causal argument what was actually an example of structural explanation triggered by an experimentally suggested conceptual model. To clarify this distinction we must look more closely at structural explanation, first in general and then in hydrology.

5.2 Models, Phenomena and Structural Explanation

Discussion of theoretical explanation among philosophers of science has benefitted considerably from Bogen and Woodward's insistence that specific data are not the target of scientific explanation.[36] Data are too idiosyncratically dependent upon unique characteristics of specimens and instrumentation to warrant explicit explanatory attention. Rather, data provide evidence for the phenomenological relationships, conditions, or laws that are the potential subject matter for explanation. Earlier we referred to the unifying importance of efforts to draw empirical correlations under a larger explanatory umbrella of more fundamental laws.

Theorists traditionally pay particularly close attention to those phenomena that are at least tentatively organized under phenomenological laws. Nancy Cartwright's discussion of the relations between phenomenological laws and fundamental laws provides some useful analytic vocabulary and initial insight, particularly when amplified by the distinction between data and phenomena.[37] In *How the Laws of Physics Lie*, Cartwright argued that the explanation of phenomenological laws in physics typically requires the application of fundamental laws such as Schrödinger's equation to an appropriately "prepared" model of the domain. "Model" is construed by her for this purpose to indicate a specific geometric *structure* in which certain processes operate; the "preparation" mentioned typically consists of simplifying geometrical idealizations of that structure. This terminology and perspective accords with our own. Starting from this combination of a fundamental law and an idealized model, physical and mathematical approximations generate the derivations that Cartwright originally referred to as theoretical explanations.

The role of models in this account has been misunderstood in a manner that requires some clarification. Kroes and Sarlemijn[38] claim that the distinction between phenomenological laws and fundamental laws is not sufficiently precise. Using the example of the Van der Waals equation, they point out that the law is phenomenological in the sense that its two parameters must be specified experimentally. On the other hand, they also claim that the law could be considered to be fundamental because it is "derived from first principles".[39] But derivation from fundamental laws does not necessarily generate additional fundamental laws. More typically, the fundamental laws are applied to idealized models together with mathematical approximations. The results of these operations are precisely the phenomenological laws that Cartwright offers as our most reasonable candidates for laws of nature. Kroes and Sarlemijn also mistakenly disagree with Cartwright's provocative thesis that the fundamental laws of physics are false. Her point was that these laws say nothing specific about the real world until they are applied to suitably contrived models. Kroes and Sarlemijn claim that "Cartwright seems to confuse the validity of the boundary conditions with the validity of the fundamental laws".[40] A more correct paraphrase would emphasize that the approximate derivation of phenomenological laws by means of idealized models does not transform those phenomenological laws into fundamental laws; nor does an acknowledgment that models are highly idealized make fundamental laws, by contrast, "true" laws of nature.

Jim Hofmann has clarified the functions of models in derivations of phenomenological laws similar to the Van der Waals equation.[41] These correspond to the two principal uses of models we discussed at the outset. On the one hand, models are a culminating stage in the description of phenomena; as such, models stipulate and emphasize selected *structural aspects* of the domain. Secondly, however, models are *idealized* to provide this description in a mathematical form amenable to the application of a fundamental law. In this sense, models are a necessary requirement for what Cartwright called theoretical explanation in *How the Laws of Physics Lie*. In the terminology of that book, the two functions of models contribute to both phenomenological description and theoretical explanation.

Cartwright herself no longer considers the derivations she described in *How the Laws of Physics Lie* to constitute explanations. In a 1989 paper she decided to "reserve the word 'explanation' for scientific treatments that tell why phenomena occur".[42] Even when successful, black box models clearly do not satisfy this criterion. Cartwright attributes her revision to the fact that influential physicists such as Edwin Kemble hold that "the function of theoretical physics is to describe rather than to explain".[43] From this point of view, the derivation of phenomenological laws from fundamental laws is a demonstration that there is an economical way to classify these laws as various applications of a few fundamental principles; explanatory causes are not addressed, and the derivation cannot qualify as a strict validation of the empirical relationship. A careful distinction between explanation and description suggests P.B. Medawar may be skating onto some thin ice as he declares: "The activity that is characteristically scientific begins with an explanatory conjecture which at once becomes the subject of an energetic critical analysis".[44] Kemble's comments suggest that theoretical physicists, at least, may proceed from a *descriptive conjecture*, instead.

There are thus good reasons to redirect this discussion by emphasizing the more specific concept of *structural explanation*. One of the points that emerged from a study of transition metal oxide models[45] is that the wide variety of modeling techniques and motivations employed make it virtually impossible to neatly summarize the relationship between chemical phenomena and related models. But in other fields, surely, the structure stipulated by a model can sometimes be acknowledged to be part of an explanation of why the associated phenomena take place. This is in fact the interpretation Ernan McMullin refers to as structural explanation:

> When the properties or behavior of a complex entity are explained by alluding to the structure of that entity, the resultant explanation may be called a structural one. The term "structure" here refers to a set of constituent entities or processes and the relationships between them. Such explanations are causal, since the structure invoked to explain can also be called the cause of the feature being explained.[46]

Several points are in order here. First, note that "structure" has been broadly construed by McMullin as model geometry, processes and any interactions between them, *i.e.*, McMullin's structure is synonymous with our definition of a process-oriented model on any scale and for any purpose. Secondly, a model alone seldom provides structural explanation. That is, phenomena typically are explained by applying fundamental laws to the structure stipulated by the model. In this sense, the combination of model and fundamental principles constitutes a *theory* of the phenomenon. Thirdly, McMullin takes the success of structural explanations to warrant interpreting the relevant model as a more or less accurate description of *reality*. In particular, when revisions of a model provide increasingly accurate derivations of phenomenological laws, McMullin argues that it is justifiable to conclude that the real structure responsible for the phenomenon is approximately known. Progressive improvements rely less and less on the simple empirical correlations of black box models. Before considering how these ideas relate to process-oriented hydrology, note the following example McMullin cites:

> Geologists assume that a successful macrostructural explanation of such surface phenomena as sonar pulses can give reason to believe in the existence of sub-surface structures like pockets of water or oil. These structures play a role in the explanation of the phenomena similar to that played by molecular structures in the explanation of chemical phenomena. But in the geological case, the existence of the water or the oil can be directly ascertained. And the geologists' belief in the ontological reliability of the retroductive form has turned out to be amply justified.[47]

Although the structure cited in this example is a macroscopic one, theoretical hydrology also constructs conceptual models of the detailed and largely unobservable structure of the subsurface and its contained fluids at the microscopic scale. At all scales, models occupy a pivotal position in the reassuring deduction from fundamental laws. The attendant complications are best considered by taking up again the specific example of Darcy's law.

5.3 Structural Explanations of Darcy's Law

Process-oriented hydrologists draw several distinctions with respect to the models employed in their discipline. Electric analogue models have been wired to investigate conductivities and flowlines, using the parallel structures of Ohm's law and Darcy's law. Physical models such as sandboxes with particular packing patterns are still sometimes assembled in attempts to replicate specific aspects of conditions encountered in the field. Small-scale fieldwork may pursue highly specific questions under heavily instrumented conditions. The utility of any of these physical models can be limited if they are highly site specific; the generality of the empirical relationships observed might then be very narrow. Seminal experiments such as those of Darcy indicate that this is not always the case; controlled experiments can generate results amenable to generalization. Support for applied models - in the form of insight into processes - is generally best facilitated by avoiding overly specific experimental conditions. Process-oriented experimentalists therefore tend to emphasize generalizable experimental work.

Theoreticians interested in the corroboration of experimental indications, on the other hand, typically emphasize the use of tractable mathematical idealizations. Mathematical models in their turn can be either deterministic or statistical. As noted earlier, statistical models include parameters that have probabilistic distributions rather than single values. Deterministic models rely on relatively simple conceptual models to approximate the structure under study. In a survey of the subject, Faust and Mercer include the following description of how conceptual models are chosen to generate deterministic models:

> The first step is to understand the physical behavior of the system. Cause-effect relationships are determined and a conceptual model of how the system operates is formulated. For ground-water flow, these relationships are generally well known, and are expressed using concepts such as hydraulic gradient to indicate flow direction.[48]

There is a clear resonance here with McMullin's insistence that descriptions of causal relationships can act as a starting point for what ultimately become structural explanations.

Before turning to more specific discussion of conceptual models within structural explanations, another set of distinctions should be noted. Hydrologists follow the conventions of physics and thermodynamics, generally classifying stipulations of structure as falling within one of three possible viewpoints: molecular, microscopic, or macroscopic. The molecular approach is the most detailed in that it stipulates the path of individual molecules within the fluid in motion. The Lattice Gas Automata (LGA) numerical method described in Chapter 6 (p.216-217) is an example of this sort of statistical dynamics based on a loose analogy to molecular dynamics. Reducing the problem to the interactions of small particles eliminates most of the concern over model geometry, but simultaneously limits the scale of the problems that can be probed (although certain effects may be scalable). Since the ultimate interests of hydrology are invariably on a larger scale, as in the case of Darcy himself, hydrologists often move to a more coarse-grained approach in which the fluid within any particular pore of material is treated as a continuum rather than a collection of localized particles.

The resulting microscopic models then represent various constraints placed upon the idealized continuous fluid in the porous medium. For example, theoretical hydrologist Jacob Bear has collaborated with Bachmat in the invention and analysis of an elaborate microscopic model to represent fluid flow in a porous medium.[49] Bear describes the initial stage in this process as follows:

> In the present text we shall adopt the continuum approach. Accordingly, the actual multiphase porous medium is replaced by a fictitious continuum: a structureless substance, to any point of which we can assign kinematic and dynamic variables and parameters that are continuous functions of the spatial coordinates of the point and of the time.[50]

Similarly, Freeze and Cherry emphasize the far-reaching ramifications of structural stipulation, as they extend the discussion to macroscopic flow considerations:

> This... may appear innocuous, but it announces a decision of fundamental importance. When we decide to analyze groundwater flow with the Darcian approach, it means, in effect, that we are going to replace the actual ensemble of sand grains (or clay particles or rock fragments) that make up the porous medium by a representative continuum for which we can define macroscopic parameters, such as the hydraulic conductivity, and utilize macroscopic laws, such as Darcy's law, to provide macroscopically averaged descriptions of the microscopic behavior.[51]

Bear's work provides an excellent illustrative example of how a conceptual model becomes the basis for both a structural explanation of Darcy's law and also for the construction of a much more general mathematical model for fluid flow. Before considering Bear's discussion of Darcy's law, his general conception of his own reasoning process should be noted. The following is Bear's description of the two stages of his procedure that follow upon the introduction of a simplified conceptual model:

> Once the model is chosen, the second step is to analyze the model by available theoretical tools, and to derive mathematical relationships that describe the investigated phenomenon. These relationships show how the various active variables (fluxes, forces, etc.) depend on each other. They also show which factors have, according to the chosen model, no influence on the investigated problem. The only way to test the validity of laws derived in this way is to perform controlled experiments in the laboratory (or to observe phenomena in nature). Such controlled experiments, which comprise the third step of this approach, will test the validity of the derived relationships among the variables. No theory developed by this approach can be accepted without first being verified by experiments.[52]

Notice that Bear mentions that the goal of this procedure is to "describe the investigated phenomenon", a passage that calls to mind Cartwright's references to Kemble's conclusions about theoretical physics (p.168-9). We will see that Bear does not entirely avoid the term "explain", but he is reticent to use it because of the way models enter into his reasoning. Let us consider his derivations of Darcy's law.

In his most thorough treatment, he begins with the Bear-Bachmat conceptual model noted above.[53] The fluid is idealized as an incompressible continuum and the medium is imagined to be a network of interconnected passages and junctions within a solid that is rigid and does not interact with the fluid. Additional idealizations include the assumption that "the fluid loses energy only during passage through the narrow channels and not while passing from one channel to the next through a junction".[54] Bear then applies a complex averaging procedure in order to be able to assign values to dynamic variables within each representative elementary volume of the idealized continuous fluid. Finally, he applies an equation stating the conservation of linear momentum for a fluid system. The result is a general equation of motion which, when simplified for a homogeneous, incompressible fluid with small inertial forces, is an extension of Darcy's law to three dimensional flow in an anisotropic medium.[55] Consequently, it is not surprising that Bear sometimes says that the law simply expresses conservation of momentum during fluid flow through a porous medium, *i.e.*, Darcy's law is a special case of a more general and certain proposition.

Although this derivation is Bear's most sophisticated analysis of Darcy's law, he also provides a review of several derivations by other researchers in which the mathematics is simplified by assuming - at the outset - that the fluid is homogeneous. These derivations utilize a wide variety of different conceptual models: capillary tube models analyzed by means of the Hagen-Poisseuille law, fissure models, hydraulic radius models, and resistance to flow models. Finally, one of the best known derivations uses a statistical model to take into account the disorder of actual porous media prior to averaging the Navier-Stokes equations over a representative elementary volume.[56] *In each case a conceptual model stipulates an idealized geometry for the porous medium.* Principles of conservation of energy or momentum are then applied to the models and theoreticians arrive at versions of Darcy's law through a series of approximations and idealizations.

It should be clear that each of these derivations provides an example of what McMullin calls *structural explanations*. Darcy's law is a phenomenological law generated by experimental data. Darcy argued that the rate of water flow through samples similar to those he employed is proportional to the hydraulic head gradient. Theoreticians are interested to explain why this relationship seems to hold; they also want to explore its limitations. For this purpose, idealized structures are postulated in order to carry out mathematical applications of fundamental physical principles such as conservation of momentum. Darcy's law follows in due course, but only through a series of approximations that may include statistical analysis; it thus remains as much a phenomenological law as it was originally. Nevertheless, analysis of structures depicted by conceptual models has more or less brought the law under the explanatory umbrella of fundamental physical principles.[57]

Structural explanation is thus an accurate description of the derivations of Darcy's law carried out by process-oriented hydrologists. We should recall at this point McMullin's position that the usual research produces a sequence of increasingly accurate structural explanations that ultimately provide an realistic description of the causal components of the structure responsible for the phenomena explained.[58] His realist account suggests that improvements in the descriptive model converge eventually to an "approximately true" explanation, inasmuch as the described structure (model) is a major part of the explanation of why the phenomenon occurs. Description thus gives way to confident explanation. The history of causal models in physics and chemistry gives some encouragement to this expectation. The development of hydrologic models differs, however, from the scenarios emphasized by McMullin mainly in that there is not necessarily a progressive modification of a single model with increasingly accurate results.

The typical gap between prototype and model in hydrologic process models - on whatever scale - is evidence of two things: 1) the weakness of the structural explanations employed; and 2) the complexity of the systems of interest. Convergence to a single model typically occurs only when the investigated physical system is extremely simple, either in nature or by design. When this occurs (usually in strictly controlled environments - as in the solute mixing experiments described in Chapter 6), hydrology has effectively been reduced to one or more of the exact sciences. In such cases, whether in the field or in controlled experiments, model and prototype may be functionally indistinguishable. Strongly structural explanations naturally result in a reliably predictive science.

In the current state of affairs, it is much more likely that real hydrologic systems of interest will defeat this convergence. On both field and laboratory scale, simplified characterization of the prototype results in many process models adequately covering the available data. While these models remain structural explanations, the failure to converge means that they only "more or less" explain empirical results, and cannot function as strict validations of those results. A structural explanation is convincing only insofar as the structure stipulated reasonably describes that of the real system; judgment is still required as to system essentials. At larger scales or in more complex systems, hydrologists lack sufficient insight into either materials or mechanisms to build powerfully explanatory models that can support confident prediction. In fact, they generally do not even attempt to quantify the complete details of natural porous media, and thus have no basis for discussing the "true" nature of those media on these larger scales. McMullin's realist interpretation must generally be modified in the case of hydrologic models.

Instrumentalists (p.12) gain confidence in their understanding of systems, however, by the creative interplay of experiment and theory. Model uncertainty can be accomodated within a comparative approach to theory acceptance. Hydrologists may therefore share the instrumentalist attitude J. Hofmann attributes to the solid state chemists and physicists, "a significant number of [whom] would interpret an increasingly accurate series of approximate derivations of phenomenological laws as an indication of the utility rather than the truth of theoretical formalism".[59] Process hydrologists typically opt for the middle-ground orientation of the solid state chemists of the 1940s, who were faced with the same decisions regarding research orientation. As we have seen, there were those who championed one approach or the other, but most were content with hybrid methods that yielded "general explanations"; Nevill Mott explains:

> most of the research workers who are working on solids at the present time are not doing so with any hope of discovering laws of a fundamental kind... It seems to me more worth while to attempt to correlate observed phenomena and to give a general explanation in terms of atomic physics, than to attempt any sort of quantitative explanation of phenomena so complicated as those that occur in solids.[60]

Given the difficulties in successfully narrowing the field of models to a single, powerfully explanatory model, we might ask what further benefit process-oriented hydrologists find in causal derivations. We have already seen that Bear sometimes refers to the derivation of phenomenological laws as a particularly extended exercise in the *description* of relevant phenomena that does not necessarily result in an *explanation* of them, echoing the comments of the physicist Edwin Kimble. However, in other passages Bear uses explanatory language. For example, in referring to his set of derivations of Darcy's law, he makes this comment (emphasis added):

> In all these cases, the model is presented as an attempt to simulate, and *thus to explain*, phenomena observed in nature or in the laboratory. Sometimes several models are equally successful in explaining the relationship between observed excitations and responses. However, we must emphasize again that the proof of the validity of a model, and the only way to determine coefficients, is always the experiment.[61]

Bear's ambivalence concerning description *versus* explanation on this scale reflects the tentative nature of insights gained from multiple models. Nevertheless, the main thrust of his position is apparent in his account of the ultimate value of derivations based upon conceptual models:

> With these thoughts in mind, a question sometimes arises as to why we bother with the model in the first place, since in any case we must eventually go back to the laboratory to determine the required

> coefficients. The answer is that in applying the conceptual model approach we gain an understanding of the investigated phenomenon and the role of the various factors that affect it. We also gain an insight into the internal structure of the various coefficients appearing in the equations that describe the investigated phenomenon. All this information is needed for planning the laboratory experiments.[62]

Hydrologists may employ a multitude of different models simultaneously or sequentially to explicate various aspects of the parameters appearing in phenomenological laws. In the case of Darcy's law, for example, the hydraulic conductivity ultimately is not a "constant", but varies with both fluid and matrix type. How the value of this coefficient varies with physical conditions can be explored through a variety of causal models without claiming that any one of them provides a full account of the actual conduction process, even with future modifications in mind.

Analysis of the components of the conductivity constant in Darcy's law illustrates the "utility of theoretical formalisms" (p.175) as a guide to further experimental research. Darcy described the conductivity constant as primarily a function of grain size and sand purity. Modern versions of this coefficient have extended this description to include additional properties of the soil matrix, such as porosity and tortuosity, and also fluid properties, such as density and viscosity. These modern expressions do not always neglect interaction between the fluid and matrix, and are frequently varied in attempts to capture heterogeneous or anisotropic behavior. Bear notes that analysis of various models shows the importance of different factors. With experience, conceptual models are chosen to reflect the *expected* significance of various parameters, forces and relationships. Mathematical analysis of specific models then generates experimental tests of these hypothetically dominant parameters.[63]

For example, one of the structural explanations of Darcy's law reviewed by Jacob Bear is based on the capillary tube conceptual model.[64] In this model, the void space within the solid matrix of the porous medium is imagined as a collection of uniform, parallel cylindrical tubes of diameter δ and length dimension s. The areal porosity, n, is the percentage of void space in a cross sectional area taken normal to the tubes. The fluid density is ρ, and the dynamic viscosity is μ. The analysis is easily generalized to tubes that differ in their diameters or change along their length, but we will confine ourselves to the simple case of a uniform and constant δ everywhere. The fundamental law to be applied to this model is the relevant version of the conservation of momentum principle, namely, the Hagen-Poisseuille law. Given a hydraulic head of h and steady laminar flow of an incompressible fluid in a single, long cylindrical tube, this law states that:

$$Q = (\pi\delta^4\rho g/128\mu)\ \partial h/\partial s$$

where Q is the volume flow rate through the tube. Applying this equation to the capillary tube model and dividing through by the model's cross-sectional area gives:

$$q = (n\delta^2\rho g/32\mu)\ \partial h/\partial s \ = \ (k\rho g/\mu)\ \partial h/\partial s \ = \ K\ \partial h/\partial s,$$

where $k = n\delta^2/32$, and $K = k\rho g/\mu$.

This relationship is in fact Darcy's law where the hydraulic conductivity, K, depends on the model structure through its intrinsic permeability, k, with the latter stipulated in terms of porosity and tube diameter. The capillary tube model thus focuses insight on the dependency of hydraulic conductivity upon two specific properties of the medium. Subsequent experimental measurements of permeability or conductivity provide information about the corresponding behavior of the system. On the other hand, with increasing experience, the choice of an appropriate model is guided in part by decisions about what aspects of the medium are *expected* to have the major impact on permeability and therefore on conductivity.

At the same time, the relevance of these insights is limited to media that can be approximated fairly accurately by the capillary tube model. Likewise, how generally reassuring we find this derivation of Darcy's law depends on how well we think the capillary tube model represents any porous medium we know. "Tube diameter", after all, is an artificial property of the model more than a true property of the medium. Although McMullin's realist approach must be modified in hydrology, surely the confidence inspired by structural explanations remains tied to the strength of the model/prototype analogy. More sophisticated models are required to narrow the descriptive gap. Pursuit of these improved models is exactly the business of process-oriented research. Although Bear is referring here to the depiction of field-scale, site-specific phenomena, his remarks apply equally to the bench scale:

> The real system is very complicated and there is no need to elaborate on the need to simplify it... . Because the model is a simplified version of the real system, *there exists no unique model* for a given groundwater system. Different sets of simplifying assumptions will result in different models, each approximating the investigated groundwater system in a different way.[65]

5.4 Conclusion

Both experiments and research-guiding derivations may be undertaken to at least indirectly support applied groundwater models. The complications attending the move from process-oriented research to applied field models have been reviewed in previous chapters. Within that discussion of applied models, the availability of multiple, mutually inconsistent models presented several problems. All of the trouble was related to the practical necessity to *choose* between models that were different but were hardly differentiable by means of the usual logic of model testing. Even though applied models may normally encompass larger or more complicated systems, the problem is less one of scale than of purpose. Process-oriented researchers somewhat more distant from applied problems find a similarly uncertain situation to be at worst an experimental puzzle calling for the exercise of discriminating ingenuity, and at best another means of exploring complicated systems. Structural explanations serve important functions within such explorations; they may be undertaken either after or before the related experiments.

In the first case, given experimentally supported empirical relationships, fundamental laws - typically conservation of mass and momentum - may then be applied to a wide variety of models to provide approximate derivations of the phenomenological insights. Such exercises more or less demonstrate that suggested relationships are special cases of fundamental laws. The imprecision in these demonstrations is proportional to 1) the artificiality of the idealizations, assumptions and approximations required to carry out the derivation; and 2) the number of competing models available and adequate to complete a structural explanation. Depending upon their objectives, hydrologists have available to them a multitude of models that are potentially and simultaneously applicable to a given system. In contrast to most of the examples emphasized by McMullin, successive hydrological models do not necessarily improve in accuracy in all respects, but may instead emphasize different aspects of the system. For this reason structural explanations in hydrology can be compelling internally, but cannot be credited with the rigor of a strict validation protocol.

In the second case, derivations are undertaken before the experimental work, but they are not expected to stand alone. They point the way to experiments that subsequently serve as an essential corroboration of theoretically derived *hypotheses*. In combination with experiment, structural explanations not only provide an analysis of the factors relevant to the value of the parameters in phenomenological laws, but also help specify the limitations within which the laws remain accurate.[66] The usefulness of this combined approach will be further illustrated in Chapter 6, in connection with numerical representations of mechanisms governing solute mixing at fracture junctions.

Analysis of the argumentative form of structural explanations also calls to our attention the mutually supportive roles of fundamental physical laws and models in an applied science such as hydrology. Fundamental laws can only be brought to bear upon models that are in an appropriate mathematical form. The idealized conditions incorporated into a model represent assumptions that permit the explanatory derivation to be carried out. Consequently, a statement of these conditions facilitates the empirical clarification of the domain in which phenomenological laws are applicable. Since hydrology has as its domain such a multitude of disparate individual systems, models function as an important scheme to classify these systems. Unless a system can be accurately represented by at least one model that functions in a structural explanation of Darcy's law, for example, there is good reason to doubt that the law can be successfully applied to that system. At the same time, it is obvious that not all structural explanations are equally compelling. As Wilfred Sellars pointed out long ago, an important characteristic of scientific explanation is the understanding it provides concerning why the phenomenological laws to be explained are in fact only approximately correct under limited circumstances.

5.5 Notes:

1. Substantial portions of this chapter first appeared as a co-authored paper: Hofmann, J.R. and Hofmann, P.A. (1992), "Darcy's Law and Structural Explanation in Hydrology", in Hull, D., *et al.* (eds.), *PSA 92*, biennial Proceedings of the Philosophy of Science Association, pp.23-35.

2. Chamberlin, T.C. (1892), "The Method of Multiple Working Hypotheses", reprinted in *Science*, 7 May 1965, pp.754-9.

3. Hofmann, J. (1990): "How the Models of Chemistry Vie", in Fine, A., *et al.* (eds.): *PSA 1990*, Vol.1, Proceedings of the Philosophy of Science Association, pp.405-419.

4. Bacon, F. (1620b), *Novum Organum*, in Burtt, E.A. (ed.) (1939), *The English Philosophers from Bacon to Mill*, The Modern Library, *cv*, p.71.

5. Hofmann, J.R. (1990).

6. Lowdin, P-O (1967), "Nature of Quantum Chemistry", in the *International Journal of Quantum Chemistry*, **1**, pp.7-12.

7. The wave function Ψ in Schrödinger's wave equation

$$\nabla\cdot\nabla\psi + 8\pi^2 m \frac{E-\Phi}{h^2}\psi = 0$$

describes the motion of an electron of mass m under the influence of an electrical potential Φ, where the kinetic energy of the particle, E, is given by $E = \frac{1}{2}mv^2 = \frac{1}{2}m\,(h/m\lambda)^2$.

8. Heitler, W. (1967), "Quantum Chemistry: The Early Period", in the *International Journal of Quantum Chemistry*, **1**:1, pp.13-36.

9. Slater J.C. (1967), "The Current State of Solid-State and Molecular Theory", in the *International Journal of Quantum Chemistry*, **1**, pp.37-102. The debate between Lowdin and Slater is a reprise of that somewhat starker and much earlier one between Humphrey Davy (1778-1829) who claimed: "The foundations of chemical philosophy, are observation, experiment, and analogy [to]... general scientific truth"; and Justus von Liebig (1803-1873), who claimed: "In science all investigation is deductive or *a priori*. Experiment is only an aid to thought, like a calculation: the thought must always and necessarily precede it if it is to have any meaning". Hacking, I. (1983), *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*, Cambridge University Press, pp.152-154, provides extended quotes in the course of discussing these contrasting views at greater length. Kuhn, T.S. (1962), *The Structure of Scientific Revolutions*, University of Chicago Press, p.29: "...since Galileo, such laws have often been correctly guessed with the aid a [research] paradigm years before apparatus could be designed for their experimental determination".

10. Bacon, F. (1620b), *xcv*, p.67.

11. This view has been challenged (ineffectively, we think) by Rothbart, D. and Slayden, S.W. (1994), "The Epistemology of a Spectrometer", in *Philosophy of Science*, **61**, pp.25-38: "If phenomena have recurring features produced regularly by some small set of factors, as Woodward states, then the notion of phenomena is that of an organized type that is instantiated by specific specimens (tokens) under scrutiny. If data assume inherently singular instances of experimental environments, then data are obviously tokens of some pattern (type). But the claim that phenomena and not data are candidates for theoretical explanation is trivialized by the contrast between phenomena as types

and data as tokens. Patterned data, such as data structures, *are* subject to theoretical explanation" (p.35). The circularity of this argument collapses completely if we recognize "patterned data" as a useful definition of "phenomena".

12. Hofmann, J. (1990), p.406.

13. Darcy, H. (1856): *Les Fontaines Publiques de la Ville de Dijon*, Paris: Victor Dalmont.

14. Freeze, R.A. and Back, W. (eds.) (1983): *Physical Hydrogeology*, Hutchinson Ross, pp.14-20; all uncited quotations from Darcy are paginated from this edition of the appendix; quotations from Darcy are therefore from Allan Freeze. There is some disagreement over Darcy's motivation for his experiments. Muskat, M. (1937), *Flow of Homogeneous Fluids Through Porous Media*, McGraw-Hill, says simply, "[Darcy] was interested in the flow characteristics of sand filters" (p.55). Hubbert, M.K. (1969), *The Theory of Groundwater Motion and Related Papers*, refers to "a problem encountered by Darcy in designing a suitable filter for the system [of Dijon]. Darcy needed to know how large a filter would be required for a given quantity of water per day and, unable to find the desired information in the published literature, he proceeded to obtain it experimentally" (pp.266-267). Freeze's translation of Darcy's report on his experiments begins: "I approach now an account of the experiments that I have carried out at Dijon together with Engineer Charles Ritter, to determine the laws of flow of water through sand". In their introductory comments, Freeze and Back (1983, p.10) say "It is clear from the main body of the report that Darcy's experiments were aimed at improving the design of filter sands for water purification". Writing some ten years later, Freeze (1994), "Henry Darcy and the Fountains of Dijon", *Ground Water*, **32**:1, p.26, notes that "the bulk of the work" on the Dijon water system that is the main subject of the report was done much earlier, from 1835-40. The Dijon system was also an aquaduct and piping delivery system from a distant spring, not a pumped groundwater supply. These facts cast some doubt on the interpretation of Watson, I. and Burnett, A.D. (1993), *Hydrology: An Environmental Approach*, Buchanan Books, who quote an uncited passage from Darcy: "I have attempted by precise experiments to determine the flow of water through aquifers". These authors also say: "The proposed well field [for Dijon] was to be in a predominantly sandy aquifer. As a follow-up to his field investigations, Darcy conducted a number of laboratory experiments that were aimed at a generic simulation of the field conditions shown [in the figure below]" (p.69). The caption to the included figure reads: "Field Conditions Simulated by Darcy's Experiment. Based on the conceptual model of flow through an imaginary cylinder in the field, Darcy designed a laboratory experiment..."

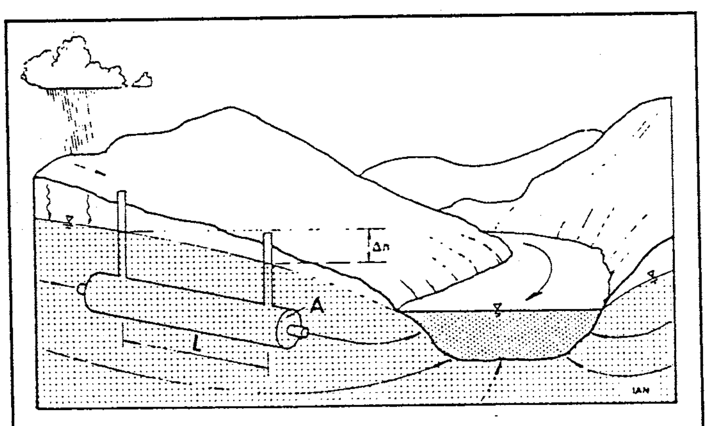


Figure 5.1: Darcy's conceptual model? From Watson and Burnett (1993), p.69.

15. Darcy states that a millimeter of mercury corresponds to 26.2mm of error in terms of water height. At atmospheric pressure, 1 atm = 760mm Hg = 10.33m H_2O; then 1mm Hg = 0.01359m H_2O. Since Darcy's data is "within one millimeter", and fluctuates above and below his mean pressure, the total error *at sea level* would double to 0.0272 meters of water. Darcy reports this number as 26.2mm, which suggests he corrected for the elevation of Dijon. Standard atmospheric pressure formulae (CRC) that are appropriate for elevations between -1,524 and +11,000 meters above sea level give:

$T \equiv$ Temp (K)	$= 288.15 - 0.0065H;$	(1)
$P \equiv$ pressure [mb]	$= 1013.25\,(288.15/T)^{5.255877}$; or	(2)
	$= 1013.25\,[288.15\,/\,(288.15 - 0.0065H)]^{5.255877}$	(3)

where $H \equiv$ height [m]. Darcy's conversion factor gives an atmospheric pressure of (26.2/27.2)atm. or 975.76 mB. Inserting this value into Eqn 3 then gives an elevation of about 315 meters at Dijon, a figure that agrees with published geographical data within the normal variability of atmospheric conditions.

Similarly, with his taps wide open, Darcy's estimate of the error is 5 times as great. The height of mercury then fluctuated over a range equivalent to 0.131m of H_2O, a figure his report mistakenly lists as 1.30m H_2O.

16. Table 5.1: Translated reproduction of Darcy's table of experiments carried out at Dijon on 29-30 October and 2 November, 1855. Adapted from Freeze and Back (1983), p.18.

Number of the Experiment	Duration (min)	Mean Flow Rate (L/min)	Mean Pressure	Ratio Between Volumes and Pressures	Observations
1st Series, with a sand thickness of 0.58 m.					
1	25	3.60	1.11	3.25	The sand has not been washed.
2	20	7.65	2.36	3.24	
3	15	12.00	4.00	3.00	The manometer column experienced only slight movements.
4	18	14.28	4.90	2.91	
5	17	15.20	5.02	3.03	
6	17	21.80	7.63	2.86	
7	11	23.41	8.13	2.88	Very appreciable oscillations.
8	15	24.50	8.58	2.85	
9	13	27.80	9.86	2.82	Strong manometer oscillations.
10	10	29.40	10.89	2.70	
2nd Series, with a sand thickness of 1.14 m.					
1	30	2.66	2.60	1.01	The sand is not washed.
2	21	4.28	4.70	0.91	
3	26	6.26	7.71	0.81	
4	18	8.60	10.34	0.83	Very strong oscillations.
5	10	8.90	10.75	0.83	
6	24	10.40	12.34	0.84	
3rd Series, with a sand thickness of 1.71 m.					
1	31	2.13	2.57	0.83	Washed sand.
2	20	3.90	5.09	0.77	
3	17	7.25	9.46	0.76	Very strong oscillations.
4	20	8.55	12.35	0.69	
4th Series, with a sand thickness of 1.70 m.					
1	20	5.25	6.98	0.75	Washed sand with a slighter coarser grain size than the preceding.
2	20	7.00	9.95	0.70	Weak oscillations as a result of the partial obstruction of the manometer opening.
3	20	10.30	13.93	0.74	

17. Darcy reports the texture as follows: "58% sand passing the 0.77mm screen; 13% sand passing the 1.10mm screen; 12% sand passing the 2.0mm screen; 17% fine gravel and shell fragments". Later he notes that the sand in the 4th Series has a "somewhat coarser grain size than the preceding", but gives no more specific description.

18. Darcy's first conclusion, that discharge through each filter [the sand column] increases proportionally with the head, is confirmed by a simple comparison of the third and fourth columns of his table. See Table 5.1, note 14, above.

After announcing this result, Darcy calculates the proportionality constant for each series in two different ways. Without elaboration, he says

> the discharge per second per square meter is related very roughly to the head by the following relations:
>
> | | |
> |---|---|
> | First series: | Q = 0.493 P |
> | Second series: | Q = 0.145 P |
> | Third series: | Q = 0.126 P |
> | Fourth series: | Q = 0.123 P |
>
> Denoting I as the head per meter of thickness of the filter, these formulae become the following:
>
> | | |
> |---|---|
> | First series: | Q = 0.286 I |
> | Second series: | Q = 0.165 I |
> | Third series: | Q = 0.216 I |
> | Fourth series: | Q = 0.332 I |

Although no details are given, Darcy's constants for the first series are apparently arrived at as follows: The length, L, is 0.58 m, while the number of experiments, n, is 10.

Then the mean flow, $\bar{Q}$, is $\Sigma Q_i/n$, i = 1,2,3,...,n;
With appropriate unit conversions, 179.64/10 [Liters/min] = $\bar{Q} = 0.2994$ [L/sec].

Then $\bar{Q}$ per unit area, $\bar{Q}/A$, is 0.2994 [L/sec] / $\pi(0.35\text{m}/2)^2$. $\bar{Q}/A = 3.11$ [L/sec·m²].
This is Darcy's "discharge per second per square meter".

If we put $\bar{Q}/A = \bar{C}_1\,\bar{P}$, where $\bar{P} = \Sigma P_i/n = 62.48/10$ [m], $\bar{C}_1 = 0.498$ [sec^{-1}].

As shown above, Darcy has $\bar{C}_1 = 0.493$ for the first series, the first of several arithmetic errors. As indicated above, he then puts I as the head per meter of thickness, P/L, and $\bar{Q}/A = \bar{C}_2 I$, and thus:

$\bar{Q}/A = \bar{C}_2\,(\bar{P}/L)$. Using Darcy's result for $\bar{C}_1$, $\bar{C}_2 = 0.493(0.58) = 0.286$ [m/sec].

The units of $\bar{C}_2$ are more generically given by

$$[\bar{C}_2] = \left[Q\cdot L \,/\, A\cdot(\text{meters of P})\right]$$
$$= [l^3/t]\,[l] \,/\, [l^2]\,[l]$$
$$= [l/t], \text{ the Darcy velocity.}$$

The "roughness" of the fit is due to the averaging involved over a wide range of pressures. The sand was the same for the ten trials (if one neglects effects of washing) but variable water hammer - increasing with increased flow, according to Darcy - lent some uncertainty to the pressure readings. $\bar{C}_1$ and $\bar{C}_2$ are, of course, average proportionality constants for the first series. Carrying out the calculations for C_1 and C_2 for each experiment in Darcy's favored units gives some information that Darcy did not provide:

Table 5.2: Proportionality constants C_1 and C_2 calculated for *individual* experiments of Darcy's First Series.

number of the experi-ment	Mean flow rate [L/min]	Mean flow rate [L/sec]	mean pressure [meters H_2O]	C_1 [per sec]	C_2 [m/sec]	Darcy's ratio of flow volume to pressure
1	3.60	0.06	1.11	0.562	0.326	3.25
2	7.65	0.128	2.36	0.564	0.327	3.24
3	12.00	0.20	4.00	0.520	0.302	3.00
4	14.28	0.238	4.90	0.505	0.293	2.91
5	15.20	0.253	5.02	0.524	0.304	3.03
6	21.80	0.363	7.63	0.494	0.287	2.86
7	23.41	0.390	8.13	0.499	0.289	2.88
8	24.50	0.408	8.58	0.494	0.287	2.85
9	27.80	0.463	9.86	0.488	0.283	2.82
10	29.40	0.490	10.89	0.468	0.271	2.70

This is the source of our comment that Darcy attributed little importance to the specific *data* values, emphasizing instead two aspects of a general *phenomena* that his data at least generally supported. In fact, the individual values of C_1, C_2 and Darcy's own ratio of flow volume to pressure suggest some non-linear behavior, at least if we take all the reported data at face value. Darcy evidently did not, to judge from his conclusions, being content with a "very rough" first approximation of the relationship.

The values Darcy listed for the Second, Third and Fourth (see Table 5.3) series are arrived at in a similar way. The third experiment of the second series, and the third experiment of the Fourth series also gave results that might seem to call for further explanation in support Darcy's first conclusion. Many hydrologists who have done versions of these same experiments will no doubt shrug off Darcy's inconsistent results as nothing more than measurement error. But this only reinforces the data/phenomena distinction, as Darcy was apparently unfazed by imperfect support for his linear hypothesis/conclusion. The admitted roughness of the fit casts doubt on the advisability of applying too fine a statistical analysis to Darcy's work, as in Davis, P.A., Olague, N.E., and Goodrich, M.T. (1992), "Application of a Validation Strategy to Darcy's Experiment", in *Advances in Water Resources*, **15** (1992), pp.175-180. This paper purports to show that Darcy's law is not confirmed by Darcy's own data, failing to appreciate the data/phenomena difference. It is more likely that their investigation has shown that their validation scheme will be of no use at Yucca Mountain, than that it mounts a serious challenge to Darcy's law under the usual laboratory circumstances. It is, however, of some historical interest that even as Darcy reduced the number of experiments in each series, the number and magnitude of his arithmetic errors increased.

Table 5.3: Calculations of Darcy's progressive errors in figuring his proportionality constants $\bar{C}_1$ and $\bar{C}_2$.

Series Number	C_1 (actual/reported)	C_1 error	C_2 (actual/reported)	C_2 error
1	0.498 / 0.493	0.96%	0.289 / 0.286	1.04%
2	0.147 / 0.145	1.36%	0.168 / 0.165	1.79%
3	0.128 / 0.126	1.56%	0.219 / 0.216	1.37%
4	0.127 / 0.123	3.15%	0.215 / 0.322	50.00%

Allan Freeze points out that the corrected value for Darcy's final and most significant error "is actually in better agreement with the other three sets of runs than the original [value reported]" (Freeze, A. (1994), p.25). The errors did not, however, deter Darcy from certain claims.

More telling than this trend is the fact that Darcy did not perform controlled experiments that directly supported his second conclusion - that the flow rate was inversely proportional to the thickness of the sand pack - even though such a conclusion is not inconsistent with his results. Although he did not perform a set of experiments with a constant head and differing column lengths, a comparison of selected experiments in Table 5.3 above gives some rough confirmation across heterogeneous samples: as the ratio of lengths increases in column 3, the ratio of flow to length in column 5 decreases, while when the ratio of lengths are virtually the same as in the last two cases, the ratios of flow to length are also nearly identical.

Table 5.4: Manipulations of Darcy's data to confirm the inverse relation of flow to column length: the second column shows data for experiments with roughly equal pressures but differing flowrates. The length of the column for series 1 was reported as 0.58m; for series 2, 1.14m; for series 3, 1.71m; and for series 4, 1.70m. As the ratio of lengths in column 3 increases, the ratio of flows in column 4 decreases.

experiments compared (series:#)	pressure @ flowrate for each experiment	ratio of lengths (longer/shorter)	ratio of flows	ratio of flows over ratio of lengths
2:6 3:4	12.34 @ 10.40 12.35 @ 8.55	1.71/1.14 = 1.5	8.55/10.4 = 0.82	0.064
1:6 2:3	7.63 @ 21.80 7.71 @ 6.26	1.14/0.58 = 1.97	6.26/21.8 = 0.29	0.146
1:9 4:2	9.86 @ 27.80 9.95 @ 7.00	1.70/0.58 = 2.93	7.00/27.8 = 0.25	0.086
1:5 3:2	5.02 @ 15.20 5.09 @ 3.90	1.71/0.58 = 2.95	3.90/15.2 = 0.26	0.087

It may not be entirely coincidental that Darcy's column lengths were very nearly in integer multiples of 0.58m. Perhaps this was intended to make the effect of column length apparent. Certainly across a roughly similar range of pressures, his ratio of volumes to pressures (see Table 5.3, column 5) decreases markedly from Series one (length equal 0.58 m) to Series 2 (length equal 1.14 m) to Series 3 (length equal 1.71 m). There is not much change from Series 3 to 4, where the lengths are virtually equal, and the only difference is in the coarseness of the sand.

Finally, Darcy performed no experiments whatsoever to support his assumption that flow rate would be proportional to the cross-sectional area of the column (all his experiments were performed on the same column). This fact, coupled with the casualness of the calculations, the very indirectly confirmed conclusion that flow is inversely proportional to column length, and even the declining number of experiments in each subsequent series, suggests that Darcy may have guessed his conclusions *a priori* by analogy to the other empirical gradient laws already established, including those of Ohm and Fourier. Hubbert said in 1956 that "it has subsequently come to be universally acknowledged that Darcy's law plays the same role in the theory of the conduction of fluids through porous solids as Ohm's law in the conduction of electricity, or of Fourier's law in the conduction of heat" (p.270). Ohm's law (1827), for example, has

$i = v/r$, where i = the current
v = the voltage drop (the potential)
r = the resistance;

but $r = \rho L/A$, where
ρ = the resistivity
L = the length of interest
A = the cross-sectional area

Thus $i = (1/\rho)A(v/L)$, an exact parallel to Darcy's law, in which resistivity goes as one over the conductivity. Ohm himself is said to have been steered to his discovery by analogy to Fourier's law of heat conduction (1822). Such conjectures are tangential to the goals of the present paper, but the facts presented do strongly suggest that Darcy's experiments were suggested and directed by prior exposure to potential theories; as a well-educated and even eminent general engineer, Darcy could hardly have remained unaware of related developments. (He graduated high in his class at L'Ecole des Ponts et Chausseés in 1826 and went on to a very distinguished career punctuated by his unanimous election in 1857 to the chair previously held by the mathematician Augustin-Louis Cauchy in the French Academy of Sciences). The facts also demonstrate that he collected only as much data as he felt necessary for his engineering purpose.

19. Table 5.5: Translated reproduction of Darcy's table of results from experiments carried out February 17-18, 1856. Pressures at the bottom of the column varied above and below atmospheric pressure. The length of the sand column was 1.10 m. Adapted from Freeze and Back (1983), p.19.

number of the experi-ment	Duration	Mean flow rate	Mean pressure		Differ-ence in pressures	Ratio between volumes and pressures	Observations
			Above the filter	Below the filter			
1	2	3	4	5	6	7	8
	[min]	[L/min]	[m]	[m]	[m]		
1	15	18.8	P + 9.48	P - 3.60	13.08	1.44	Strong oscillations in the upper manometer
2	15	18.3	P + 12.88	P 0	12.88	1.42	"
3	10	18.0	P + 9.80	P - 2.78	12.58	1.43	"
4	10	17.4	P + 12.87	P + 0.46	12.41	1.40	Weak
5	20	18.1	P + 12.80	P + 0.49	12.35	1.47	Rather weak
6	16	14.9	P + 8.86	P - 0.83	9.69	1.54	Almost none
7	15	12.1	P + 12.84	P + 4.40	8.44	1.43	Very strong
8	15	9.8	P + 6.71	P 0	6.71	1.46	Very weak
9	20	7.9	P + 12.81	P + 7.03	5.78	1.37	Very strong
10	20	8.65	P + 5.58	P 0	5.58	1.55	Almost none
11	20	4.5	P + 2.98	P 0	2.98	1.51	"
12	20	4.15	P + 12.86	P + 9.88	2.98	1.39	Quite strong. The cause of these oscillations was explained by this time

20. Hofmann, J.R. (1990).

21. Darcy (1856), p.593.

22. The graphical representation of Darcy's data could have been simply flow v. head to demonstrate one of his conclusions, or it could have been flow v. hydraulic gradient to demonstrate the composite determining factors. In the latter case, Q = k(s/e)(h+e) = ks*i*, where *i* is the hydraulic gradient. Then k(s/e)(h+e) = ks*i*, or *i* = [(h/e)+1]. As would be expected from the dependence

of gradient on head for a given series, the two plots are nearly identical but for slope. Similar results can be obtained for the other series, again with differing slopes.

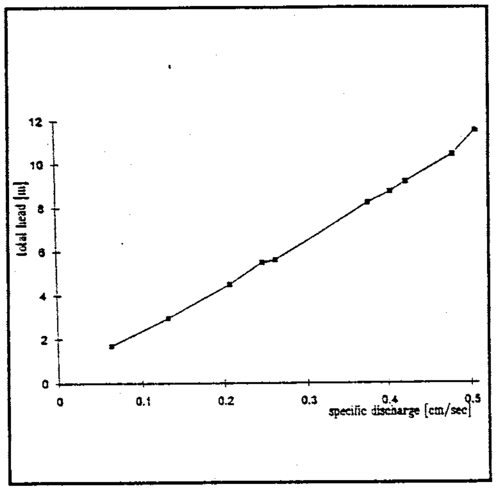


Figure 5.3: Flow v. total head, Darcy's First Series.

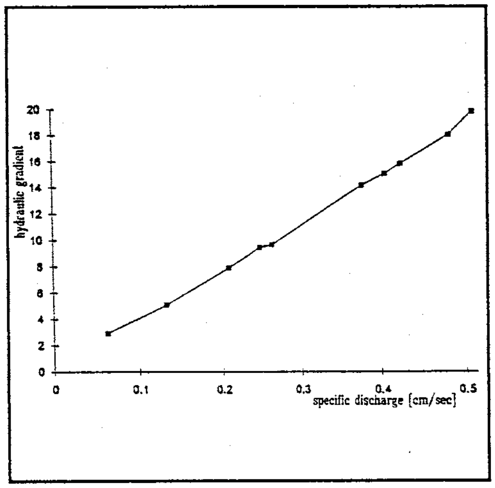


Figure 5.4: Flow v. hydraulic gradient, First Series.

23. Hubbert, M.K. (1969): *The Theory of Groundwater Motion and Related Papers*, Hafner Publishing Co. Pagination is from the 1969 collection.

In the introduction to the 1969 collection that includes *Theory of Groundwater Motion*, Hubbert recounts the origins of that paper (originally printed as an article in the *Journal of Geology*, **8**, pp.785-944): "At the time [it] was written I was without previous experience in groundwater hydrology or related subjects..." (p.11). He was doing earth-resistivity surveys with the Illinois State and United States Geological Surveys 1931-1936, and began to extrapolate from the flow fields governed by Ohm's law. He wondered what would happen if the electrodes were replaced by wells in a saturated homogeneous sand, and the battery by a pump connecting the wells. "Intuitively it was suspected that the water flow lines would be identical with those of electrical current in a medium of uniform conductivity" (p.12). The fluid analogue of electrical current was clearly the mass of water per unit area per unit time of water (or volume). The more difficult problem was to identify the analogue of electrical potential. In due course, taking an academic job at Columbia that afforded opportunity for research 1935-1937, he found that the groundwater potential "Φ is an energy per unit mass and is the sum of the two terms, a gravitational potential energy and a pressure energy" (p.13). "These results, achieved in 1936, were the consequences of intellectual curiosity. Next ignorance played its part. It was taken for granted that the foregoing results must be common knowledge and in daily use by the principal groups working with the flow of fluids through porous solids" (p.14). Hubbert notes, however, that O.E. Meinzer's "authoritative treatise" (1923), *Outline of Ground-Water Hydrology With Definitions*, USGS Water-Supply Paper 494 does not "even mention Darcy's law". Charles Slichter's (1899) "Theoretical Investigation of the Motion of Ground Waters" comes in for more extended criticism. Supposing that hydrology would be relatively primitive, Hubbert recalls his surprise that so many petroleum engineers also relied on a faulty understanding of Darcy's law. "At this stage, I concluded that my work of 1936 evidently was not, as I had supposed, common knowledge in ground-water hydrology and petroleum-reservoir engineering, and resolved to prepare a modest paper... The net result was a ream-sized manuscript entitled 'The Theory of Ground-Water Motion'..." (p.19). He recounts the resistance his analysis encountered among reservoir engineers in the petroleum industry, who were convinced (on the basis of Muskat, M. (1937), *The Flow of Homogeneous Fluids Through Porous*

Media, McGraw-Hill) that Darcy's law relied on the pressure head alone: "For nearly a decade, [my paper's] influence on either [hydrologists or petroleum engineers] was almost imperceptible... Such groups, it was found, can be peculiarly impervious to purely theoretical arguments, especially if the conclusions are at variance with prevailing opinions or dogmas" (pp.19-20). As a result, Hubbert and his graduate assistant Jerry P. Connor performed further experiments to extend the empirical basis of Darcy's law and to "test the validity of the principal theoretical deductions of the original paper". In particular, these experiments showed good agreement with theory in the area of tilted interfaces of fluids, whether freshwater/saltwater or oil/water in the form of hydrodynamic traps, the subject of a 1953 paper included in the 1969 collection.

Within the 1940 paper itself, Hubbert recaps the essentials of Darcy's experiments, with the difference that Hubbert's cylinder is "mounted so as to pivot about a horizontal axis perpendicular to its own axis", thereby allowing the direct investigation of the effect of tilting the column (there is none). He then assembles equations similar to Darcy's:

$$Q = -KA\frac{dh}{dl} \qquad (1); \qquad \Rightarrow q = -KA\frac{dh}{dl} \qquad (2).$$

After some discussion of the significance of, and differences between, elevation and pressure head, Hubbert comments: "If it is not elevation [since the column can be tilted], perhaps it is the fluid pressure that is the determining factor in the flow, with the flow always directed away from regions where the pressure is higher and toward those where it is lower. In fact, the great majority of all writers upon this subject have stated that this is so, and many have employed equations of the form:

$$q = -K'\frac{dp}{dl} \qquad (3)$$

as a statement of Darcy's law, presumably under the impression that equations [(2) and (3)] are physically equivalent statements" (Hubbert, M.K. (1940), pp.791-792). Hubbert then goes on to show that the two expressions are only equivalent for "nearly horizontal flow, or when dh/dl is very large" relative to the cosine of the angle of tilt (p.794). Hubbert is looking for some "physical quantity, capable of measurement at every point of a flow system, whose properties are such that the flow always occurs from regions in which the quantity has higher values to those in which it has lower, regardless of the direction in space. What we have demonstrated so far is that neither elevation nor pressure is such a quantity. Formally, the manometer height h satisfies this condition entirely..." Hubbert is not satisfied with this empirical measure, however, and quickly derives the Bernoulli equation and simplifies it for the negligibly slow flow of groundwater. He then uses a surface and body force argument to derive from fundamental mechanics the basic potential equation, and to break down the lumped parameter K into its component parts, arriving in due course at the familiar expression:

$$q = -Nd^2 \cdot \frac{\rho}{\eta} \cdot \frac{\partial\phi}{\partial l} = -Nd^2 \cdot \frac{\rho}{\eta}g \cdot \frac{\partial h}{\partial l} \qquad (4)$$

In a final point of interest here, Hubbert then demonstrates the difference between what he calls R*, the value of the Reynolds' number at which inertial forces become effective, and R_{crit}, the Reynolds' value at which turbulence sets in. In a pipe, these values are the same; they differ by more than two orders of magnitude within porous media.

Hubbert's efforts thus show the same interplay of conjecture and experiment, and the same analogy to electrical conductance, as the original experiments of Darcy, with an added component of theoretical deductions that we will see constitutes a *structural explanation*, capable of both increasing confidence in empirical relations and of guiding further research.

24. See note 21, above, for Hubbert's analytical treatment of the difference in porous media between the Reynolds' number at which inertial forces become effective, and the Reynolds' number at which turbulence begins.

25. Hubbert, M.K. (1940), p.792.

26. See Darcy, H. (1956), p.570: "J'avais déjà entrevu ce curieax résultat daus mesrecherches sour l'écoulement de l'eau dans les tuyaux de conduite de très-faible diamètre, lorsque la vitesse de l'eau ne dépasse pas 10 à 11 centimètres par seconde" ("We have already seen this curious result before regarding flow of water through pipes of small diameter, so long as water velocity did not exceed 10-11 cm/sec"). This footnote occurs early in his appendix where Darcy summarizes his results before presenting the details of the famous porous media experiments.

    A specific discharge of 10 cm/sec through Darcy's column would require a volumetric flow of almost 580 L/min.; if seepage velocity (Q/nA) is what is meant, 10 cm/sec still implies a volumetric flow of about 220 L/min. None of the experiments listed in Darcy's appendix used flow rates above 29.4 L/min.

27. Bear, J. (1972), pp.125-126, puts the range in which Darcy's law remains valid at between 1 and 10. Some of this range can be attributed to the different lengths used in the calculation. Bear mentions the mean grain diameter (Hubbert's choice), d_{10} and d_{50}, and various functions of k, permeability, or n, porosity, such as $(k/n)^{1/2}$, or $k^{1/2}$. In virtually all these cases, Bears says, Darcy's law remains valid for $1 < N_{Re} < 10$. Besides the variation in definition, there is also random variation within the samples. In the "Theory of Groundwater Motion" (1940), Hubbert says that use of the mean grain diameter "obviously gives only an order of magnitude, because different sediments may depart widely from geometrical similarity" (p.62). Darcy does not provide sufficiently detailed information on his sand samples to more than tentatively calculate the associated Reynolds' numbers.

28. Davis, P., *et al.*, (1992) note other early studies done on Daracy's law. These include Stearns, N.D. (1927), "Laboratory Tests on Physical Properties of Water-Bearing Materials", United States Geological Survey, Water-Supply Paper 596, pp.144-159, in which the hydraulic gradients ranged from $9x10^{-4}$ to $5x10^{-2}$; Meinzer, O.E. and Fishel, V.C. (1934), "Tests of Permeability with Low Hydraulic Gradients", in *Transactions of the American Geophysical Union*, **15**, pp.405-409, in which the reported gradients ranged from $3x10^{-5}$ to $7x10^{-3}$ (this study was motivated by the observation that: "Many of the water-bearing formations in the United States have gradients of much less than 20 feet to the mile [0.0038], and some may have gradients of less than one foot per mile". The results "strengthened the presumption" that "Darcy's law holds precisely for flow through permeable materials under indefinitely low gradients"). Fishel, V.C. (1935), "Further Tests of Permeability with Low Hydraulic Gradients", in *Transactions of the American Geophysical Union*, **16**, pp.499-503, extended these studies to gradients of less than one inch per mile, with reported gradients ranging from $5.7x10^{-6}$ to $1.9x10^{-3}$. Fishel concluded that "the rate of flow varies directly as the hydraulic gradient, down to a gradient of two or three inches to the mile, and there are indications that Darcy's law holds for indefinitely low gradients". In Darcy's original series of experiments, the hydraulic gradient ranged from 1.9 to 19 (see Figure 5.4 in note 22, above).

29. Shrader-Frechette, K. (1989): "Idealized Laws, Antirealism, and Applied Science: A Case in Hydrogeology", in *Synthese*, **81**, pp.329-352.

30. Shrader-Frechette (1989), p.335.

31. Shrader-Frechette (1989), p.337.

32. There are those who will insist that even the most commonplace observation is inherently "theory-laden". Hacking, I. (1983) comments: "Feyerabend says that observational reports, etc., always contain or assert theoretical assumptions. This assertion is hardly worth debating because it is obviously false [a claim Hacking supports with many famous examples], unless one attaches a quite attenuated sense to the words, in which case the assertion is true but trivial" (p.174). In particular, this assertion threatens to obscure altogether the distinction between empirical and fundamental laws, which would be a loss. Out of the hundreds of observations offerred by Francis Bacon to show, with no theoretical framing except the most primitive, that heat and light are not always the same thing, consider just one barrage: "And it is a well-known fact that, and looked upon as a sort of miracle, that a few years ago a girl's stomacher [an ornamental bodice insert], on being slightly shaken or rubbed, emitted sparks; which was caused perhaps by some alum or salts used in the dye, that stood somewhat thick and formed a crust, and were broken by the friction. It is also most certain that all sugar, whether refined or raw, provided only it be somewhat hard, sparkles when broken or scraped with a knife in the dark. In like manner sea and salt water is sometimes found to sparkle by night when struck violently by oars..." (1620b*, **2**:*xii*,11, p.152).

    In contrast to the theory-laden interpretation, Hubbert, M.K. (1969), p.270, discusses the evident confusion about what Darcy had done, and comments: "Darcy's own statement of the law was in an empirical form which conveys no insight into the physics of the phenomenon. Consequently, during the succeeding century many separate attempts were made to give the law a more general and physically satisfactory form, with the result that there appeared in the technical literature a great variety of expressions, many mutually contradictory, but all credited directly or indirectly to Henry Darcy".

33. Freeze, R.A. and Cherry, J.A. (1979): *Groundwater*, Prentice Hall, p.17.

34. Bacon, F. (1620b), *cxvii*, p.77: "My course and method, as I have often clearly stated and would wish to state again, is this - ...from works and experiments to extract causes and axioms, and again from those causes and axioms new works and experiments, as a legitimate interpreter of nature".

35. Shrader-Frechette, K. (1988): "Values and Hydrogeological Method: How Not to Site the World's Largest Nuclear Dump", in Byrne, J. and Rich, D. (eds.): *Planning for Changing Energy Conditions*, Transition Books, pp.101-137.

36. See the discussion of scientific explanation above at p.157. Bogen, J. and Woodward, J. (1988), "Saving the Phenomena", in *The Philosophical Review*, **97**, pp.303-352; Woodward, J. (1989), "Data and Phenomena", in *Synthese*, **79**, pp.393-472; Cartwright, N. (1983), *How the Laws of Physics Lie*, Oxford University Press.

37. See Bogen, J. and Woodward, J. (1988); Woodward, J. (1989); Hofmann, J. (1990).

38. Kroes, P.A. and Sarlemijn (1989): "Fundamental Laws and Physical Reality", in Sarlemijn and Sparnaay (eds.): *Physics in the Making*, Elsevier Science Publishers, pp.303-328.

39. Kroes and Sarlemijn (1989), p.324.

40. Kroes, P.A. and Sarlemijn (1989), p.326.

41. Hofmann, J. (1990).

42. Cartwright, N. (1989): "The Born-Einstein Debate: Where Application and Explanation Separate", in *Synthese*, **81**, pp.271-282.

43. Cartwright (1989), p.275.

44. Medawar, P.B. (1967), *The Art of the Soluble*, Methuen and Co, Ltd., p.154.

45. Hofmann, J. (1990).

46. McMullin, E. (1978): "Structural Explanation", in *American Philosophical Quarterly*, **15**, pp.139-147, (p.139).

47. McMullin (1978), p.147. "Ontological reliability" means entities whose existence is indicated by the test are usually or always really there. "Retroduction" is the backward inference common to all inverse methods. In short, McMullin is saying that in this case the inverse method generally works (doesn't give too many false positives).

48. Mercer, J. and Faust, C. (1981): *Ground-Water Modeling*, Reston: National Water Well Association, p.2.

49. Bear, J. (1972), *Dynamics of Fluids in Porous Media*, American Elsevier Publishing Co., p.92.

50. Bear (1972), p.24.

51. Freeze, R.A. and Cherry, J.A. (1979), p.17.

52. Bear, J. (1972), p.91.

53. Bear, J. (1972), p.92.

54. Bear, J. (1972), p.93.

55. Bear, J. (1972), pp.104-6.

56. See Hubbert, M.K. (1956), pp.277-289 for the details of this structural explanation of Darcy's law.

57. In a later paper (1956) honoring the centennial of Darcy's work, Hubbert [(1969), "Darcy's Law and the Field Equations of the Flow of Underground Fluids", pp.265-30] reprises some of his analysis from 1940. After again describing the basics of Darcy's experiment, and generalizing it to tilted columns, M. King Hubbert immediately moves to a generalized theoretical approach: "Darcy's work... forms a solid experimental foundation for... a field theory" (p.266). In an earlier paper, Hubbert (1969, p.792 [32]), Hubbert explains that

> the method of obtaining the equation has been empirical, and, expressed in this primitive form, the equation is of slight usefulness because it expresses only what we have learned already and gives us no insight into the deeper mechanism of fluid flow. What determines the direction of the fluid flow in the first place? What would be the effect if we changed from a finer to a coarser sand? How would the flow rate be altered if we changed the viscosity or the density of the fluid? These and other similar ones are questions that [the primitive form of the equation] does not answer; yet they are questions to which it is highly important that answers be known.

In this 1956 paper, Hubbert pursues structural explanations by first breaking the proportionality constant K down into its component bits. This "extension of the empirical method" is necessary insofar as Darcy's law "remains an empirical equation devoid of dynamical significance since there is no obvious reason why the flow of a viscous fluid through a porous solid should be proportional to a dimensionless quantity, -grad *h*" (p.272). Showing that $K = Nd^2\rho g/\mu$, Hubbert finds his "dynamical factor g" that has "evidently been concealed in the original factor K", and that "the final factor of proportionality", N, "must be related to the only remaining variable, namely the shape of the passages through which the flow occurs". After some discussion of the "regrettable" definition of the darcy unit of permeability by means of an incomplete statement of Darcy's law, Hubbert says: "Having thus achieved the desired generalization and a proper physical statement of Darcy's law by an extension of the empirical method which Darcy himself employed, let us now see if the same result can be derived directly from the fundamental equation of Navier and for the motion of a viscous fluid" (p.275). Having achieved this goal, Hubbert then comments on the results: "The direct derivation of Darcy's law from fundamental mechanics affords a further insight into the physics of the phenomena involved over what was obtainable from the earlier method of empirical experimentation" (p.282). Among these insights, Hubbert mentions that Darcy's law is not a special case of Poiseuille's law, but "Poiseuille's law is in fact a very special case of Darcy's law". As a result, Hubbert is prepared to clarify "deductions concerning the Darcy-type flow made from the Poiseuille flow [that] are likely to be seriously misleading". In particular, it becomes clear that the Reynolds' number at which Darcy's law fails (now put at 1, instead of 4 as in 1940) is not the Reynolds' number at which turbulence begins (about 600).

In the conclusion of this paper (p.58), Hubbert reiterates his program to pay his respects to Darcy by 1) clarifying what Darcy did and to "give his results a more general, but still equivalent physical formulation"; 2) deriving Darcy's law directly from the Navier-Stokes equation of motion; and 3) to develop at least preliminary field equations. Of particular interest to our discussion of structural explanation is the way Hubbert characterizes the results of his efforts as leading to a "comprehensive unification" of hydrologic field theory:

> These objectives have now been accomplished, and the result is that, despite a number of troublesome complexities such as those arising from thermal convection and from waters of variable salinity, the field theory of the flow of underground fluids, is capable of being brought into the same kind of a comprehensive unification as that already achieved for the more familiar phenomena of electrical and thermal conduction.

58. McMullin, E. (1987): "Explanatory Success and the Truth of Theory", in Rescher, N. (ed.): *Scientific Inquiry in Philosophical Perspective*, University Press of America, pp.51-73. See pp.59-60.

59. Hofmann, J. (1990), p.411.

60. Mott, N. (1941), "Application of Atomic Theory to Solids", in *Nature*, **147**, pp.623-624.

61. Bear, J. (1972), p.92.

62. Bear, J. (1972), p.92.

63. See the earlier note (56) on Hubbert's analysis and its relation to experiment.

64. Bear, J. (1972), pp.162-3. See page 171, above.

65. Bear J. and Veruijt, A. (1987): *Modeling Groundwater Flow and Pollution*, D. Reidel Publishing Co., p.12.

66. See, *eg*., Hubbert, M.K. (1969), pp.283-287, where Hubbert suggests we "now see if the value of N [the shape factor], at least within an order of magnitude, can be determined theoretically". Beginning from the governing differential equations, Hubbert derives an expression for $N_{\bar{\lambda}}$, the

"shape factor corresponding to $\bar{\lambda}$ as the characteristic length of the system". $N_{\bar{\lambda}} \approx f/6$, where f is the porosity. Several approximations serve to expedite this process, since "to attempt to do this in detail would be a statistical undertaking beyond the scope of the present paper". The mean half-gap width, $\bar{\lambda}$, (which Hubbert notes could be measured directly) is then estimated via an indirect conceptual method in which a rectangular prism is imagined to pass through a macroscopically homogeneous and isotropic porous solid. A general result for the intersection of the prism with voids and solids is then applied to uniformly packed spheres, and Hubbert arrives at a theoretical value of $\bar{\lambda}$ as a function of the porosity and mean grain diameter for packs of uniform spheres. Hubbert uses this expression to derive a relation between $N_{\bar{\lambda}}$, the shape factor related to the characteristic length within the system, and N_d, a function of sphere diameter. Experiments carried out by Jerry Conner resulted in a typical value of $N_d = 6 \times 10^{-4}$, which, when coupled with a porosity of 37%, in turn gave an experimental value of $N_{\bar{\lambda}} = 26.2\ N_d = 0.0157 = 0.25 N_{\bar{\lambda},\ \text{theoretical}} = f/6 = 0.0617$. Echoing Galileo's reply to Simplicio (p.126), Hubbert comments on the "error":

> Since our object at the outset was to merely gain some insight into the nature of the shape-factor N occurring in Darcy's law, no particular concern is to be felt over the the discrepancy... between the observed and theoretical values. All this really indicates is that the system of averaging required should be better than the oversimplified one actually used... The fact that our approximate analysis yields a result in error by only a factor of 4 makes it appear promising that if account is taken of the variability of λ and of $\bar{u}$ [mean velocity] as a function of λ, much better approximations may be obtainable.

6

Ripples in Still Water: A Diffusion Hypothesis

> when [the geometrical philosopher] wants to recognize in the concrete the effects which he has proved in the abstract, [he] must allow for the impediments of matter, and if he is able to do so, I assure you that things are in no less agreement than are arithmetical computations. The errors lie, then, not in the abstractness or concreteness, not in geometry or physics as such, but in a calculator who does not know how to keep proper accounts.
>
> - Galileo Galilei, *Two New Sciences*[1]

> Puzzles mount on puzzles the more we consider details.
>
> - Stephen Jay Gould, *Wonderful Life*[2]

6.0 Background

In Chapter 3 we saw how the availability of more than one thoroughly articulated basin-scale model at Pojoaque has complicated regulatory and legal procedures by forcing an attempt at model comparison that both the science and the courts are ill-equipped to make. There are immediate methodological implications. According to Ernan McMullin, it was one of Galileo's great contributions to science to point out that systems could be approached in piecemeal fashion, and the cumulative result assembled from the component bits. Francis Bacon made such an approach the centerpiece of his methodology. Process orientation is usually characterized by narrowly defined and carefully exclusive problems; the exclusivity is effected by either *crucial assumptions* in theoretical work, or by *controlled experiments* in the field or laboratory. Although the critical reconsideration in moving from applied to process-oriented work is less one of scale than of purpose, reduced scale is often among the most straightforward ways to eliminate or highlight (as needed) the role of Galileo's *causal impediments*.[3] Chosen internal problems are also for obvious reasons ones the researcher has some idea he can solve with existing tools, or slight extensions of the same.[4]

The impasse over the Tesuque aquifer models should make us sympathetic to the position of the molecular biologists described by Evelyn Fox Keller: "The mood in biology [around 1960] had grown impatient with the complexity of higher organisms".[5] In Keller's account, the enthusiasm of those researchers for more constrained investigations was both justified and fueled by the conviction that "what was true for *E. coli* was true for the elephant".[6] Part Two of this study is devoted to process hydrologists' equivalent retreat from "higher organisms" to the study of separate pieces of hydrologic puzzles. Utilizing the history and subsequent treatment of Darcy's law, the interplay of theory, model and experiment have already been highlighted in recounting the discovery and exploration of the most general of empirical hydrologic laws. While Darcy's work was directly applicable to the problem of sizing in-line water filters, it clearly functions in a different way in relation to aquifer scale problems. It remains to further illuminate these points with a contemporary example of process research aimed at the component bits of Galileo's *impediments of matter*.

We therefore next consider the typical methodology of modern research programs in which the results may be less sweeping but more detailed. A representative thread will be presented of recent research into solute mixing at fracture junctions, culminating in work performed at the New Mexico Institute of Mining and Technology. Fractured systems pose special problems both in terms of simple flow and with respect to solute transport. A recent National Research Council committee noted that: "In many bedrock aquifers, fractures provide an appreciable contribution to the capacity of the medium to transmit a fluid. At greater depths, they may account for the primary contribution, with a much smaller fluid flux occurring within the rock matrix".[7] To default by treating fractured media as a type of porous medium governed by Darcy's law may therefore by unacceptable; a special description of fluid distribution via fractures may be deemed an essential component of hydrologic models. Unfortunately: "The scientific basis for describing fluid flow and solute transport in fractured media lags behind the state of knowledge for describing these processes in porous media systems".[8] One aspect of this unresolved general problem is the controls on mixing behavior at fracture intersections or junctions; *i.e.*, what happens when the fluid streams from two or more fractures meet at one point, and in particular, what governs the distribution of incoming solutes?

When their codes do attempt to capture the special effects of fractures, modelers typically resort to one of two simple descriptions of junction effects on incoming streams: no mixing, or, more often, complete mixing. In the first case, solute transport is treated as entirely advective and non-turbulent; *i.e.*, the solutes of interest are simply carried along the flow streamlines at the mean flow velocity; solutes must therefore be conservative. Under laminar conditions, streamlines do not cross, and solutes cannot escape a streamtube; describing solute mixing at fracture junctions is therefore essentially a matter of describing the distribution of flow within the system.[9] Transport of this sort is often termed *streamline routing*. Assuming that advection dominates solute behavior relies on the fact that flow through fractures can be very fast relative to both 1) flow through the adjoining rock matrix; and 2) the diffusive time constant within the bulk fluid. In the second case, an alternative modeling assumption treats solute transport as a diffusive process independent of the flow regime; transport is not advective within the junction. Exiting flows carry identical averaged solute concentrations due to diffusion across streamlines within the junction that leads to *complete mixing*. Transport through the channels to and from channel junctions is typically again by advection. Describing junction behavior in this way reflects a belief that groundwater movement is generally slow, even through fractures.

The assumption of complete mixing is especially computationally convenient, but neither assumption had been put to any crucial tests before the mid-1980s.[10] After commenting on the relative maturity of groundwater flow models (p.56), Mary Anderson extends her 1983 imagery to a second class of modeling problems of interest here:

> However, in the case of solute transport models the Emperor's underwear, the theory upon which the model is based, is full of holes. There is continuing controversy regarding the proper field measurement and quantification of the dispersion term in solute transport models.[11]

Fractured media transport modelers necessarily lapsed for some time into the posture of the Tesuque aquifer modelers: further model embellishments were not justified by available understanding or data. In contrast to the confounded decisionmaking of that case, however, there was every reason to think that the specific question of fracture effects could be answered in due course. We propose to take a detailed look at the means by which process hydrologists have since constrained the problems under study in the hope of eventually assembling accounts of larger phenomena of practical import. Along the way, we will gain a thoroughly modern perspective on the mutually supporting roles of experiment, models, and theory introduced in Chapter 5.

Research on the process level proceeds as experimental work, mathematical derivations and numerical simulations. Progress is often measured by the evolution of conceptual models. The significance of process research, on the other hand, is usually tied to the generality of the conclusions. The urgency of practical concerns is not irrelevant to process research; the retreat from applied aquifer-scale problems by process hydrologists is nominally tactical, in that the justification for their research is invariably the possible elucidation of real problems. The essence of investigated systems is still at trial, even though the systems may be small and generic. The hope within process research is often that systems are reasonably linear and reasonably additive; in short that the system is only complicated, and not complex, as the latter term has come to be used in non-linear dynamics (chaos). Chastened by the experience of applied modelers, however, we will do well to bear in mind throughout this account the final moral of Keller's story about reductionist molecular biologists of the 1960s; namely, that not only wasn't their particular theory true for the elephant, it wasn't even true for *E.coli*.[12]

6.1 Experimental Benchmarks

In 1986 Laurence Hull and Karen Koslow published a report on their experimental investigation of streamline-routing through fracture junctions.[13] The significance of the research problem was attributed to its relevance in the siting of nuclear waste repositories (Yucca Mountain), the exploitation of geothermal and petroleum assets, and hazardous waste remediation. In environmental applications, the degree of mixing at fracture junctions is important because of its effect on both lateral spreading of contaminant plumes and the maximum concentration within plumes. Hull and Koslow explain: "If mixing at junctions is overestimated, then the lateral spread of the plume will be overestimated and the peak plume concentration will be underestimated".[14] Thus if streamline routing is a reasonable description of junction behavior, the common assumption of complete mixing at fracture junctions could have a misleading dispersive effect within applied models of fractured media.

Hull and Koslow do not rely on theory in planning their experiments. Their interpretation of the results, as we will see, also invokes a minimum of fluid mechanical theory, a fact reflected in their work's publication as a Technical Note in *Water Resources Research*. They are interested in mounting an experimental challenge to the typical assumption by numerical modelers of complete inflow mixing at fracture junctions. For example, in 1984, Leslie Smith and Frank Schwartz published a paper on the effects of fracture density, length and aperture on solute transport within fractured systems.[15] Among the simplifying assumptions in this stochastic analysis is that "there is complete mixing of mass at fracture intersections".[16] The challenge by Hull and Koslow takes the form of a straightforward demonstration of streamline routing, in which minimal mixing occurs. They note the existence of a few fractured media codes that assume streamline routing, and they are prepared to delineate the conditions under which this is a reasonable assumption.

The fracture junction employed for this purpose is highly idealized. It consists of orthogonal channels 0.16 cm wide by 0.95 cm deep cut in an acrylic sheet which is then capped with a second sheet to enclose the system. Flow is thus in the form of a plus sign; any two channels can be used as inlet channels and the other two as outlets. Flow in all four channels is controlled independently, allowing flow through the system to be varied in several ways. Hull and Koslow examine the effects of two flow configurations. They call the first arrangement *continuous junctions*, meaning that the inlet channels are adjacent and form a right angle (see Figure 6.1a, in which channels 1 and 2 are the inlets). They also observe the behavior of what they call *discontinuous junctions*, meaning that the inlet channels directly oppose one another and form a straight line (see Figure 6.1b, in which channels 2 and 3 are the inlets). Furthermore, within either configuration, four ratios of inlet channel velocities are used, ranging from

1:1 to 3.2:1. For each inlet flow ratio, the outlet flow ratio is then set at 1:2, 1:1 and 2:1, resulting in a total of 12 basic experiments for each configuration. Absolute flow velocities range from 0.06 to 0.58 cm^3/min, giving a Reynolds' number range of 0.2 to 2.0. Additional experiments examine the effect of holding flow ratios constant but increasing the absolute flow rate.

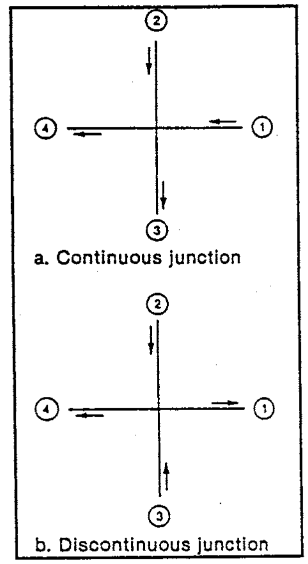


Figure 6.1: Conceptualization of junctions: a) continuous; b) discontinuous. From Hull and Koslow (1986).

In every case, one inlet channel carries a solution of 17 mg/L KCl, while the other contains an equal concentration of $KMnO_4$. The densities of these solutions are the same, so flow characteristics are unaffected by the solutes. $KMnO_4$ has a lower electrical conductivity than KCl; mixing of the inlet streams under different conditions can therefore be calculated from measurements of the electrical conductivity of the solutions downstream of the junction. Platinum probes installed flush with the channel walls for this purpose are 0.3 cm in diameter, or about 1/3 of the total channel depth. Electrical conductivity data are collected after the system reaches an apparent steady-state. $KMnO_4$ also serves as a visual tracer of mixing behavior; these observations play no direct role, however, in the quantitative analysis. Although Hull and Koslow do not mention the Peclet number, it is a useful alternative measure of experimental conditions; for their experiments, $300 < Pe < 3000$. [17]

The hypothesis that streamline routing and velocity ratios are the major controls on the behavior of interest is readily translated into a quantitative model of continuous fracture junctions. Assuming only that streamlines cannot cross at continuous junctions (see Figure 6.2a), Hull and Koslow use a very simple mass balance analysis to calculate expected solute concentrations in the downstream channels. Referring again to Figure 6.1a, using Q to denote flow, C to denote concentrations, and subscripts to denote channels, if $Q_1 \geq Q_3$, then $C_3 = C_1$, and $C_4 = C_2(Q_2/Q_4) + C_1[1 - (Q_2/Q_4)]$. Under these conditions, the flow in channel 3 is entirely from channel 1; channel 4 may receive flow from channel 1 (if $Q1 > Q3$). On the other hand, if $Q_1 < Q_3$, then $C_4 = C_2$, and $C_3 = C_1(Q_1/Q_3) + C_2[1 - (Q_1/Q_3)]$. Now the flow in channel 4 is entirely from channel 2, while channel 3 may receive flow from channel 2. A case that will be of particular interest for our later purposes occurs when $Q_1 = Q_2 = Q_3 = Q_4$. Under these flow conditions, the dividing streamline occurs diagonally across the fracture, and $C_1 = C_3$; $C_2 = C_4$.

The conceptual model of flow and solute mixing at <u>dis</u>continuous junctions is slightly complicated by the possibility of adjacent streamlines moving in opposing directions at the dividing streamlines. Hull and Koslow expect that in such a case: "Viscous forces would cause flow along these streamlines to slow until [a] stable configuration was achieved", in which the occurance of adjacent and opposing streamlines is minimized (mechanical mixing is also thereby minimized). Despite uncertainty about the shape and location of the dividing streamlines, calculation of the expected downstream concentrations is simplified by assuming a stable configuration will be maintained at steady-state. Referring to the discontinous junctions in Figure 6.1b, Hull and Koslow expect that under all flow conditions outlet concentrations would be equal. The concentration is given by: $C_1 = C_4 = [(C_2Q_2+C_3Q_3) / (Q_2 + Q_3)]$, again by simple mass balance. The position of the horizontal dividing streamline is controlled by the ratio of inflows; the position of the vertical dividing streamline is controlled by the ratio of outflows (see Figure 6.2b). The rule developed for discontinuous fractures has since been termed *proportional routing* by other investigators.

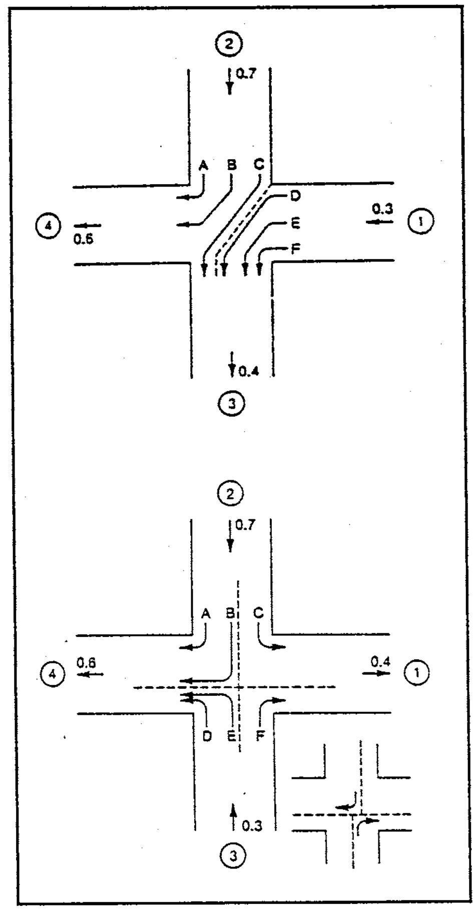


Figure 6.2: "Possible" dividing streamlines in junctions: a) continuous; b) discontinuous. From Hull and Koslow (1986).

In summary, to account for their results within continuous fractures Hull and Koslow assume only that streamlines cannot cross; within discontinuous fractures, minimization of adjacent streamlines moving in opposite directions is also required. We turn now to the experiments themselves. Independent measurements of inflow and outflow during all experiments revealed slight mass balance errors in both continuous and discontinuous junction experiments, ranging up to 9.3% in the former and 6.6% in the latter. Hull and Koslow nevertheless report very good agreement between their measured outflow concentrations and the assumed streamline or proportional routing incorporated into the mass balances presented above. They report a maximum error of 8% in continuous junction experiments (10.5% in discontinuous experiments), with most of the errors considerably less. $KMnO_4$ served as a visual tracer to confirm the minimization of adjacent streamlines flowing in opposite directions within discontinuous junctions: "this was, indeed, what was happening in the junction". Referring to the conditions implicit in their mass balance equations, the investigators conclude: "The excellent agreement... indicates that the two criteria are sufficient to define solute transfer through a junction under laminar flow conditions".[18]

A year later, Hull was the lead author of a report on additional physical experiments and numerical simulations.[19] The new experiments examine solute transport through an orthogonal fracture network of channels 0.318 cm wide by 1.91 cm deep, again manufactured from acrylic sheeting. All channels meet in continuous junctions; channel spacing is a uniform 10.2 cm. Preliminary results provide experimental evidence that head loss occurrs only within the channels and is negligible within the junctions; the alert reader may recall this assumption figured in the Bear-Bachmat derivation of Darcy's law within porous media (p.170). In this case, the physical model is intended only to "validate" hydraulic and transport algorithms of the numerical code.

The numerical model of Hull, *et al.*, treats streamline routing and "complete homogenization" as end-members. The code includes a diffusion coefficient (set to zero for streamline routing), resulting in "three different sets of assumptions".[20] This numerical model was run for both a uniform model resembling the physical model and also for networks of variable apertures and spacing. These simulations bear out the 1986 contention that the mixing rule governs both lateral spread and maximum concentrations of plume. Based only on their numerical model, Hull, *et al.*, also conclude that fracture aperture is a significant variable: "In crystalline rocks with fracture apertures on the order of microns, complete mixing will occur in fractures and in fracture junctions at velocities up to several meters per day. For more permeable fracture systems with apertures on the order of millimeters the assumption of complete mixing at junctions is probably not justified".[21] This assessment relies on the absolute magnitude of fracture apertures and does not consider factors in combination. As aperture width decreases, however, so does the Peclet number for a given velocity, so (diffusive) mixing should increase.

It remained to be seen which mixing rule is most reasonable within larger systems. There was also as yet no theoretical foundation for any of these experimental observations and extrapolative conjectures, but theoreticians were not far behind. The work of Hull and Koslow (1986) promptly attracted the attention of hydrologists willing and able to apply fluid mechanical theory to the problem of flow through idealized fracture junctions.

6.2 Theory of Discontinuous Fracture Junctions

In 1988, noting that the 1986 work of Hull and Koslow had "eschewed the application of quantitative fluid mechanics", J.R. Philip published a report with just such an emphasis.[22] In it we can see the usual subdivision of labor within process research clearly at work. First fracture flow was distinguished from flow through porous media; later, specific fracture geometries were found to call for separate treatment. The 1987 report of Hull, *et al.*, shows an eagerness to move forward to conclusions about practical problems of interest, such as flow through proposed nuclear waste respositories. Philip, however, wants to consider carefully aspects of the earlier work.

Philip accepts the "obvious and unambiguous" treatment of continuous junctions by Hull and Koslow. He concentrates on the fluid mechanics of the more problematical discontinuous junctions and advances a theoretical exploration of the proportional routing suggested by Hull and Koslow. Philip states his purpose clearly; he is ultimately looking for a "stringent test of experiment and theory, oriented to the critical practical case of a small inlet flow of an undesirable pollutant". This purpose is subdivided into two aspects: 1) "establish the dividing streamline and use it to test the validity of streamline routing"; and 2) find a "measure of the error of proportional routing for the particular discharge pattern".[23] In short, he is looking for both a structural explanation of the empirically obtained relationship and also some idea of the practical bounds on the validity of that relationship.

Philip begins by pointing out that "our analysis depends on the same physical assumptions as those of Hull and Koslow". In particular, 1) flow is laminar and steady; 2) there is no diffusion across streamlines within the junction; and 3) diffusion leads to complete mixing within the channels. It is characteristic of process research that where possible the conceptual model of the theorist is identical to the physical model of the experimentalist (within experimental noise). Such is the case here (see Figure 6.3). Philip uses conformal mapping to generate "canonical" solutions for both Laplace and Stokes flow through his model. The Laplacian (plug flow) is simpler and applies to Hele-Shaw models in which the channels are shallow; Stokes flow occurs when channels - like those of Hull and Koslow - are deep relative to their width.[24] Philip notes that real-world fractures tend to be narrow and deep, and that we should expect flow within them to approximate Stokes behavior rather than Laplacian.[25] In this respect, at least, the physical model of Hull and Koslow and the conceptual model of Philip are thus potentially relevant to the study of many practical problems.

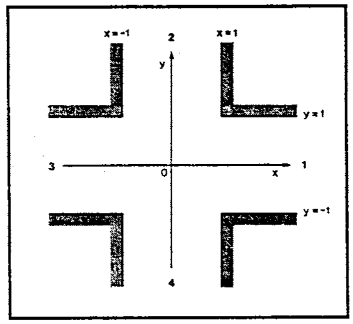


Figure 6.3: Model junction. From Philip (1988).

Combining his mathematical model with the reported data of Hull and Koslow, Philip finds inexplicably that the data appear to indicate the existence of Laplace flow, instead of the Stokes flow that ought to prevail under the experimental conditions. Nevertheless, analyzing the effects of different flow magnitudes in the different channels, Philip concludes that proportional routing is reasonably accurate under many circumstances. It is when "the two inlet discharges and the two outlet discharges both differ greatly in magnitude" that "the errors of proportional routing may be unacceptably large". Philip bases this conclusion not only on his own analyses, but also on the data of Hull and Koslow. This introduces some reservations about the claim of Hull and Koslow of "excellent agreement" between their proportional routing model and their observed downstream concentrations, since Philip finds: "The fate of the solute downstream of the junction depends critically on the ratio of the outflow discharges". Within the conventions of his mathematical analysis, Philip then provides a quantitative assessment of just when "the errors of proportional routing are likely to be of practical significance". Under certain conditions, these errors can be "serious".[26]

In the same spirit earlier illustrated with regard to Darcy's law in Chapter 5, Philip had thus demonstrated a theoretical basis for - and the expected limits on - an empirical relation developed by Hull and Koslow. His conclusion points out the generality of his method: "The present analysis can be extended to junctions of fractures with different widths, with different number of branches, and different angles of intersection..." Despite some reservations about details of the experimentalists' methods and analysis,[27] Philip's structural explanation strongly advanced the claim of streamline routing through discontinuous junctions under specified conditions. He also provided a mathematical description of mixing behavior that would prove useful to others. The strength of this support and the usefulness of the description were both critically related to the identity of experimental and theoretical conditions. However, as Hacking says, "The ability to explain carries little warrant of truth".[28] Philip's work could not be described as a validation of events and conjectures related to the particular apparatus of Hull and Koslow. The reasonableness of this assertion is reflected in the fact that Philip's work did not lay the matter to rest, but instead led to another cycle of experiment and theory.

6.3 Additional Continuous Fracture Experiments

In 1990, James Robinson and John Gale published additional experimental arguments in favor of no mixing at fracture junctions. They review the work to date, including that of Hull and Koslow. They link their purpose to a summary of Philip's conclusions, which "showed... that proportional routing is erroneous when the magnitude of the inlet and outlet fractures differ greatly. No serious attempt, however, has been made to test these conclusions in the laboratory and to incorporate the findings into a fracture network model".[29] Robinson and Gale are unequivocal in assessing the results of their own work: "The mixing algorithm for open fractures, used until recently in most transport models for fractured rock systems, is in error".[30]

The findings of Robinson and Gale rely on experiments with 14 separate acrylic models, of which 12 are "fully-intersecting" fractures of the sort studied by Hull and Koslow. Channels are again much deeper than wide, but Robinson and Gale investigate more complicated scenarios in addition to the simple orthogonal junction of Hull and Koslow. The channels have differing apertures (0.28 - 0.5 mm) and angles of intersection (22.5° - 157.5°). Moreover, some of these models are "intersecting fracture sets"; *i.e.*, systems of many fractures, rather than the single fracture junction of Hull and Koslow. Robinson and Gale do not give absolute flow rate information; instead, they describe their experiments as using three hydraulic gradients: 3.33, 1.67, 0.33. The resulting Peclet numbers are greater than 3,000. [31]

The reader should be aware that Philip's reservations relate to a large disparity in outlet *discharges*, and say nothing specifically about outlet *apertures*. The published analytical solutions of Philip in fact apply only to junctions with four equal apertures. Furthermore, although Robinson and Gale label their fracture junctions as discontinuous, they do not use this term in same sense as others; their figures clearly depict junctions that are continuous in the sense used by Hull and Koslow and later by Philip.[32] The experiments of Robinson and Gale are therefore reasonably viewed as a relaxation of certain conditions in the continuous junction experiments of Hull and Koslow, rather than as the claimed test of Philip's analysis of discontinuous junctions. None of this makes the results of less interest to the general problem of solute mixing at fracture junctions; as Philip pointed out, his solutions could be generalized to different boundary conditions, etc. It is also very likely that continuous junctions are far more common in nature than discontinous ones. The cycle is thus from the initial empirical relationship of Hull and Koslow, to the theoretical explanation and reservations of Philip regarding aspects of the proposed relationship, to additional experimentation by Robinson and Gale that considers new conditions and couples the results to another numerical simulation.

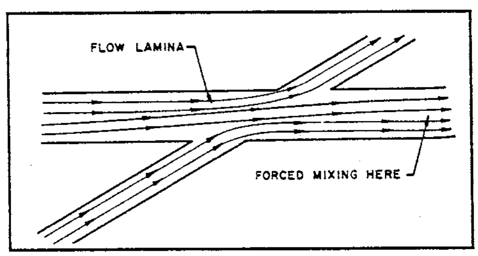


Figure 6.4: Forced mixing downstream of a junction due to unequal apertures. From Robinson and Gale (1990).

Summarizing more than 300 tests conducted with their acrylic models, Robinson and Gale find that "the overall average adjusted mixing percentage is less than 3%".[33] The small amount of mixing causes them to ascribe little significance to "minor trends" observed in their results: 1) less mixing at lower hydraulic gradients;[34] 2) less mixing in models with unequal apertures; 3) the most mixing at junctions whose channels meet at angles between 67.5-112.5°; and 4) the least mixing at the most acute or obtuse junction angles.[35] They are unambiguous, however, in their major findings: "It may be concluded, on the basis of these tests, that no mixing occurs at fracture intersections except that which is forced to take place due to the flow differential that may exist in the two fractures".[36] Based on their experimental results, Robinson and Gale emphasize the mechanical mixing of streams downstream of a junction (see Figure 6.4); in fact, they deny any other cause of mixing.

Robinson and Gale also construct a two-dimensional numerical model of non-reactive advective transport. This model assumes complete mixing within each fracture segment (outside junctions), even though "the laboratory tests indicate that these streams do not mix in open fractures".[37] Synthesized random fracture networks are used to analyze model sensitivity to mixing rule, aperture, angle of fracture intersection, and flow rate. Preliminary test results align with the expectation of Hull and Koslow, showing: "When mixing is assumed to be perfect at the intersections, the solute quickly spreads out over the whole model", while on the other hand, "When no mixing is allowed at the fracture intersections, except that dictated by the variations in flow rate, the spread is very much less..."[38] The authors again refer to "forced mixing" occurring downstream of the junction of channels with unequal apertures.[39] What little mixing occurs within their numerical simulation is attributed to this mechanism: "Essentially no mixing takes place at fracture intersections. The only exception was the mixing that is forced to take place due to different-sized apertures... The perfect mixing model of Castillo (1972) and Smith, *et al.*, (1985)...has been shown by this study to be incorrect".[40]

Robinson and Gale then update their simulations by incorporating their experimentally and numerically determined mixing rules: "The model was designed to reflect the findings of the mixing tests... It was assumed that no mixing occurred at the fracture intersections unless the flow rates in the elements [channels] were different, in which case forced mixing was assumed to take place". In one case, little mixing occurs despite flow through six junctions.[41] Robinson and Gale conclude with an application to the Stripa mine, a Swedish site much studied for insights into proposed nuclear waste repositories.[42] The authors recognize the uncertain relevance of their work to this application, since the Stripa system contains relatively few intersecting fractures: "Under these conditions... the effects of using the correct mixing algorithm would be minimal...". As a result: "It is important to establish the proportion of four-way intersections in a fracture network in order to apply the appropriate mixing algorithm".[43]

In summary, Robinson and Gale substantiated in fracture networks what Hull and Koslow (1986) had anticipated based on singular junctions: no mixing leads to less spreading and higher outlet concentrations. The work of Robinson and Gale also supported the conclusions of Hull, *et al.* (1987) on the relation between mixing behavior, flow rate, and aperture width. At the flow rates considered, very little mixing occurred at fracture junctions. Robinson and Gale attributed the small remaining mixing behavior to the mechanical effect of forced mixing; they also identified their results loosely with the reservations of Philip regarding proportional routing. Accordingly, Robinson and Gale eliminated diffusive transport from their numerical model, allowing only mechanical mixing due to different outlet flow rates or aperture widths. This constitutes a simplification from the work of Hull, *et al.*, who incorporated a diffusion coefficient into their model despite their suspicion that it might be unnecessary; such a simplification might be considered preparatory to practical applications, if basic questions had in fact been settled. Remarkably, Robinson and Gale did not mention or seemingly consider what flow rates may actually be attained in fracture flow. They treated the problem at hand as basically a pipe flow problem of fluid distribution at relatively high flow rates.

6.4 A More Inclusive Numerical Study

In 1994, Brian Berkowitz, Curt Naumann and Les Smith published a numerical study of fracture junctions that selectively built on the previous work.[44] Their introductory comments identify the main thread in the evolving research, a thread that shows relatively little interest in discontinuous junctions. They recognize the forced mixing of Robinson and Gale, but set it aside: "Our focus in this paper is on mixing within the intersection, and not on forced mixing downgradient from the intersection".[45] Robinson and Gale had ruled out mixing within fracture junctions; their models (both physical and numerical) showed a slight mixing that occurred entirely within fractures, not at junctions. Berkowitz, *et al.*, is thus another restraining study that tends to rein in the eagerness of some to consider basic questions settled and to move on to at least tentative extrapolations to real-scale problems. Although this is reminiscent of Philip's reaction to Hull and Koslow, Berkowitz, *et al.*, do not take a rigorously theoretical approach; theirs is a numerical study that incorporates approximations of existing theory. The model developed is a general one that includes both streamline routing (which they prefer to call streamtube-routing) and diffusion: "The objective of this paper is to examine mixing processes at a fracture intersection, using a numerical approach..., and to determine the conditions under which the streamtube-routing and complete mixing models may be valid".[46]

The conceptual model employed for the fracture junction is the familiar orthogonal one of previous researchers. The authors decide to "intentionally restrict our investigation to [this] idealized geometry... in order to examine in as simple a framework as possible current concepts used in modeling mass transfer at fracture intersections" (see Figure 6.5). Additional simplifications are also familiar, and include that fractures are smooth, parallel-plate openings, that no turbulent flow occurs, that solute transport is only two-dimensional, that the matrix is impermeable, and that "a uniform concentration profile exists across the width of the fracture as a consequence of diffusive mixing in the fracture branches".[47] True to the process orientation, "these idealizations are not viewed as limiting for the particular issues addressed in this paper".[48]

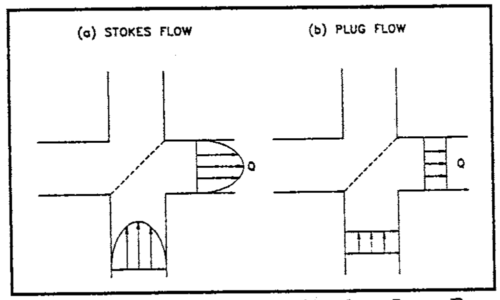


Figure 6.5: a) Stokes flow; b) plug flow. From Berkowitz, *et al.*, (1994).

In reviewing the work to date, Berkowitz, *et al.*, quantify the Peclet numbers for previous experimenters, and use these to lump considerations that have been attacked separately before - aperture width and flow velocity. Use of the Peclet number emphasizes that it is the ratio of the two parameters that is critical, and introduces a further measure of generality to results, since absolute velocity and aperture width are no longer determining.[49] The mixing behavior of a multitude of diverse systems can thus be categorized according to the ratio of advection and diffusion expressed by the Peclet number. Berkowitz, *et al.*, link calculated Peclet numbers to previously reported results: "Wilson and Witherspoon (1975) is > 10,000; for Hull and Koslow (1986), 300 to 3000; and for Robinson and Gale (1990), > 3000. Because these experiments were carried out at relatively high flow rates, it is not surprising that they observed stream tube routing at fracture intersections"[50]. In other words, such high Peclet numbers indicate that advection should dominate diffusion, which is exactly what the previous experiments showed.

Berkowitz, *et al.*, begin by synthesizing a fracture network and calculating the local Peclet number at each fracture junction, in order to get some idea of a natural range of values for *Pe*.[51] A range of Peclet numbers from 10^{-6} to 1 results from the parameter values chosen for this trial, a range far lower than those employed by previous researchers. In theory, a Peclet number of 1 indicates that advective and diffusive transport are of equal importance in a given system; it may therefore be of particular interest to investigators. The system simulated by Berkowitz, *et al.*, corresponds to a mean hydraulic conductivity of 10^{-9} m/sec, a low value typical of a porous medium composed of shale or glacial till. Despite their focus on single junctions, the authors point out that: "No single value may be representative of the mixing process... when the network is viewed at the large scale" that is ultimately of interest.[52] After this initial network simulation, the remainder of the work by Berkowitz, *et al.*, is devoted to describing the mixing at single junctions associated with values of *Pe* considerably lower than those examined by previous workers.

In their most detailed analysis, Berkowitz, *et al.*, simulate transport through a fracture junction that has been divided into a computational grid 240 x 240. This discrete network approach represents an evolution from the continuum models used previously.[53] Separate algorithms determine advective and diffusive movement at each node. Flow and transport equations are solved at each of these nodes to determine the distribution of flow and solute to each of the outlet channels. For this purpose, the velocity is obtained by solving the analytical stream function solutions of Philip (1988) and linking the results to a particle-tracking algorithm within the numerical code. 5000 particles are released in a typical particle-tracking simulation; each channel has an "exit gate" at which exiting particles are counted to determine the distribution.[54] Further indication of the importance ascribed to Philip's work is seen in the expression of outflow as a ratio of one outlet branch to another. Simulations are run for different ratios of outflow volumes - 20/80, 50/50, and 80/20 - for both Stokes and plug flow. The effects of flow type are limited to small changes in the microscale velocity vectors, even though: "Stokes flow is a more general description of the flow field within a fracture".[55] For purposes of evaluating their numerical models, Berkowitz, *et al.*, arbitrarily use the velocity of a particular outflow branch to calculate the junction Peclet number. Results are reported for Peclet numbers as low as 10^{-4}, and as high as 10^{3}.

Berkowitz, *et al.*, reach several conclusions from their simulations. Even though diffusion is recognized to control particle distribution at Peclet numbers less than 10^{-2}, "In this diffusive regime, particle movement into the outflow branches is [still] not given by a complete mixing ratio"; and hence: "As a general observation the concept of complete mixing within a fracture intersection does not properly represent the mass transfer process at any value of the Peclet number".[56] Mixing is not very sensitive to differences in Stokes and plug flow, but the Peclet number at which advection dominates diffusion varies somewhat with flow type and outflow ratios. These different scenarios still typically result in a 2:1 solute distributions at small Peclet numbers (see Figure 6.6a,b). Berkowitz, *et al.*, advance a brief argument based on geometry and probability to account for the absence of complete mixing even at very low Peclet numbers. Despite the absence of complete mixing, their work does show a mixed-mechanism zone in which diffusion contributes to mass transfer: "There is a range in the Peclet number, of roughly 3 orders of magnitude, where mixing varies as a function of the interplay between advective and diffusive mass transfer. The extent of mixing is dependent upon the value of *Pe* within the intersection".[57] Finally, Berkowitz, *et al.*, provide a general estimate of the lower bound on Peclet numbers that support the streamline routing assumption: "Streamtube-routing models provide a good approximation for Peclet numbers greater than 1; and in some cases this limit may be as low as 10^{-2}." Their findings are thus completely consistent with (if somewhat tangential to) the experimental findings of Hull and Koslow, and Robinson and Gale. Berkowitz, *et al.* also translate their conclusions into an "initial indication" of operational conditions: "A histogram of the local Peclet number should be calculated prior to a transport simulation... it is suggested that if the effective permeability of the fracture network is greater than 10^{-13} m^2 for the hydraulic gradient, then a streamtube-routing model

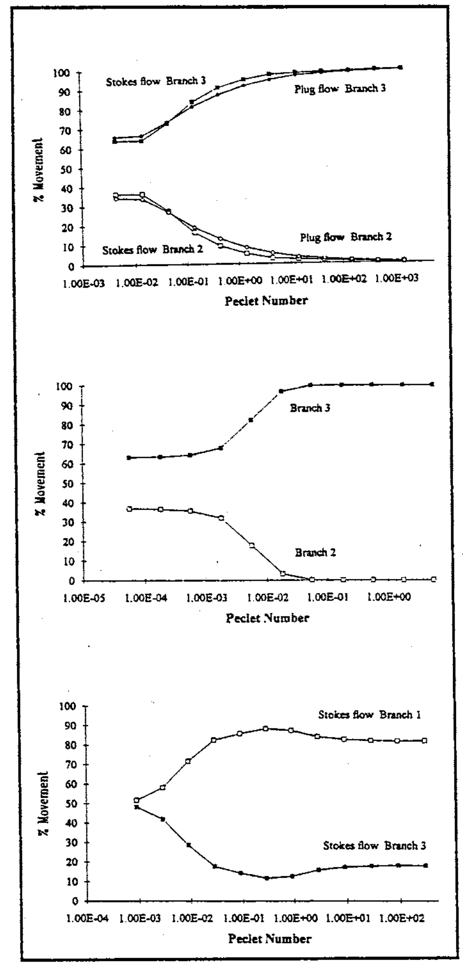


Figure 6.6: Numerical results from Berkowitz, *et al.* (1994) for mixing at fracture junctions. a) continuous, 50/50 outflow; b) continuous, 80/20 outflow; c) discontinuous, 80/20 outflow.

can be applied at the fracture intersections".[58]

In summary, despite the conclusions of Robinson and Gale, Berkowitz, *et al.*, included a range of mixing behavior in their numerical simulations. Streamline routing and complete mixing were taken as endmembers of a "more complete analysis" that allowed for diffusion within continuous fracture junctions. (They also briefly consider discontinuous junctions; see Figure 6.6c). Even though simulations showed a mixed-mechanism zone, results tended to further discredit the idea of complete mixing, even at very low flow velocities neglected by earlier workers. Pointing out that their numerical method is not practical at network scale, Berkowitz, *et al.*, exited their report with the comment: "There is a need to develop an efficient computational procedure that properly represents mass transfer at the fracture intersections". The beauty of process research is that there is often someone listening who is willing to take up such a challenge.

6.5 Additional Single Fracture Conjectures

Despite the tide of findings moving away from the idea of complete mixing at fracture junctions under any conditions, Li Chunhong, John L. Wilson and Paul Hofmann took a revisionist tack in a series of preliminary reports in 1993.[59] This work again examines flow through an orthogonal channel junction (see Figure 6.7). Li, *et al.*, agree with the reported results of other investigators that at high velocities and high Peclet numbers there will be little mixing across the dividing streamline - a situation we term *hydrodynamic control*. At small velocities and Peclet numbers, however, Li, *et al.*, still expect diffusion to dominate and result in complete mixing, this despite earlier reports purporting to rule out significant diffusion within junctions. The early reports mention mixing within single pores in porous media, but quickly focus on a discussion of mixing within fracture junctions. As is clear from comments within these preliminary reports, Li, *et al.*, treat mixing behavior as a basic question that, when answered, will allow applications to flow within different sorts of networks. Unlike some of the other investigators, they does not consider these questions settled.

Like Berkowitz, *et al.* (1994), Li, *et al.*, are interested in relatively slow-flow events within the fracture junction itself. The forced downstream mixing and very high flow rates of Robinson and Gale give results that are inconclusive for this purpose. The relevance of the somewhat lower Peclet number experiments performed by Hull and Koslow is also slight: "In their junction experiments... volumetric flow rates and velocities were very high, higher than might be expected in most real groundwater systems".[60] Moreover, the fracture apertures employed by Hull and Koslow are relatively wide. Within the conceptualization of the system held be Li, *et al.*, the high Peclet numbers of earlier work indicate exactly the dominance of advection observed by Hull and Koslow and confirmed by Robinson and Gale. Their experimental results do not, therefore, argue very forcefully against the hypothesis of complete mixing under certain other conditions. It is possible that these earlier experimentalists had - in Bacon's terms - unnecessarily framed a no-mixing interpretation "only to the measure of those particulars from which it is derived". The inclusion of low Peclet number simulations by the numerical study of Berkowitz, *et al.* poses a different set of questions than previous experimental results. The new research program is intended to pursue both numerical and physical experiments at the lower Peclet numbers considered more representative of real systems.

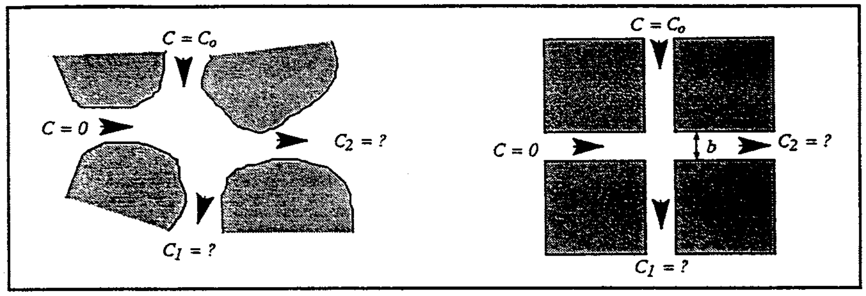


Figure 6.7: "Natural" and idealized pore bodies or fracture junctions. From Wilson, *et al.* (1993).

Assuming for the moment that the complete mixing hypothesis can be substantiated, the operational interest of Li, *et al.*, is in defining the transition zone between the two mixing end-members. If the transition zone is narrow, a simple two-rule numerical code would be adequate; either a no-mixing streamline algorithm (hydrodynamic control) or a complete-mixing algorithm (diffusion control) could be used as needed. If the transition zone is broad and encompasses a significant portion of the reasonable Peclet number distribution, then an adaptive mixing rule is required to estimate dispersion and concentrations within networks, calculated from local conditions at each junction within the network.

Again like Berkowitz, *et al.*, Li, *et al.*, seek a "more complete analysis" that defines the regimes within which one or the other of the mixing rules or a combination of the two applies. Despite the ten orders of magnitude over which the Peclet number might vary, it doesn't seem unreasonable 1) to expect or conjecture that a third mechanism might govern the possibly broad middle ground; nor 2) to hope that this third mechanism might be robustly enough described to encompass the two extant theories as endmembers - special cases of the new theory.[61] A continuum of mixing rules is therefore hypothesized that might apply at intermediate values of the Peclet number, and expected to be centered somewhere near $Pe = 1$ (see Figure 6.8). This is the "larger and wider axiom" that Bacon urges we confirm by observing any "new particulars" it may indicate.

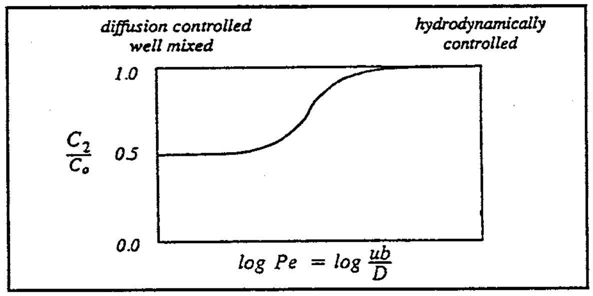


Figure 6.8: Hypothesized continuum of mixing rules within the idealized pore body or fracture junction of Figure 6.7. From Wilson, *et al.* (1993).

If the transition zone is narrow, or the transition Peclet numbers improbably low, then the importance of any new particulars is minimized. Previous investigators, such as Hull, *et al.*, and Berkowitz, *et al.*, had already considered the idea that there is a continuum of mixing rules. The latter found a simulated transition zone extending over about three orders of magnitude of *Pe*. They located the upper limit of the transition zone at about $Pe = 1$; in some cases this limit was reduced to $Pe = 10^{-2}$. As described above, they did not find less than a 2:1 mixing ratio under any conditions. To date there was neither experimental nor numerical support for significant mixing under any conditions. These questions were first approached with another numerical simulation resting on a more fundamental physical basis.

As an alternative to continuum and discrete fracture network codes, flow and transport within a fracture junction can be simulated using Lattice Gas Automata (LGA). LGA is a particle tracking method that relies on the conservation of mass and momentum. It is thus a fundamental theoretical approach to fluid flow and solute transport. Instead of relying on a governing equation for the flow field, LGA incorporates arrays of discrete variables that follow interaction rules. The basic

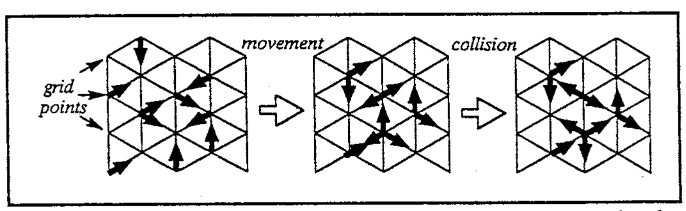


Figure 6.9: Typical lattice gas rules and evolution (two and three body rules only). From Wilson, *et al.* (1993).

assumption is that particles are in a local equilibrium state and there are no correlations between them. Particles are initially distributed at random within a hexagonal lattice to simulate a fluid. The basic physics of momentum is expressed as collision rules (see Figure 6.9), and particles are re-distributed accordingly. Solute particles have mass but do not react with the other particles; their progress is instead determined by the local momentum field. Solutes do not affect the flow field; both flow and transport can thus be simulated simultaneously. Macroscopic fluid phenomena can be recovered by calculating average fluid densities and velocities over suitable regions of the lattice.[62]

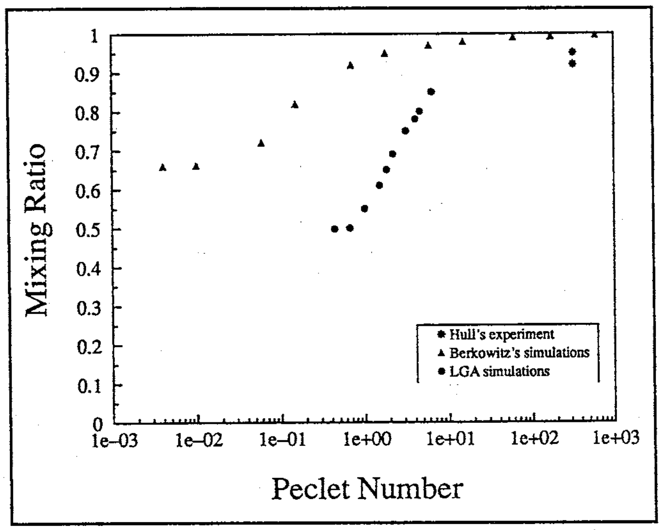


Figure 6.10: Numerical (LGA) simulation results plotted against the numerical simulation of Berkowitz, et al. (1994) and the low end of the experiments of Hull and Koslow (1986). Courtesy of Li Chunhong.

Results of these simulations are shown in Figure 6.10. Since there are currently numerical obstacles to simulating high velocity fields with LGA, the upper asymptotic curve is incomplete. It nevertheless suggests that a transition zone combining advective and diffusive control occurs over no more than three orders of magnitude of *Pe*, perhaps as little as two orders of magnitude. The results clearly differ from those of Berkowitz, *et al.*, in two important respects. The upper limit of the LGA transition zone appears to be between 10 and 100, not *Pe* = 1; this has implications for when an adaptive mixing rule is needed. Most striking is the LGA claim that complete mixing will in fact occur as *Pe* approaches unity (there was therefore no point in running simulations at lower Peclet numbers). The numerical results thus pose a significant challenge to the simulations of small Peclet number behavior reported by Berkowitz, *et al.* Since no one had as yet run physical experiments to observe mixing behavior within analogous physical systems, there was no obvious means of judging the two simulations.[63] The experiments of Hull and Koslow could be taken to support either of the numerical simulations (see Figure 6.10).

Sources of error (if error it is) are often difficult to isolate within nested numerical strategies. The mutually supportive roles played by theory (simulation) and experiment are sometimes particularly clear in the presence of competing results. In this, process research generally enjoys an advantage over applied work. For example, after Galileo (still in the guise of Salviati, see p.126) has produced a idealized geometric demonstration of uniformly accelerated motion, his interlocutor Simplicio reprizes his earlier reservations: "But I am still doubtful whether this is the acceleration employed by nature in the motion of her falling heavy bodies".[64] We find our practical fears concerning the physical significance of idealized (and unsupported) theory clearly expressed some 350 years ago. On this occasion, Salviati offers to describe the pertinent confirming experiments, conceding to Simplicio that

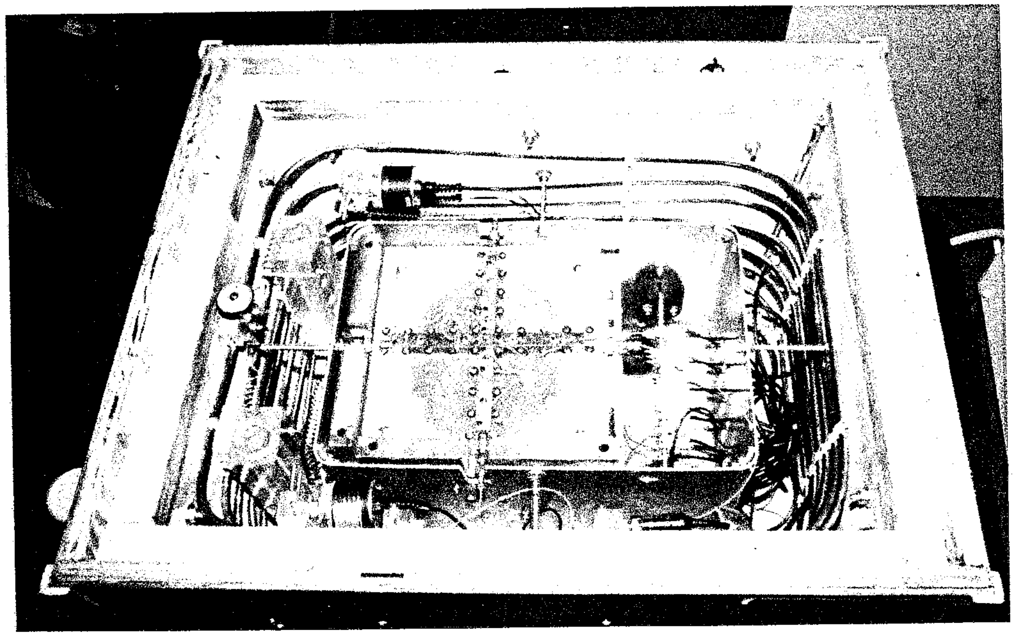
> Like a true scientist, you make a very reasonable demand, for this is usual and necessary in those sciences which apply mathematical demonstrations to physical conclusions, as may be seen among writers on optics, astronomers, mechanics, musicians and others who confirm their principles with sensory experiences, those being the foundations of all the resulting structure... Therefore, as to the experiments: The Author has not failed to make them...[65]

6.6 Experimental Confirmation

Our first acrylic model used a method of quantifying mixing similar to that described by Hull and Koslow, and depended on measuring electrical conductivity at 12 points in the model (see Figure 6.11a). Channel dimensions were 203 x 1400 microns deep; channels were about 10 cm long. The numerical limitations of LGA in simulating high velocity flow fields have been mentioned above; the early physical experiments, on the other hand, worked well at high Peclet numbers, but had problems at low Peclet numbers; attempts to calibrate electrical conductivity signals to standard solute solutions were marred by signal drift and leaking. These are obviously critical concerns since low Peclet number experiments require several days to run; Hull and Koslow report mass balance errors of up to 10% in their relatively rapid experiments.

A second acrylic model has abandoned the idea of collecting transient channel concentration data electronically in favor of simply collecting discharged effluent for analysis by High Pressure Liquid Chromatography (HPLC) (see Figure 6.11b). Elimination of the platinum probes greatly reduces

a) The original model attempted to measure electrical conductivity, and used one pump with a flow splitter.



b) the final model produced effluent to be analyzed by HPLC, and used two independently calibrated pumps.

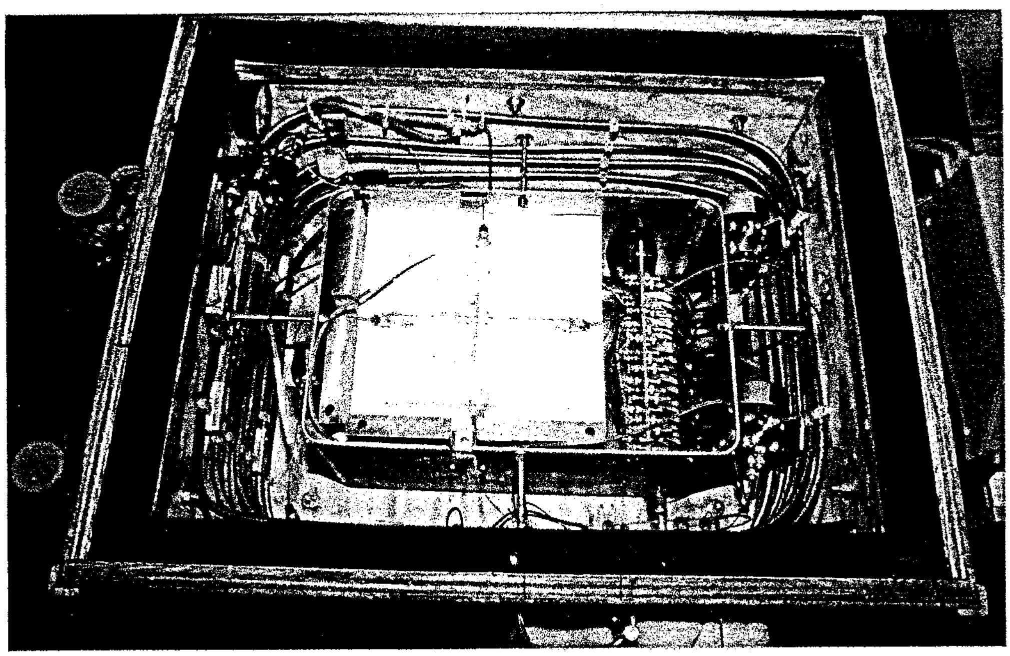


Figure 6.11: The acrylic fracture mixing models, inside the constant temperature box. a) the original; b) the final.

perforations in the model and the potential for leaking. Adjustments to channel dimensions allow a higher flow rate at the same Peclet number, resulting in more reliable pump discharge.[66] Painstaking attention to both pump and pressure transducer calibration has sought to reduce the physical experiment to the level of the perfectly controlled numerical experiment. Wild nighttime temperature swings in the laboratory - that could affect both flow rates and diffusion coefficients - were finally discovered and controlled as well.[67]

Experimental data is plotted in Figure 6.12, along with the simulations and the low end of the experiments by Hull and Koslow, giving strong support to the *Velveteen Rabbit Effect* of crucial demonstrations: the convenient but merely hypothetical phenomenon of complete mixing has become real. Regarding the transition zone, the accuracy of the simulations is especially impressive since they were constructed without the fitting advantage of experimental data. This suggests that the principles incorporated into the LGA methods may be a reliable analogy to actual events. (As usual, the numerical results can be readily adjusted to reflect data).

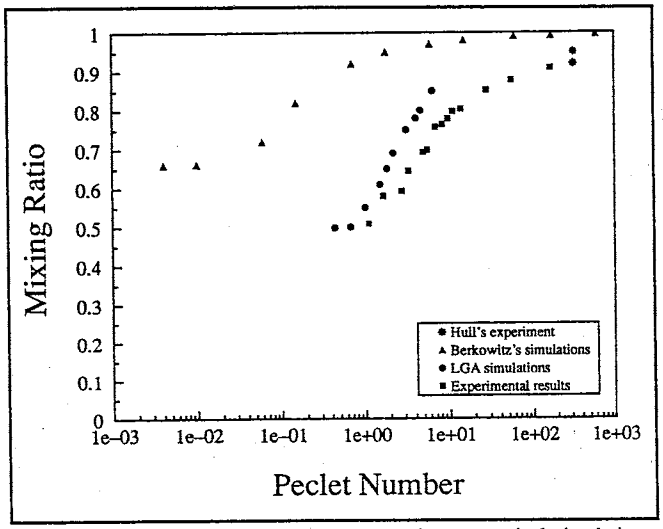


Figure 6.12: Experimental results contrasted to numerical simulations. Courtesy of Li Chunhong.

The heretofore evanescent (virtually-)complete mixing has now been demonstrated at realistic Peclet numbers, but a new puzzle has appeared: Why hasn't anyone ever reported anything like no-mixing? The new experiments appear asymptotic to about a 0.9 mixing ratio. Not even the previous experiments run at high Peclet numbers avoided mixing altogether. Several suggestions have been advanced to account for this phenomenon, including experimental flaws (such as imperfections in model geometry, or unsteady flow rates), or other unrecognized processes at work (such as eddies or other circulation patterns within the junction). The latter suggestions reflect the sense of Hacking's comment that: "Although it also has a precise meaning, [experimental] *noise* often means all the events that are not understood by any theory".[68]

The consideration of these secondary effects reminds us that the investigation of solute mixing at fracture junctions is hardly brought to a close by a demonstration of complete mixing and a significant transition zone under certain conditions. Many factors carefully isolated or excluded from extant studies may exert significant effects on mixing. These include unequal inlet flow rates, surface roughness, variable angles of intersection and, most generally, non-idealized systems in which these parameters naturally vary. On another level altogether, the practical application of theoretical or experimental insights clearly requires a great deal of information on the orientation and density of natural fractures that is difficult to obtain. Among others, Shapiro and Nicholas (1989), Berkowitz, *et al.* (1994), and Y.W.Tsang (1992) point out that the parallel plate analysis common to all of the studies surveyed here is not realistic, complicating the more important transition to practical problems.[69] Ultimately, the most important research question regarding mixing at fracture junctions may be how to scale up from bench-scale work to the systems of practical interest; *i.e.*, how to recover macroscopic behavior from more or less controlled experiments.

6.7 Conclusion

In discussing paleontology, Stephen Jay Gould says: "Puzzles mount on puzzles the more we consider details". There is a danger that - in the pursuit of detail - process research can become caught up in what David Pramer calls *terminal research*. Work becomes terminal, he says, when it leads only to more research, never approaching conclusions with practical applications. Robert Peters has leveled a similar criticism at much of ecology.[70] The *few significant facts* for which the historian clamors may not be few after all, either within large hydrologic systems of practical interest or even for very small and seemingly simple systems. Where this is the case, the Galilean program of decomposition, inspection and reassembly cannot quickly offer enough insight to guide practical choices. Nevertheless, progress is marked by experimental and theoretical benchmarks. In the case reviewed here, complete mixing eventually ceased to be something merely convenient or theoretical; it has become experimental - *real*. The immaturity of a science may be measured by the distance from hard-won benchmarks to waiting applications.[71]

Although Peter Medawar chooses to associate crucial experiments with Galileo, at this point

Francis Bacon's comments are more revealing. Bacon's taxonomy of experiments includes one type he calls *instantiae crucis*. This is sometimes rendered as "crucial experiment", but early translators (such as the 1939 Modern Library edition used for the most part here) sometimes give it as "instances of the fingerposts",[72] after the stylized roadsigns of the time. As Ian Hacking points out, such experiments are not viewed by Bacon as necessarily decisive, though they "afford very great light, and are of high authority, the course of interpretation sometimes ending in them and being completed." Hacking comments on this passage:

> Bacon is truer than the more recent idea... Bacon claimed only that crucial instances are *sometimes* decisive. It has recently become fashionable to say that experiments are only crucial with hindsight, that they never decide anything at the time. Imre Lakatos says just that. Hence a false confrontation has arisen. Had philosophers stuck with Bacon's good sense we might have avoided the following pair of contraries: (a) Crucial experiments decide decisively, and lead immediately to the rejection of one theory; (b) "There have been no crucial experiments in science" (Lakatos[73]). Certainly Bacon disagrees with Lakatos, and rightly so, but he also dissents from (a).[74]

Bacon is not so anxious to rule out either turning of the road, even in a crucial experiment; for one thing, he recognizes that roads that meet at an intersection and force a temporary choice of directions might still meet again further on. In this sense a maturing hydrologic collection of instances of the crossroads may fail to be decisive in the ways insisted upon by legal constraints, and yet be *compelling* to investigators looking into a particular model/prototype analogy. Although we recognize the powerful utility of mathematical derivations (see Chapter 5) and consider them essential in any account of process-oriented progress, an immature science still gravitates to the view of Bacon. Referring to controlled physical experiments, Bacon called them "*Striking Instances*",

> which I also call *Shining Instances*, or *Instances Freed and Predominant*. They are those which exhibit the nature in question naked and standing by itself, and also in its highest exaltation or highest degree of power; as being disenthralled and freed from all impediments, or at any rate by virtue of its strength dominant over, suppressing and coercing them. For since every body contains in itself many forms of natures united together in a concrete state, the result is that they severally crush,

depress, break and enthrall one another, and thus the individual forms are obscured. But certain subjects are found wherein the required nature appears more in its vigor than in others, either through the absence of impediments or the predominance of its own virtue. And instances of this kind strikingly display the Form. At the same time in these instances also we must use caution, and check the hurry of the understanding...[75]

6.8 Notes

1. McMullin, E. (1985), "Galilean Idealization", in *Studies in the History of the Philosophy of Science*, **16**:3, p.251.

2. Gould, S.J. (1989), *Wonderful Life: The Burgess Shale and the Nature of History*, Norton, p.59.

3. McMullin, E. (1985), p.264-5: "The really troublesome impediments, Galileo said more than once, are the causal ones. The unordered world of Nature is a tangle of causal lines; there is no hope of a 'firm science' unless one can somehow simplify the tangle by eliminating, or otherwise neutralizing, the causal lines which impede, or complicate, the action of the factors one is trying to sort out". The influence of these "impediments" is, of course, reduced or calculated by means of controlled experiments.

4. See, *eg.*, Kuhn, T.S. (1962), *The Structure of Scientific Revolutions*, University of Chicago Press, 2nd Edition (1970), p.164.

5. Keller, E.F. (1983), *A Feeling for the Organism: The Life and Work of Barbara McClintock*, W.H. Freeman & Co., p.9.

6. Keller, E.F. (1983), p.6.

7. National Academy Press, (1991), *Opportunities in the Hydrologic Sciences*, p.78. *Fractured media* are thus distinguished from *porous media* for which continuum assumptions can more reasonably be made.

8. National Academy Press (1991), p.78.

9. Viewed in two dimensions, the streamtube is just the area between two streamlines.

10. See Berkowitz, *et al.*, (1994), p.1766: "Complete mixing... is straightforward to implement within a network-scale simulation model because it is not necessary to consider the details of the flow pattern within the intersection... A streamtube-routing model requires that attention be paid to the relationships between the stream tubes carrying mass into the intersection and the distribution of the stream tubes among the possible outflow branches". The simplicity of complete mixing algorithms no doubt accounts in part for the popularity of such models; much of the work recounted in this historical overview relates to the requirements identified for streamline routing.

11. Anderson, M.P. (1983), "Ground-Water Modeling: The Emperor Has No Clothes", in *Ground Water*, **21**:6, pp.666-669. Referring to Hans Christian Andersen's tale of the Emperor's New Clothes, she says: "Models may be divided into ground-water flow models and contaminant transport models. In a ground-water flow model, we are fortunate in that the theoretical framework upon which the model is built has been well verified and the physical meanings of the parameters are well understood. In this case the model (Emperor) has a good set of underwear and the directions for tailoring his clothes are clear" (p.667). The present constrasting comment follows.

12. Keller, E.F. (1983), p.6.

13. Hull, L.C. and Koslow, K.N. (1986), "Streamline Routing Through Fracture Junctions", in *Water Resources Research*, **22**:12, pp.1731-1734.

14. Hull, L.C. and Koslow, K.N. (1986), p.1731.

15. Smith, L. and Schwartz, F.W. (1984), "An Analysis of the Influence of Fracture Geometry on Mass Transport in Fractured Media", in *Water Resources Research*, **20**:9, pp.1241-1252.

16. Smith, L. and Schwartz, F.W. (1984), p.1242.

17. Pe = vb/D, where $Pe \equiv$ the Peclet number; v $\equiv$ average fluid velocity in the intersection; b $\equiv$ a characteristic length of the system, often taken to be the radius of the fracture intersection, measured diagonally from the center; and D $\equiv$ the diffusion coefficient of the solute of interest at the given concentration. The Peclet number is thus a dimensionless measure of the relative importance of advective versus diffusive transport in a given system. The Peclet number range for these experiments is given by Berkowitz, *et al.*, (1994), p.1766.

18. Hull, L.C. and Koslow, K.N. (1986), p.1734.

19. Hull, L.C., Miller, J.D. and Clemo, T.M. (1987), "Laboratory and Simulation Studies of Solute Transport in Fracture Networks", in *Water Resources Research*, **23**:8, pp.1505-1513.

20. Hull, L.C., *et al.* (1987), p.1511.

21. Hull, L.C., *et al.* (1987), p.1512. The relation of flowrate and aperture width is seen directly in the Peclet number: Pe = vr/D. At higher flow rates, the Peclet number can be held steady (and the relative importance of advection and diffusion) by reducing the aperture width. As apertures widen, a slower flow accomplishes the same thing.

22. Philip, J.R. (1988), "The Fluid Mechanics of Fracture and Other Junctions", in *Water Resources Research*, **24**:2, pp.239-246.

23. Philip, J.R. (1988), p.242.

24. Laplacian plug flow: $\nabla^2\phi = 0$. Stokes flow: $\nabla^4\phi = 0$.

25. Philip, J.R. (1988), p.245.

26. Philip, J.R. (1988), p.245.

27. Philip criticizes the method employed by Hull and Koslow in reporting deviations from their expected outlet concentrations: see Philip, J.R. (1988), p.243.

28. Hacking, I. (1983), *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*, Cambridge University Press, p.271.

29. Robinson, J.W. and Gale, J.E. (1990), "A Laboratory and Numerical Investigation of Solute Transport in Discontinuous Fracture Systems", in *Ground Water*, **28**:1, pp.25-36 (p.26).

30. Robinson, J.W. and Gale, J.E. (1990), p.35.

31. The Peclet number can be calculated using the one-dimensional cubic law:

$$\bar{v} = \frac{b^2}{12}\left(\frac{g}{\nu}\right)\frac{\partial\phi}{\partial x}$$

where $\bar{v}$ $\equiv$ mean velocity; b $\equiv$ aperture radius; $\nu \equiv$ kinematic viscosity; and $\partial\phi/\partial x$ is the hydraulic gradient. Then $Pe = \bar{v}$ b/D. Using the information given in Hull and Koslow (1986) for the smallest aperture and gradient, $Pe \geq 3075$. Fracture geometry is described on p.26, and hydraulic gradients on p.27; the description indicates that every junction contained at least one aperture 0.5mm wide, so the minimum Peclet number pertains only to particular branches of certain junctions. Absolute volumetric flow rates can be calculated from $Q = \bar{v}$ (2bd), where d $\equiv$ depth of the channel; and 2bd $\equiv$ the cross-sectional area normal to flow. Finally, an effective permeability could then be calculated from Darcy's law.

32. See their figure 3, p.27. They define their term as: "Individual fractures are discontinuous within their own planes" (p.25).

33. Robinson, J.W. and Gale, J.E. (1990), p.28.

34. Lower gradients result in lower Peclet numbers for a given geometry and solute concentration; if these results were significant, they should show more diffusive mixing, not less.

35. Robinson, J.W. and Gale, J.E. (1990), p.28.

36. Robinson, J.E and Gale, J.E. (1990), p.29.

37. Robinson, J.W and Gale, J.E. (1990), p.35. See also Gale, J.E. (1987), "Comparison of Coupled Fracture Deformation and Fluid Flow Models With Direct Measurement of Fracture Pore Structure and Stress Flow Properties", *29th Symposium on Rock Mechanics*, Tucson AZ, suggests this is true of real fractures that are "intimately pressed together" (p.35).

38. Robinson, J.W. and Gale, J.E. (1990), p.32.

39. Robinson, J.W. and Gale, J.E. (1990), p.28.

40. Robinson, J.W. and Gale, J.E. (1990), p.31. Smith, *et al.* (1985) builds on the work presented in Smith and Schwartz (1984) mentioned at the outset.

41. Robinson, J.W. and Gale, J.E. (1990), p.29.

42. Local hydrogeological conditions have forced Sweden to look into repositories located in saturated, fractured media.

43. Robinson, J.W and Gale, J.E. (1990), p.35.

44. Berkowitz, B., Naumann, C. and Smith, L. (1994), "Mass Transfer at Fracture Intersections: An Evaluation of Mixing Models", in *Water Resources Research*, **30**:6, pp.1765-1773.

45. Berkowitz, B., *et al.* (1994), p.1766.

46. Berkowitz, B., *et al.* (1994), p.1766.

47. Berkowitz, B., *et al.* (1994), pp.1766, 1769.

48. Berkowitz, *et al.* (1994), p.1766.

49. For example, decreases in velocity can be balanced by increases in aperture width, and vice versa (see note 21, above). Berkowitz, *et al.* (1994), p.1768, comment: "The partitioning of mass to the two outlet branches is transparent to the actual values assigned to fracture aperture, volumetric flow, or the diffusion coefficient, provided results are compared at the same value of the Peclet number".

50. Berkowitz, *et al.* (1994), pp.1766-1767.

51. Berkowitz, *et al.* (1994), p.1767. Hydraulic gradient was a uniform 0.005 (*cf* Robinson and Gale: smallest hydraulic gradient was 0.33). The mean aperture width was 40 microns; the diffusion coefficient was a uniform 1.2 x 10^{-9} m^2/sec.

52. Berkowitz, B., *et al.* (1994), p.1768.

53. See, *eg.*, Berkowitz, B., Bear, J. and Braester, C. (1988), "Continuum Models for Contaminant Transport in Fractured Porous Formations", in *Water Resources Research*, **24**, pp.1225-1236.

54. Berkowitz, B., *et al.* (1994), pp.1769-1770.

55. Berkowitz, B., *et al.* (1994), p.1768.

56. Berkowitz, B., *et al.* (1994), p.1771.

57. Berkowitz, B., *et al.* (1994), p.1771.

58. Berkowitz, B., *et al.* (1994), p.1772.

59. Li, C., Wilson, J.L. and Hofmann, P. (1993a), "The Influence of Pore Size on the Diffusion Coefficient Inside Porous Media: A Pore Scale Numerical Experiment", Proceedings of the Society for Industrial and Applied Mathematics *(SIAM) Conference on Mathematical and Computational Issues in the Geosciences*. Houston, Texas.

    Li, C., Wilson, J.L. and Hofmann, P. (1993b), "Fickian and Non-Fickian Phenomena in Porous Media: A Microscopic Numerical Experiment", in H.J.Morel-Seytoux (ed.): Proceedings of the 13th Annual American Geophysical Union, *AGU Hydrology Days*, Fort Collins, CO, pp.149-160.

    Wilson, J.L., Li, C. and Hofmann, P. (1993), "Mixing Rules of Pore and Fracture Junctions", in Proceedings of *3rd Annual WERC Technology Development Conference*, Waste Education and Research Consortium, Las Cruces, NM.

    My own involvement was limited to building some of the laboratory apparatus and minor assistance with the experimental confirmations. Details of the numerical simulations and experimental procedures can be found in Li, C. (forthcoming in 1995), "Solute Mixing at Fracture Junctions: Low Peclet Number Simulations and Experiments", Ph.D. dissertation, New Mexico Institute of Mining and Technology, Socorro, NM.

60. Wilson, J.L., *et al.* (1993), p.2.

61. The empiricist quantum chemist Lowdin mentioned in Chapter 5 (p.155) says: "In reality all 'theories' are tools for correlating one set of experimental data with another set, and theories of various types are essentially unalike in that they correlate data of different orders of magnitude" (Lowdin, P-O. (1967), "Nature of Quantum Chemistry", in the *International Journal of Quantum Chemistry*, **1**:1, p.10).

62. A structural explanation of Darcy's law that utilizes LGA was reported by Balasubramanian, K., Hayot, F. and Saam, W.F. (1987), "Darcy's Law From Lattice-Gas Hydrodynamics", in *Physical Review A*, **36**:5, pp.2248-2253. It may also be noted in passing that the authors make no effort to keep quiet the shade of M.King Hubbert, as they repeatedly state Darcy's law as $\mathbf{v} = -A\, \nabla p$, and

$u_y(x) = -(1/\rho\alpha)(dp/dy)$.

63. We met Nevill Mott earlier in our discussion of Darcy's law (p.174), as he argued for a middle ground utilizing both theoretical derivations and experiment; his interest is clearly directed, however, as he continues: "I can hardly doubt that some of these concepts may be falsified over again, but I hope they will suggest worthwhile experiments. That is, after all, what theories are for" (see Hofmann, J. (1990), p.416). Note contrast to the research reported in Glen, W. (1982), *Road to Jaramillo*, Stanford University Press, p.26: in the early days of Potassium-Argon dating, the dating of older rocks was uninhibited by alternative explanations, since no one else had any idea.

64. Galileo, G. (1638), *Two New Sciences*, Drake, S. (trans., ed.), Wall and Thompson, Toronto, 2nd edition (1989), p.169.

65. Galileo, G. (1638), p.169. The inclined plane experiments are described on pp.169-170. See also McMullin, E. (1985), pp.265-267, in which this exchange between Salvati and Simplicio is considered at length and in historical and philosophical context. Salviati is at pains to minimize the "impediments" in his experiment by using "a very hard bronze ball, well-rounded and polished"; the surface of the inclined plane itself is "vellum as much smoothed and cleaned as possible". Simplicio regrets missing these demonstrations, but is content to rely on Salviati's report.

66. Channel width is 500 microns. A much deeper channel meant that $Pe < 1$ could be had with much higher volumetric flow rates due to greater cross sectional area of the junction.

67. The details of the experimental struggles can be found in Li Chunhong's dissertation. Whatever else these experiments showed, they did nothing to disprove what Thomas Kuhn said: "The operations and measurements that a scientist undertakes in the laboratory are not "the given" of experience but rather "the collected with difficulty" (*The Structure of Scientific Revolutions*, p.126). Li's struggles in this regard also bring to mind Hacking's observation that "Often the experimental task, and the test of ingenuity or even greatness, is less to observe and report, than to get some bit of equipment to exhibit phenomena in a reliable way" (*Representing and Intervening*, p.167); or: "Good experimenters guard against the absurd" (p.269).

68. Hacking, I. (1983), p.265. If the acrylic model is marred by a small asymmetry, the data might be reflecting a systematic effect. Preliminary flow visualizations appeared to rule out this possibility, although inspections continue. Visualizations of the dividing streamline gives further confirmation of experiments dating back to Wilson and Witherspoon (1972), but more importantly, suggests that the complete mixing hypothesis is fairly robust in the presence of imperfect geometry. The transport of both dyes and colloidal particles was visualized. No circulation was evident at moderate flow and very high Peclet numbers. The flow rate was the same as at $Pe = 300$ with the original solutes, but Peclet number is vastly increased due to extremely small diffusion coefficient of the one micron particles.

69. Shapiro, A.M. and Nicholas, J.R. (1989), "Assessing the Validity of the Channel Model of Fracture Aperture Under Field Conditions", in *Water Resources Research*, **25**:5, pp.817-828: "Usually, a single fracture is modeled as the space between two parallel plates. Recent investigations, however, have demonstrated that a parallel plate fracture is not consistent with field and laboratory observations of fluid and solute movement... Consequently, there have been several recent modifications... that attempt to describe the heterogeneity in the fracture aperture" (p.817). Tsang, Y.W. (1992), "Usage of 'Equivalent Apertures' for Rock Fractures as Derived From Hydraulic and Tracer Tests", in *Water Resources Research*, **28**:5, pp.1451-1455: "Until the 1980s, researchers in the field of fracture flow and transport used to consider a rock fracture as a pair of

parallel plates separated a constant distance, *b*, so that the aperture of such an idealized aperture is unambiguously defined. In the last decade, increasing experimental evidence has led many researchers to revise their conceptualization of a rock fracture and now regard the fracture as having rough walls and being therefore constituted of apertures of many different values, which may be mathematically characterized by an aperture density distribution and a spatial correlation" (p.1451). Tsang also provides a table of fracture aperture terminology in an effort to clear up usage non-conformities. Berkowitz, *et al.*, (1994): "Of course, it is important to recognize that 3-D mass transfer through a network of rough-walled fractures is considerably more complex and varied than the simple geometry we analyze" (p.1766).

70. Pramer, D. "Terminal Science", in *Bioscience*, **35**, p.141. Peters, R.H. (1991), *A Critique for Ecology*, Cambridge University Press, pp.144, 185-6, uses similar language.

71. Hacking, I. (1983), pp.262-272, discusses experimental efforts regarding polarized electrons, saying: "Some 10^{11} events were needed to obtain a result that could be recognized above systematic and statistical error", but in the end: "hunches are based on a hard-won sense of the kinds of things that electrons are".

72. Bacon (1620b, *lii*, p.122-123) lists several general categories of investigations. Of these, Perogative Instances are further subdivided into 27 sub-instances, and could be: Solitary, Migratory, Striking, Clandestine, Constituitive, Conformable, Singular, Deviating, or Bordering; there were as well, Instances of Power, Instances of Companionship and of Enmity, Subjunctive Instances, Instances of Alliance, Instances of the Fingerpost, Instances of Divorce, Instances of the Door, Summoning Instances, Instances of the Road, Instances Supplementary, Dissecting Instances, Instances of the Rod, Instances of the Course, Doses of Nature, Instances of Strife, Intimating Instances, Polycrest Instances, Magical Instances. The interested reader is referred to Bacon's treatment of each. Here we are interested in Instances of the Fingerposts which "guard [the understanding] against false forms and causes". Fingerposts (*instantiae crucis*) were set up at crossroads to indicate the traveler's several options.

73. Lakatos, I. and Musgrave, A. (1970). See the discussion in Hacking, I. (1983), p.279.

74. Hacking, I. (1983), p.250.

75. Bacon, F. (1620b), *xxiv*, p.121.

7

Progress: "The Believing Spirit"

> No wonder, then, that in the early stages of the development of science different men confronting the same range of phenomena, but not usually all the same particular phenomena, describe and interpret them in different ways. What is surprising, and perhaps also unique in its degree to the fields we call science, is that such initial divergences should ever disappear.
>
> - Thomas S. Kuhn, *The Structure of Scientific Revolutions* [1]

> Our road does not lie on a level, but rises and falls; rising first to axioms, then falling to works.
>
> - Francis Bacon, *Novum Organum* [2]

7.0 Introduction

Peter Medawar notes the slippery vagueness - even inaccuracy - of scientists' own descriptions of what they do. Albert Einstein said: "If you want to find out anything from the theoretical physicists about the methods they use, I advise you to stick closely to one principle: don't listen to their words, fix your attention on their deeds".[3] It seems very likely at this point that hydrologists' allusions to Popper and the philosophy of science are more rhetorical devices than firmly held methodological rules. It remains only to synthesize what insights we have gained into the actual operative methodology of hydrologists, as a measure of confidence within an uncertain science.

A certain amount of pessimism is understandable when hydrologists attempt to analyze and/or predictably impact groundwater systems. Since models distill - and then extrapolate from - a modeler's limited understanding of the system under study, it is not surprising that predictions based on models cannot be taken at face value, nor that models can in fact be actively misleading. The complexity and idiosyncratic nature of model construction means, moreover, that a claim of objective fact for a simple output, "without further qualifying phrases", is fundamentally deceptive, particularly in the relatively complicated context of applied models. The situation worsens if we think indirect history-matching or infrequent post-audits are the only possibly significant tests of model predictions. Caution in the use of predictive models recognizes that our understanding of the subsurface environment is inadequate for certain purposes. Among these purposes are definitive assessments of many practical problems. *A priori* validation at best leads to a qualitative dependence on expert insight that cannot be entirely satisfying within a supposedly rational and quantitative regulatory environment.

The managerial implications of modeling uncertainty are most apparent in two areas. First, if hydrology is not yet a strongly predictive science, then the present fascination with "strategic research" is a matter of putting Descartes before the horse: fundamental experimental research is needed to pull this wagon. Secondly, if hydrologists may be unable to predict the course of contaminant plumes and effectively remediate them, the proper regulatory response should obviously include a far greater insistence on the prevention of contamination problems and avoid an undue reliance on cleanup projections. A similar caution should accompany water supply projections. The state of the art will often not support the closed, sequential decisions that bureaucracies favor.[4] At present, simple and final answers are often demanded for unforeseeably complicated scientific questions. To the extent that our system requires such pronouncements at every step, the system is self-defeating. Vastly different notions of truth are highlighted as, on the one hand, scientists advance plausible explanations from an admittedly inexact science, and on the other hand, possibly "uninitiated" outsiders seek to assign liability, establish and enforce regulations, or make economic decisions. A better understanding of the foundations of groundwater modeling can promote realistic expectations and compatible procedures in environmental science, engineering and law. These matters have provided the background for our study of hydrologic methodology.

Regulatory and legal pressure has had a significant impact on what would otherwise be a more disinterested, "normal", and considerably less generously funded science. Freeman Dyson voiced a related complaint among the physicists when he said that "no one has fun anymore with reactor design" since the accountants and managers took over nuclear physics,[5] a comment that calls to mind something Garrison Keillor said in another context: "It is hard to put your finger on, but guys are in trouble. Guys are gloomy...Years ago, manhood was an opportunity for achievement and now it's just a problem to be overcome".[6] It is evidently important to recognize the current limitations of the science in addressing important questions. These limitations are implicit in hydrologic methodology, although the general problem is certainly not the invention of hydrologists. Some additional historical and multidisciplinary perspective on methodology will later lend itself to a more specific discussion of hydrology.

7.1 Experiment and Theory

We began our discussion of methodology in science by recognizing the need to relate and rank four things: observation, experiment, theory and models. Absolutely simple observation is rather rare among modern day investigators, surfacing mostly in accounts of serendipitous discoveries. Ian Hacking still likes, however, the prescription of the physicist George Darwin,

> who used to say that every once in a while one should do a completely crazy experiment, like blowing the trumpet to the tulips every morning for a month. Probably nothing will happen, but if something did happen, that would be a stupendous discovery.[7]

Darwin suggests occasionally trying to evade the usually useful guidance of generally accepted theory. Independence from both theory and critical intent characterizes simple observation. Most often, intuitions and inquiries are influenced by a conceptual framework, as were our exploration of mixing behavior at fracture junctions and Darcy's experiments at Dijon. Theories are usually incomplete or wrong to some degree, however. An excessive or exclusive reliance on existing theory to frame our research programs must be at least a little myopic, even if difficult to avoid.[8]

Random observation did not figure much in our account. Needless complications were avoided when we quickly took method in hydrology to be nearly equivalent to the procedures involved in building models. *Structures* (geometry and processes) in turn defined models. A methodological inquiry into hydrology essentially became an effort to build a taxonomy and critique of hydrologic models on all scales. Models are an appropriate focal point, since they are the essential intermediary between the remaining items of interest: experiment and theory. Before we reconsider the interplay of experiment and theory within hydrologic models, a more general historical and multi-disciplinary survey of the balance between experiment and theory can help put that analysis in perspective.

We found a useful beginning in the comments of Francis Bacon, whose preoccupation with the shortcomings of Scholastic deduction led naturally to an emphasis on basic experimentation. Our review of *Novum Organum* revealed Bacon as a committed (if not altogether simple) empiricist. He underestimated the power of mathematical deduction and did not use it, though he hinted at an eventual "closer and purer league" between the experimental and the rational from which "much may be hoped". His insistence on "True Induction" included experiments as contrived experience that served specific purposes. They were not unthinking observation; they were intended to build theory. Moreover, he expected previous insights to inform new work. He did, of course, stress the need for "sufficient negatives" to ensure that the "understanding [be] hung with weights, to keep it from leaping and flying"[9] from specific observations to intermediate axioms. His description of crucial experiments (p.221-2) and the subtleties of their interpretation remains an insightful commentary. He recognized the pitfalls of too-ready immunization of tested theories; he expected that such tests might *sometimes* be decisive, and deplored the blunting effect of blaming or adding auxiliary hypotheses:

> And if some opposite instance, not observed or not known before, chance to come in the way, the axiom is rescued and preserved by some frivolous distinction; whereas the truer course would be to correct the axiom itself.[10]

Above all, Bacon elevated practical progress as the ultimate standard of scientific methodology. He opposed unwarranted inductions, spurious deductions and a neglect of foundational experiments on precisely these grounds: the "unfruitfulness of the way". The immaturity of science in his day naturally favored a methodical gathering of basic information that would both build and constrain emerging axioms.

In a comparison of quantum mechanics and biology, Lewis Thomas, medical doctor and popularizer of science, reaches a similar conclusion about both the long-term importance of theory and the immediate need for experiment within biology:

> One big difference is that biology, being a more difficult science, has lagged behind, so far behind that we have not yet reached the stage of genuine theory - in the predictive sense in which theoretical physics drives that field along. Biologists are still principally engaged in making observations and collecting facts, trying wherever possible to relate one set of facts to another but still lacking much of a basis for grand unifying theories.[11]

P.B. Medawar, also a biologist, takes a considerably leaner view of Bacon's science than we do; in particular, Medawar denies any critical attitude to Baconian experimentation. Experiments lacking an underpinning of critical theory cannot, in his view, lead to significant progress. Medawar's admiration for Karl Popper stems in part from a shared conviction that the guidance of theory and a reliance on crucial tests characterize maturing sciences. Medawar is unimpressed with Bacon's *instantiae crucis*:

> [Baconians] describe and annotate the phenomena when they are made to take place under certain well-defined and well-regulated conditions. In the meanwhile we begin to form opinions about the nature of the causal mechanisms at work and the relationship of the phenomena to others, and only critical experimentation can discriminate between them. Sciences which remain at a Baconian level of development... amount to little more than academic play.[12]

Our earlier discussion (Chapter 4) of Popper's philosophy of science repeated the common criticism that decisive tests are not so readily contrived, generally weakening Popper's logic of falsification;[13] more importantly here, such tests were shown to be particularly elusive in the context of predictive hydrologic models. Popper's view nevertheless reflects the central importance of theory - not discovery - in his logic of science. When he considers the "epistemological theory of experiment" he says:

> The theoretician puts certain definite questions to the experimenter, and the latter, by his experiments, tries to elicit a decisive answer to these questions, and to no others... But it is a mistake to suppose

that the experimenter proceeds in this way "in order to lighten the task of the theoretician", or perhaps in order to furnish the theoretician with a basis for inductive generalizations. On the contrary, the theoretician must long before have done his work... He must have formulated his question as sharply as possible. Thus it is he who shows the experimenter the way. But even the experimenter is not in the main engaged in making exact observations; his work, too, is largely of a theoretical kind. Theory dominates the experimental work from its initial planning up to the finishing touches in the laboratory.[14]

Such an orientation, if tenable anywhere, presupposes a very high level of theory. Popper's attitude reflects the overweening interest in physics displayed by many philosophers of science. In contrast, it could be said that the greatest weakness in the less mature sciences is an unfortunate reliance on unsubstantiated and untestable theory. In this respect Medawar's admiration for Popper's method is just a little odd since, referring to his own field in 1969, Medawar said:

Biologists work very close to the frontier between bewilderment and understanding. Biology is complex, messy and richly various, like real life; it travels faster nowadays than physics or chemistry (which is just as well, since it has so much farther to go), and it travels nearer to the ground.[15]

It is easily seen from the collection of comments above that the roles of theory and experiment are not the same in all fields of science; what is more, these roles evolve as a particular science matures. Hacking says, more reasonably than Popper: "The relationships between theory and experiment differ at different stages of development [of a science]." Hacking reports on an important generational change that figures in the opposing methodologies of Davy and Liebig that we noted in opening our direct discussion of theory and experiment in Chapter 5 (see note 9, p.181):

When Davy wrote [1812], the atomic theory of Dalton and others had only just been stated, and the use of hypothetical models of chemical structures was only just beginning. By the time of Liebig [1863], one could no longer practice chemistry by electrically decomposing compounds or identifying gases by seeing whether they support combustion. Only a mind fueled by a theoretical model could begin to solve mysteries of organic chemicals.[16]

Versions of this methodological debate continued into the modern quantum chemistry era, as we saw in the exchanges between Lowdin, Heitler and Slater in Chapter 5.

Medawar's estimate of biology in 1969 was reinforced ten years later by Lewis Thomas, whose historical perspective suggests that progress, as assessed by the participants, is not likely to be smooth or uniform; progress is often made visible by the revisiting of what had earlier been considered settled portions of scientific problems:

> In biology, it is one stupefaction after another. Just thirty years ago we called it a biological revolution when the fantastic geometry of the DNA molecule was exposed to public view and the linear language of genetics was decoded. For a while things seemed simple and clear; the cell was a neat little machine, a mechanical device ready for taking to pieces and reassembling, like a tiny watch. But in just the last few years it has become almost imponderably complex, filled with strange parts whose functions are beyond today's imagining... It is not just that there is more to do, there is everything to do.[17]

Due to these sorts of hitches and starts, in which confusion and confidence alternate, the relative importance of experiment and theory might also be expected to fluctuate in response.[18] Thomas also calls to mind Bacon's description of the world as a "labyrinth, presenting as it does on every side so many ambiguities of way, such deceitful resemblances of objects and signs, natures so irregular in their lines, so knotted and tangled". On a larger historical scale, mature sciences do show, however, a general progression from simple observation to contrived and decisive experimentation to a more and more sound reliance on theory. Thomas Kuhn reviews the history of work on electricity, and notes the same kind of generational divides that Hacking found in the history of chemistry. Moreover, he reinforces the idea that the nature and pace of progress can change:

> Writers on electricity during the first four decades of the 18th century possessed far more information about electrical phenomena than had their 16th century predecessors. During the half-century after 1740, few new sorts of electrical phenomena were added to their lists. Nevertheless, in important respects, the electrical writings of Cavendish, Coulomb, and Volta in the last third of the 18th century seem farther removed from those of Gray, Du Fay, and even Franklin than are the writings of these

> early 18th century electrical discoverers from those of the 16th century. Somewhere between 1740 and 1780, electricians were for the first time enabled to take the foundations of their field for granted. From that point they pushed on to more concrete and recondite problems...[19]

These historical remarks can be used to highlight the situation within hydrology. We begin by recapitulating our argument to this point.

7.2 Experiment, Theory and Models in Hydrology

Applied hydrologic modelers have certainly "pushed on to more concrete and recondite problems" - perhaps inappropriately so, if they cannot yet "take the foundations of their field for granted". Their efforts utilize conceptual models within a hydrogeologic theory of a particular location and application. Recognizing the critical importance of strongly predictive models for water resource management purposes, early in our report we reviewed the associated details and difficulties of these models (Chapters 2, 3). That review soon bore out the accuracy of Mazur's remark that: "We may define 'experts' as two or more people who can authoritatively disagree with one another".[20] The situation recalls Freeman Dyson's account of the early days in nuclear physics, *Disturbing the Universe*, in which he says:

> The physicist Leo Szilard once announced to his friend Hans Bethe that he was thinking of keeping a diary: "I don't intend to publish it; I am merely going to record the facts for the information of God". "Don't you think God knows the facts?" Bethe asked. "Yes," said Szilard, "He knows the facts, but he doesn't know *this version of the facts*.[21]

Identifying a defensible set of significant facts became the central issue in our discussion of directly applied hydrologic models. Both data and theory typically underdetermine the conceptual and causal models of hydrologists. The sometimes divergent implications of non-unique but plausible models were seen to confound a more than qualitative effort at validation by either the modelers themselves or by responsible regulatory agencies. Recourse to professional judgment seemed the only available option in assigning weighted meaning to possibly abundant but inconclusive field data. For example, Glenn Hearne observed that although there was plenty of evidence to be sure aquifer characteristics varied spatially in the Tesuque aquifer, "the data are not adequate to describe this variation as a general pattern". Similarly, Mary Anderson characterizes the inadequate understanding of dispersion within both porous and fractured media as holes in the Emperor's underwear; the Emperor (theory) is just barely clothed at all. The significance of various details, both in the field and in models, is often poorly understood and difficult to incorporate into, or reconcile with, a conceptual model. As a result, hydrologists may yet be an exception to Medawar's confident claim that

> the ballast of factual information, so far from from being just about to sink us, is daily growing less... In all sciences we are being progressively relieved of the burden of singular instances, the tyranny of the particular. We need no longer record the fall of every apple.[22]

The tale of two USGS models at Pojoaque, New Mexico illustrated that conflicts over identifying the "few significant facts" are not only the usual result of adversarial litigation; nor does trouble arise only in complicated contaminant transport problems. Rather, as typically constructed, a groundwater model is usually inconclusive: it appeared that models might often be dark grey even when above reproach as a good faith effort. Obvious practical problems plague attempts at description and prediction as hydrologists struggle to identify the important apples within their interdisciplinary orchard. It remained to be seen if these problems were only practical in nature; if so, they might be amenable to resolution.

The standard validation practice equates reasonable history-matching with predictive capability. Our examination of logical issues in the validation of theories in hydrology failed to find any consistent and defensible basis for the strict validation of a model as a "correct representation of a process or system for which it is intended" that could be tested to "ensure an acceptable level of predictive accuracy". There was as yet no answer to the consequences of non-uniqueness; more than one model might well pass the usual tests but give seriously different projections. The widespread (and often completely reasonable) practice of immunizing models to falsification further compromised the situation. Poor history-matching is typically blamed on or solved by auxiliary hypotheses; such model changes cannot be equated with model "improvements", since they are often *ad hoc* and do not necessarily reflect any improved understanding of the system. Even negative post-audit results may provoke only a re-calibration of the model. This practice essentially reduces the post-audit to a secondary history-matching exercise with updated data.

On the positive side, groundwater models were recognized to be numerically and causally verifiable, in that known conditions and forces produce a limited and repeatable range of model responses. The mean of these responses becomes invariant over repeated model runs, and indicates the model's most probable response to the given conditions. Models are not open systems, even though their application to real systems and their regulatory validation are fraught with practical difficulties. All of which left open at least a theoretical possibility that strongly predictive causal models of real groundwater systems could be constructed.

On the other hand, when Anderson and Woessner defined model validation as a confirming post-audit, they did not leave open much hope that applied hydrologists could ever meet typical regulatory, legal and planning needs for *a priori* validation. They and other observers have thus deflated the idea of strict validation for applied models, but they have not usually appended an account of the prospects for progress. This is an important omission, since the idea that decisions based on models require valid models is only common sense. While validation expectations may be naïve and unrealistic, the general idea of validation is inescapable. Decisionmakers of all sorts need to know how good their tools actually are. Anderson and Woessner and some other prominent modelers counter that the only relevant tool is the expert opinion of an investigating hydrologist - not some computer-generated output. Like others, Anderson and Woessner point to the modeler's expert opinion as the proper basis of decisionmaking rather than the numerical model.[23] They go further and stress the use of models as *research tools* that can clarify modelers' thinking about systems. The numerical model is thereby relegated to a background role of uncertain importance.

A dilemma had thus emerged in our discussion, in which the goal of validation could not be set aside, but the means had been reduced toward a qualitative reliance on expert opinion. Allowing that applied hydrology may not yet be a reliably predictive science for at least some purposes, we wanted to take Einstein's advice and look at what hydrologists actually do in constructing their opinions, with less emphasis on either their non-scientific motivations or their rhetorical devices. Further exploration of validation required some insight into how hydrologists construct compelling accounts - at any level - in the absence of a protocol for strict validation. These qualitative validation procedures were most easily observed in connection with process-oriented research one step removed from the extra-scientific pressures that cloud the direct application of groundwater models. At the process level, progress is evident, promising us a window on how hydrologists gain confidence in their growing understanding.

Process-oriented research is ultimately aimed at supporting applied models, by eventually improving critical portions of model/prototype analogies. Bacon refers to this more disinterested level of research as *Experiments of Light*; he likes to point out that even God resorted to a day of this sort of thing before He took up the planned engineering projects. Questions arise in three main areas. First, given the reduction of validation to expert opinion, can the ways in which process-oriented investigators gain confidence in their growing understanding lead to a more productive applied methodology, or at least shed new light on the nature of the problem? That is, might the insights provided applied science by process investigations be procedural, as well as practical? Secondly, having surveyed the shortcomings in the solution of concrete hydrologic problems by means of applied theory, what is the proper relation between theory and experiment at this stage of hydrological development? How do the observations by Hacking, Kuhn, Medawar and Thomas on methodological evolution reflect on an immature empirical science like hydrology? Just what is the relationship of theory and experiment within process research? Lastly, Popper and Medawar are sure that discovery and testing are very distinct activities, with completely separate methodologies. Maybe this separation is unnecessary and even ill-advised; Putnam thought so. Is there anything in an act of discovery that might be linked to the possibility of testing, or at least the building of confidence? To pursue all of these questions, we looked at different aspects of *application-deferred* process-oriented hydrologic research, beginning with a history and analysis of Darcy's law, and concluding with original work on solute mixing at fracture junctions.

In our account of process-oriented experimentation and corroboration, the relation between theory and experiment is mediated by models. This is not a simple relationship; models play two inverted but complementary roles. In instances of discovery, the model is an early *goal* of experimental investigations that leads to phenomenological relationships, as in Darcy's 1855-1856 work at Dijon. Contra Popper, Darcy's law is an example of an empirical generalization from data having no deductive reliance on fundamental laws. In such cases, pattern recognition operates on accumulating data to generate reasonable mathematical descriptions of structures generally responsible for behavior of the observed type: the model is the culmination of phenomenological description, a summary of experience. Models play a second and different role in theoretical derivations of such relationships. Such derivations may precede or follow the experimental work, but in either case the model is now the *means* by which specific results are more or less shown to be special cases of a more general point of view. This more general view may or may not have been previously substantiated and accepted.

Whatever the sequence of events, theoreticians favor causal models in the form of structural explanations. Ernan McMullin emphasizes in this concept the direct causal link between model structure (geometry and processes) and explained behavior. A regular model geometry is a virtue, inasmuch as an "appropriately prepared model of the domain" is a structure whose mathematical form allows the investigator to apply fundamental laws to the structure stipulated; all of this is readily seen in connection with the various models used in derivations of Darcy's law, or in the different numerical models of solute mixing. The geometry stipulated need not be simple nor uniform, but it must be readily describable. The structure (model) coupled to fundamental laws together constitutes a theory or structural explanation of the phenomena investigated, the implications of which can be pursued through the resulting governing equations.

Multiple mutually inconsistent models are generally available to process-oriented investigators (and still impede convergent realism), but as Jacob Bear demonstrates with respect to Darcy's law, researchers on this level find some promise and even utility in this situation. Lack of strict validation is not fatal on this level of research as these structural explanations are mainly 1) a guide to experiment; or 2) a theoretical imprimatur that the suggested empirical correlation contains nothing in violation of more fundamental laws. We will refer to these two roles as indicative and corroborative, respectively.

In the first case, theoretical derivations precede experimental evidence and causal reasoning can be <u>indicative</u> of the nature of critical experimentation, just as Popper says; possibly testable hypotheses are the result. That is, causal arguments can indicate hypothetically likely behavior. The experimentalist then attempts to observe or deny the hypothesized behavior, results that may respectively confirm or falsify the original theory. This was the sequence of events in our investigation of solute mixing at fracture junctions, in which a numerical indication of complete mixing at low Peclet numbers preceded experimental confirmation. Experiments may also modify or leave indeterminate a proposed theory. In no event, however, can idealized causal reasoning directly prove anything about the non-ideal earth. And thus in every case, as Bear so pointedly insists, the "proof of the validity of a model, and the only way to determine coefficients, is always the experiment". "Validity", Bear says, cannot be demonstrated via logical deductions alone; experiments inevitably (and usefully) reintroduce a measure of the "qualitative richness of Nature" that mathematical treatments can eliminate. Experimentation routinely serves to reinforce or winnow theoretical conjectures. *Hydrologists require causal reasoning borne out or constrained by observation; or, alternatively, a form of critical induction buttressed by structural explanations.*

Our second case proceeds from the latter direction, as structural explanation is triggered by an empirical proposition. Theorists are interested in why certain phenomena occur and in any limitations on stated relationships. As in the treatment of Darcy's law - or in Philip's exploration of discontinuous fractures - the process ends in <u>theoretical corroborations</u> that are best described as semi-qualitative. Derivations attempt to display clearly what can or cannot be true if the reported empirical findings are true; these claims can then be compared with more generally accepted theory. It is also possible for sound theory to cast doubt on the soundness of reported experimental results. Derivations of empirical relationships from fundamental laws thus contribute to a satisfying though rather qualitative overall theoretical unity within the physical sciences. The gathering and generalizing of phenomena under relatively few general laws is not exact due to the idealized nature of the models employed and because broad approximations are frequently embedded in the derivations. Non-unique models also facilitate the theoretical exploration of alternative structural explanations for the same investigated phenomenon. Within these procedures, models replace the complexities of real systems with bundles of tubes, buckets of spheres, grid blocks, etc.; the nature and behavior of fluids, solutes and mixtures may also be severely simplified. For all these reasons, the derivations that follow experimental findings do not constitute more than a qualitative validation of the investigated relationship, with the imprecision being proportional to the liberties taken in simplifying or approximating the original system.

In Chapter 5, we mentioned several structural explanations of Darcy's law, and illustrated one in detail. *If there were no other basis for Darcy's law, it would not be the cornerstone of hydrology that it is.* But in fact Darcy's law rests primarily on Baconian experimental data and experience, and only secondarily on idealized theory and reasoning. Again, the results of experimentalists inevitably reflect some of the "qualitative richness of Nature" that mathematical treatments necessarily overlook; within limits, this adds weight and robustness to the demonstrated empirical relations. The experimental evidence for Darcy's law, combined with structural explanations of various kinds, give some indication that *the law is both empirically sound and causally plausible* within some limitations. The two primary functions of models (as goal and means) are thus again mutually supportive in the pursuit of compelling accounts of phenomena - in this case as theory is applied after experimental results are in hand.

The deductive reasoning in this latter case serves mainly a reassuring purpose after the fact. The degree of assurance naturally rises as the models employed in the structural explanation become more realistic; *i.e.*, as the geometry and processes of the model more closely approximate those of the real system. McMullin takes this fact to support convergence to an "approximately true" structural explanation, in which a faithful structural description merges into an explanation of why events occur as they do. McMullin expects a re-iterative theory/experiment cycling to lead to not only improved, but literally realistic models. This expectation had to be slightly modified in hydrology, however, where the systems are diverse and hydrologists do not benefit from repeated and detailed examination of nearly identical systems: prohibitively complicated systems rule out a final and optimal model.

The process-oriented research examined in our case studies is typical, in that it is characterized by narrowly defined and carefully exclusive problems; the exclusivity is effected by either *crucial assumptions* in theoretical work, or by *controlled experiments* in the laboratory. Thus research on the inner workings of a numerical model might be properly placed in this category, even though the application of that model to real systems is clearly something else. In addition, chosen internal problems are often of no immediate strategic importance, and for obvious reasons are typically ones the researcher has some idea he can solve.[24] As we have seen, hydrologists engaged in internal process-oriented research gain confidence (though not certainty) by a creative use of both theoretical derivations and experimental investigations. Theories without experimental corroboration, and empirical correlations without a link to theory are both generally considered incomplete (though not useless, due to the cooperative nature of such work). Since both theory and experiment are required, designation of either as the corroborating exercise simply depends on the order of events. Finally, the labor of process research may be divided among many workers with a preference for one aspect or another. It is therefore rare for anything of substance to suddenly appear as a coherent whole. The usual process is normally both incremental and iterative, with the many steps subject to peer review as results are communicated to colleagues as personal notes and journal articles.

The discussion of Darcy's law showed how process hydrologists employ multiple perhaps contradictory structural explanations that generally did not converge to a single optimum model. The situation clouded a literal and realist subscription to the structure stipulated in any single model. But since all models so used were seen as potential contributors in various ways to an improved understanding of the physical system, this was not treated as a major difficulty within the context of process-oriented models; in fact, the discussion may have seemed to suggest that final choices between models are unnecessary and even ill-advised in light of the different insights obtained from different structural explanations.

The general non-convergence of models remains a fact within the arena of directly applied models. The motivations, methods and issues internal to hydrology do not, however, carry over unadapted as hydrologists tackle external problems. Both the "cumulative uncertainty" and the consequences of error increase for applied hydrologists: they build models of inherently inclusive situations with practical considerations at stake. They model sites chosen on the basis of external criteria. They make their best guess at the influence of what are often very poorly understood structures and processes which they certainly did not select for study because they thought they were the next logical increment in the slow progress of hydrologic science. Applied modelers also often work in relative isolation within a short timeframe; they thus have correspondingly less benefit of peer review or iterative improvement for many aspects of their projects. In short, they routinely pursue what in another venue would be described as very poorly defined research projects.

Looking at things in this way, we again wonder if the methodology of process discovery is relevant to the logical problems in directly applied modeling. To what extent are the conventions of directly applied modeling analogous to the procedures of process-oriented researchers? In both settings, the hunt is for the empirically sound and causally plausible. It is tempting to view the usual applied causal models as an attempt at structural explanation, inasmuch as they argue from basic geometric and mechanistic conditions and relations to expected behavior. (It is not important for our purposes that applied models usually predict specific data, rather than more general phenomenological laws). As it happens, on occasion they have been described as something at least similar: theoretical corroborations of expert opinions. This was the case when George Pinder claimed his computer models were only illustrations of his opinion, intended to show the plausibility of that opinion in the context of general hydrologic theory.[25]

Applied modelers often arrive at intuitive professional opinions that the models they build can qualitatively "explain". Their models can help identify conservative and reasonable sets of assumptions and initial conditions capable of leading to observed conditions. Numerically validated causal models can reveal the implications of various structural choices. Standard history-matching "validations" rely heavily on this fact; the calibrated model, as a theory of a particular site, is never willingly trusted without reasonable history-matching.[26] As we have seen, this is often a manifestly inadequate corroboration of applied theory, since it is not difficult to construct many contradictory models that can all be weakly corroborated by a reasonable history match. It is clear, moreover, that the typical applied model is not the culmination of anything, and certainly not the culmination of phenomenological description. It rarely reflects an intimate understanding of the structure and processes of the region of interest. Typically, the only empirical findings at test in the verification/validation of models are historical records of head levels in wells, fluxes in surface bodies, or contaminant concentrations in monitoring wells.

The fatal flaw in applied modeling, therefore, is that deductive reasoning has been elevated far beyond either the <u>corroborative</u> or <u>indicative</u> roles assigned to it in process research. The fatal difference is that like a bird with one wing, causal models can *only* clarify the interaction of the *given* geometries, processes and other conditions. While the internal manipulation of these factors can be numerically validated, such models lack the second wing of a sound process-oriented investigation: phenomenological demonstration. When predictive modelers use reasoning as their primary tool, the indicated experiments or measurements are *in the future* and hence cannot immediately support the theoretically derived expectations. The most relevant experimental results are certainly not available to water managers faced with a decision. When modeling from the present to the past in so-called inverse problems, the indicated experiments require data that is the central issue under investigation and, possibly, in dispute; the empirical results of interest are in the unrecoverable past.

The very purpose of applied modeling is normally taken to be to tell us something we have no other reason to know. This is a major change from the more convincing structural explanations we reviewed in our discussions of process modeling of Darcy's law or solute mixing. "Proper" structural explanations either provide theoretical support for experimental evidence of phenomena, or they point to experiments that can be done to support or test the emerging theory. It is our contention that this interplay of experiment and theory, as mediated by models, must be a major part of any account of progress within process-oriented research. The overall result is that, compared to their process brethren, applied modelers rely on a weaker logical method but labor under weightier practical expectations with correspondingly sterner validation standards. One-sided modeling methodology is naturally inconclusive, and not only because of practical limitations. The situation recalls Anderson and Woessner's definition of model validation as a confirming post-audit. Experiments of fruit that rely on calibration and indirect verification are largely fruitless on complicated questions of law and policy in the absence of confirming experience.

7.3 In Pursuit of Structural Explanations

Our intent was to see whether the process investigations that provide detailed support for applied models might also suggest improvements in the *methodology* of applied models; further consideration has so far mainly highlighted the weaknesses in the usual predictive practices. Reliance on indirect history-matching tests of conceptual models leads to both practical problems (it often doesn't work) and logical ones (there is no reason to think it should work). There is good reason, therefore to stress the need for powerfully structural explanations over more indirectly verified, generically causal models.

This is clearly what Vit Klemes had in mind when he made a strong case for what he called causal models within the context of the surface water modeling debate. He noted in particular the potential to "derive the behavior of a hydrologic process for a given set of states of nature... from the dynamic mechanisms... without recourse to model calibration by empirical fitting".[27] Klemes was later forced, however, to recognize an unsatisfactory state of affairs, in which a pale process light had so far led only to boxes that were at best dark grey:

> The attendent danger of this difficulty [shortfalls in basic research] is the temptation of reductionism: to make shortcuts and to fill the void between the data and the goals with logically plausible assumptions that are sometimes correct but often wrong and, more often than not, individually untestable.[28]

In our discussion of applied models in Chapters 2, 3 and 4, we found that the bundles of "individually untestable" hypotheses within the conceptual model can confound overall model validation. Despite the claims of certain observers, the problem was not due to the "openness" of the uncertain model. On the contrary, the model is part of an numerically testable and verifiable theory of a particular problem. We characterized the overall effort as a perfectly controlled experiment that is not open at all. The only problem is that the validated model typically is not a precise representation of the physical system of interest, due to the idealizations, approximations and mistakes included in the model. As a result, *the incontrovertible model answer applies only to a different and usually far simpler question than the one actually posed.* Making these causal models more relevant (analogous) to practical concerns obviously requires extended insights into causes.

If it is important that hydrology become a predictive science in any but the most simple circumstances, the strength of the model/prototype analogy must be bolstered directly. Reasonable history-matching is a necessary but not sufficient condition for confidence. The only way out of the validation box is to shift the focus to directly closing the gap between model structure and that of the prototype. Here we find Tsang's recognition that "a thorough understanding represents actually the major part of validation" translated into both an organizing principle and a methodological rationale: improved insight and understanding is largely the province of experimental investigators in collaboration with "process-oriented applied modelers" of the sort we met in Chapter 4 (pp.135-137). Hence - contrary to the current calls for the major part of funding to go to "strategic research" geared to immediate applications - the rationale within an immature science is strong for experimental research into both improved geophysical methods and better understanding of physical, chemical and biological subsurface processes. Any progress in these areas can reduce the insupportable burden of proof placed on the usual "verification/validation" of causal models.

Process-oriented researchers take as their province the investigation of *individually testable* aspects of groundwater movement, behavior or models of the same. No one imagines this to be a brief project.[29] Causal factors are principally in the form of geologic structure and properties, processes of all sorts, and boundary conditions. Better prototype insight could take advantage of the numerically validated model response, as the model becomes a better facsimile with respect to system essentials. Although McMullin's realist notion is an unlikely reach for hydrologists, comparatively better structural explanations based on better models are certainly possible, given appropriate process-oriented experimental, theoretical and numerical work. Within an immature science, the balance of theory and experiment therefore naturally inclines toward the latter.

We saw above how Popper relegates experiments to a supporting (even though important) role in which experimenters are said to always be guided by theoretical overseers; his view of experiment is not surprising given the prominence of theory in his logic of science and his tendency to view a mature physics as the standard-bearer for all science. Jacob Bear describes the use of experiment in an ancillary way in connection with structural explanations. After noting the "understanding of the investigated phenomena and the role of the various factors that affect it" to be gained from the conceptual model approach, Bear also says: "All this information is needed for planning the laboratory experiments". As practitioners of an immature empirical science, process-oriented hydrologists do not, of course, only employ experiment in this way.

Experimental evidence often precedes theoretical treatments. Experimentation is thus not merely the adjunct to theory that Popper would have it. In this respect, the strategy of hydrologists is reminiscent of Thomas' comments about biology. He states bluntly that biologists have yet to "reach the stage of genuine theory", and are therefore "principally engaged in making observations and collecting facts, trying wherever possible to relate one set of facts to another", a thoroughly Baconian approach. Within hydrology, experiment is at least the equal of theory; theory cannot take routine precedence because hydrologists cannot as yet "take the foundations of their field for granted". As we noted before, whether theory or experiment is the corroborating activity simply depends on the order of events.

Much has been said about the difficulty of arranging and accepting truly crucial tests, which might seem to the Popperian to reduce the role of experiment. Hacking points out, however, that we shouldn't let the slop in so-called crucial experiments make us:

> play down the role of experiment too much. Certain types of experimental findings serve as benchmarks, permanent facts about phenomena which any future theory must accomodate, and which, in conjunction with comparable theoretical benchmarks, pretty permanently force us in one direction.[30]

This is no unqualified endorsement of *crucial demonstrations*, but it serves to announce their central importance in the immature sciences. Whether forcing water through sand-filled columns or idealized acrylic fracture networks, process researchers proceed by laying a foundation of pertinent observations, leading us "pretty permanently" in one direction or the other in the development of sound theory. Models remain non-unique, but they are thereby *constrained* in important ways by theoretical, experimental and measurement benchmarks, in which models figure as described in Chapters 5 and 6. Each such constraint acts as a further requirement within a process model, and, most importantly, reduces the field of significantly different non-unique models. In their own way, hydrologists thereby pursue what Francis Bacon calls the "collateral security" of a sound bi-modal scientific method. Building models on firm principles rather than fitting models to observations follows Bacon's maxim that "we must also examine and try whether the axiom so established be framed to the measure of those particulars only from which it is derived, or whether it be larger and wider".[31]

A recent Chapman Conference (AGU) was devoted to "Hydrogeological Processes: Building and Testing Atomistic-to-Basin-Scale Models". In his *EOS* report on this meeting, Tom Torgersen describes the fault line between attendees who "wanted to extract and explore the broad generalities", and those who "preferred to examine the specifics of hydrogeologic processes". He says the groups eventually found common ground, and comments shifted to:

> "I didn't know you could include that level of detail in models..." and "I didn't know you could supply that information for my model..." The shift came with the realization that field and laboratory studies supply valuable constraints on the initial conditions, boundary conditions, and parameterizations needed by hydrogeological process modelers and that models supply the observationalist with valuable insight into the implications and requirements of the system as it is envisioned. Much of what the contrasting schools do and conclude is valid, but one school is not necessarily asking the same questions as the other school.[32]

Torgersen highlights both 1) the constraining role of experimental evidence on theory development; and 2) the guidance for experiment provided by theory. As a result of such collaborations, a sense of optimism occasionally surfaces that runs counter to the gloom we mentioned at the outset. In particular, Torgersen reports from the same conference that a cautious optimism prevails among those those committed to building bridges between process-oriented modelers and experimentalists:

> Eventually, an overall feeling began to emerge that the issue is not intractable. The wealth of new techniques and disciplines now available and focused on specific components of these processes and parameters in space and time has reached a point where these new bridges can and are being built. The leaders in the field are now asking, "How can I help?", "What can I do?", and "Who can I ask...?" This exchange of information allows the Earth sciences to focus on the right questions. And as everyone knows, once you ask the right question, any graduate student can turn it into a good thesis.[33]

The "bridges" Torgersen applauds push process models toward the sort of experimental confirmation and information that we claim is essential in the building of compelling accounts in normal science. Similarly, when Torgersen says "models supply the observationalist with valuable insight into the implications and requirements of the system as it is envisioned", he echoes both the methodology of structural explanation and our view of the proper symbiosis between theory and experiment. He finds grounds therein for some optimism regarding significant progress in the construction of process models. The physicist Max Planck said:

> *Man muss optimist sein.* We must be optimists. ...Science demands also the believing spirit. Anybody who has been seriously engaged in scientific work of any kind realizes that over the entrance to the gates of the temple of science are written the words: *Ye must have Faith.* It is a quality which the scientists cannot dispense with... Imaginative vision and faith in the ultimate success are indispensable. The pure rationalist has no place here.[34]

There is plenty of evidence of a growing impatience among those watching from the margins of hydrologic practice. The likely erosion of water managers' operational realism - blind faith in simple and defensible answers - calls for a constructive response. Hydrologists could do much worse than to adopt the attitude attributed to Charles Slichter by his biographer; they might all become "reverent iconoclasts" - fully aware of the beauty and promise of science, but bringing to bear on their own results a useful and skeptical rigor. It is really up to hydrologists to keep the faith, and recognize their professional opinions as no more - and at the same time no less - than what they are.

7.4 Conclusion

In Vonnegut's *Bluebeard*, a painter is asked how he can tell a good painting from a bad one, and he answers that there isn't much to it: "All you have to do, my dear, is look at a million paintings, and then you can never be mistaken".[35] This is not an option available to hydrologists, who share the methodological difficulties of other historical sciences investigating complicated and remarkably variable systems. The major flaw in the standard applied model has been identified as a dependence on only weakly confirmed hypotheses; what should be corresponding experimentation or measurements are typically in the future, and cannot help a decisionmaker evaluate the soundness of that model. At best, the conscientious applied modeler can claim: "A more precise adjustment of the model is not justified by available data" (Hearne), or, somewhat more positively, "The representation of the physical hydrogeological system in the model is substantiated by available data" (McAda/Wasiolek). Recognition of shortfalls in strict validation and strict falsification mean that science becomes at once harder for insiders and less accessible to outsiders. Good judgment is not so readily attained by practitioners nor recognized by others, much to the chagrin of observers, funders, and would-be debunkers or policers of science. In short, the groundwater modeler is, as Nietzsche says, much more of an artist than he knows.

There is thus reason to think that both the degree and the practical consequences of uncertainty loom larger in hydrology than in many other sciences. If applied groundwater scientists cannot readily defend what they claim to know, they can lend only an uncertain weight to political, legal and economic deliberations.[36] The idea that model validation consists of a confirming post-audit highlights the immature status of hydrology and the near-term futility of predictive causal models in many settings. Current practices nevertheless routinely solicit opinions that focus on inappropriate, inadequate and ultimately indefensible tools of the groundwater trade. Not all of these predictive exercises can be delayed; not all of the pressing and specific questions of water managers can be deferred. If hydrology is not a reliable predictive science for some purposes, then there are obvious implications for water managers; rather than a stubborn insistence on unknowable answers, the questions asked should be re-evaluated. There ought to be policy repercussions in the form of 1) routine careful documentation of modeling procedure; and 2) a commitment to standard post-audit reports. In particular, given the uncertainty of the applied science, managers should directly consider the consequences of model failure. If the consequences of model failure are unacceptable, the result should be a short-term caution in water resources management and a long-term renewed commitment to the process research confidence requires. If it is important to expand the class of problems for which hydrologists can prognosticate, then more direct insight into physical systems is needed, providing some constraint on a bewildering array of non-unique model/prototype analogies, while avoiding the pitfalls of terminal research.

Hydrologists, though practitioners of an immature science, cannot always wait until they are completely satisfied with their tools, even if their primary tool is their experience and understanding of the subsurface. This limited expertise serves as the major instrument in the "validation" of models. With respect to process research, we have tried to unravel how compelling (if still unvalidated) accounts result from a creative and fascinating symbiosis of theory and experiment. The iterative and cooperative efforts of process modelers circumvent *to a degree* the logical difficulties of applied modelers. The more "proper" structural explanations on this level assign models a role mediating between theory and experiment, with model improvement proceeding from advances at either end.[37] Within this procedure, discovery and validation cannot be so simply demarcated as Medawar and Popper prefer. Since we must be content with confidence and not proof, the source and development of ideas is not irrelevant to a judgment on the confidence they instill. In addition to process research largely abstracted from practical problems, another intermediary between process and application is the development of what can be termed site-specific process models. Torgersen expresses optimism that meaningful progress is being made as hydrologists take advantage of peer-reviewed, iterative and cooperative combinations of theory, experiment and numerical simulations.

Similar and much more important progress can be expected as better process models are incorporated into directly applied models. Experimental and theoretical constraints can augment the standard history-matching guides, forcing model design in one direction or another. For example, detailed and reliable information on the actual mountain-front recharge to the Tesuque aquifer from the Sangre de Cristo Mountains would act as a significant constraint on model construction by fixing a major part of the water budget. Or, a more detailed understanding of solute behavior at fracture junctions could replace the computationally convenient but hydrodynamically naïve models used previously. Uncertain optimization can surely benefit from being subject to more constraining conditions and more relevant information. In urging the surface water modelers to adopt causal methods, Klemes points to the possibility that models might be so significantly constrained that their practical utility will overwhelm their logical underdetermination. This is essentially the situation in the theoretically mature and (correspondingly) predictive sciences; their progress is slower, but less important, now that good structural explanations are available for functional purposes.

Any analysis of the current technical limitations of groundwater science must attempt to distinguish between temporary and more resistant inadequacies. Our analysis leads inevitably to the idea that the current balance of theoretical and experimental needs is neither fixed nor amenable to adjustment by bureaucratic priorities. Environmental strategists must make a similar distinction between fixed goals and contingent, short-term decisions. The resulting insights could form the basis of a sound approach to the development and application of groundwater models. Kuhn asks: "Is it really any wonder that the price of significant scientific advance is a commitment that runs the risk of being wrong?"[38] If it is important that hydrology become a predictive science, then mistakes must also be allowed, and modeling cannot remain an "risk-free indoor sport". Kuhn echoes Bacon, who said, "Truth emerges more readily from error than from confusion". If the potential consequences of mistakes are unacceptable, then the questions asked must again be modified or abandoned, in effect limiting the options discussed.

Sharing Bacon's sentiment, we consider the major importance of hydrologic science to be its practical utility. The utility of applied models, real or imagined, is the public's only vested interest in the methodology of hydrologists. This methodology has not developed within a vacuum. Early on in motivating a detailed look at groundwater methodologies, we stressed the impact of an evolving regulatory situation. Managers of public works have apparently assumed hydrology is a strongly predictive science because it is important that it be such a science. This assumption makes it easy to misrepresent the proper balance of theory and experiment at this stage of scientific development. The regulatory climate within which many hydrologists have come of age has littered their path with Atalanta's apples in the form of pressing engineering concerns. In the process, an immature science has been warped away from its proper attention to fundamentals. The growing emphasis on strategic research reminds us that research funding does not come without strings attached. At the same time, *acknowledging the importance of reliable hydrologic predictions has actually provided the motivation for an intense investigation of fundamentals.* This is most emphatically not an excuse for Medawar's academic play (p.233), nor for Pramer's terminal science (p.220).[39] On the contrary, like Bacon, hydrologists should resist the "unseasonable and premature tarrying" over direct applications - simply because of the manifest "unfruitfulness of the way".

☼

The growing sophistication of her female wine customers recently encouraged Adrienne Hertler, co-owner of a bistro in San Fransisco, to hope "they are getting over that white zinfandel plateau".[40] Whatever the state of feminine enology, one can hope that hydrologists will surmount the primitive Popperian positions that unfortunately obscure much of the current discussion of methodology, validation and progress. Technical insight is not enough in the present regulatory and funding atmosphere. While the social, legal and scientific aspects of environmental protection and remediation are clearly individually complex, they are also highly interwoven. The validation debate carried on in leading scientific journals has yet to disentangle these spheres of influence. A more sophisticated approach to a general philosophy of hydrology is needed to clarify debates over such diverse issues as the justification of research programs and the validation of groundwater models with public policy implications. The semi-scientific margins of hydrologic practice need the defining and defending contributions of hydrologists themselves, for both short- and long-term purposes. A sudden awareness that "all their lives they have been speaking prose" is not sufficient today; a commitment is needed from hydrologists to speak to these points intelligently and, maybe, persuasively.

7.4 Notes:

1. Kuhn, T.S. (1962), *The Structure of Scientific Revolutions*, University of Chicago Press, 2nd edition (1970), p.17.

2. Bacon, F. (1620b), *Novum Organum*, in Burtt, E.A. (ed.) (1939), *The English Philosophers from Bacon to Mill*, The Modern Library, *ciii*, p.70.

3. Einstein, A. (1935), "On the Method of Theoretical Physics", in *The World as I See It*.

4. See, *eg.*, Freeze, R.A. and Cherry, J.A. (1989), "Guest Editorial - What Has Gone Wrong?", in *Ground Water*, **27**:4, pp.458-464: "The practicality of much of the current ground-water-quality legislation can also be questioned. Most of this legislation was promulgated in the early heady years of the environmental movement. In many cases it was designated with little hydrogeological input, and without recognition of the true complexities of the problems... In summary, future legislation should be more cognizant of available technology, more flexible in its technical requirements, and more applicable to nonpoint sources and DNAPL migration" (pp.459-460).

5. Dyson, F. (1979), *Disturbing the Universe*, Harper and Row, p.104-5.

6. Keillor, G. (1993), *The Book of Guys*, Penguin Books, p.11.

7. Hacking, I. (1983), *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*, Cambridge University Press, p.154.

8. The guidance and fleshing out of existing theory is the essence of what Thomas Kuhn (1962) would call normal science.

9. Bacon, F. (1620b), *civ*, p.71.

10. Bacon, F. (1620b), *xxv*, p.32. Bacon's essay on "Superstition" includes the famous line: "... the schoolmen were like astronomers, which did feign eccentrics and epicycles, and such engines of orbs, to save the phenomena, though they knew there were no such things" (Bacon, F. (1587), *Essays*, Everyman Classics (1986), p.52.)

11. Thomas, L. (1980), *Late Night Thoughts on Listening to Mahler's Ninth Symphony*, Viking Press, pp.70-71.

12. Medawar, P.B. (1968), *Induction and Intuition in Scientific Thought*, American Philosophical Society, p.38. Elsewhere (Medawar, P.B. (1967), *The Art of the Soluble*, Methuen, p.119), he says: "Baconian experimentation is not a critical activity, but a kind of creative play".

13. Popper himself (*The Logic of Scientific Discovery*, Routledge (1992), p.42) says: "I must admit the justice of this criticism, but...".

14. Popper, K.R. (1959), *The Logic of Scientific Discovery*, Routledge (1992), p.107.

15. Medawar, P.B. (1969), p.1.

16. Hacking, I. (1983), *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*, Cambridge University Press, p.154-5.

17. Thomas, L. (1980), The Viking Press, pp.19-20.

18. When a science that is accustomed to confidently relying on existing theory is forced to reconsider these fundamental guidelines, Thomas Kuhn (1962) would say it is "in crisis". One result might be that: "Previously completed work on normal projects would now have to be done again because earlier scientists had failed to recognize and control a relevant variable" (p.59). More broadly, "Some old problems may be relegated to another science or declared entirely 'unscientific'. Others that were previously non-existent or trivial may, with the new paradigm, become the very archetypes of significant scientific achievement" (p.103). From this it is clear that "the data are not unequivocally stable" (p.121), and "sensory experience is [not] fixed and neutral" (p.126), but is subject to reinterpretation in the wake of major theoretical upheavals.

19. Kuhn, T.S. (1962), p.21. It is no coincidence that these advances were fueled in large part by technological and experimental gains, as Kuhn reports: "The post-Franklinian developments include an immense increase in the sensitivity of charge detectors, the first reliable and generally diffused technique for measuring charge, the evolution in the concept of capacity and its relation to a newly refined notion of electric tension, and the quantification of electrostatic force" (p.21, note 12).

20. Mazur, A. (1981), *The Dynamics of Technical Controversy*, Communications Press, p.10.

21. Dyson, F. (1979), p.*xi*.

22. Medawar, P.B. (1991), *The Threat and the Glory: Reflections on Science and Scientists*, Oxford University Press, p.*xv*.

23. Although our discussion of applied modeling quickly moved to the usual numerical exercises, *model* in the sense we defined it certainly includes all analytical and less formally mathematical conceptual models.

24. See, *eg.*, Kuhn, T.S. (1962), p.164. Accordingly to Kuhn, this is part of the reason normal, autonomous science is so productive. "Normal research, which *is* cumulative, owes its success to the ability of scientists regularly to select problems that can be solved with conceptual and instrumental techniques close to those already in existence" (p.96).

25. In that particular trial, Pinder was employing an inverse argument, but not modeling back in time; rather, he assumed certain things and demonstrated that these assumptions could result in conditions similar to those observed in the present. Not much has been made in this study of the fact that applied modelers model both forward and back in time. Although the technical issues differ, the logic of model testing is essentially the same in arguing from a known condition forward or backward to an empirically unknown and presently unknowable condition.

26. Anderson, M.P. and Woessner, W.W. (1992a), *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*, Academic Press, pp.253-256: "Owing to uncertainties in the calibration, the set of parameter values used in the calibrated model may not accurately represent field values. Consequently, the calibrated parameters may not accurately represent the system under a different set of boundary conditions or hydrologic stresses... In a typical model verification exercise, values of parameters and hydrologic stresses determined during calibration are used to simulate a transient response for which a set of field data exists... In the absence of a transient data set, the calibration may be tested using a second independent set of steady-state data... Unfortunately, it is often impossible to verify a model because usually only one set of field data is available. That data set, of course, is needed for calibration. If this is the case, the model cannot be verified. A calibrated but unverified model can still be used to make predictions as long as careful sensitivity analyses of both the calibrated model and the predictive model are performed

and evaluated. Predictions resulting from calibrated but unverified models generally will be more uncertain than predictions derived from verified models".

27. Klemes, V. (1982), "Empirical and Causal Models in Hydrology", in National Academy Press (1982), *Studies in Geophysics: Scientific Basis of Water-Resource Management*, p.102.

28. Klemes, V. (1982), p.102.

29. Einstein said: "Now I know why there are so many people who enjoy chopping wood. In this activity one immediately sees the results". Weber, R.L. (ed.) (1973), *A Random Walk in Science*, The Institute of Physics, London and Bristol, p.7: "When a theoretical physicist is asked, let us say, to calculate the stability of an ordinary four-legged table he rapidly enough arrives at preliminary results which pertain to a one-legged table or a table with an infinite number of legs. He will spend the rest of his life unsuccessfully solving the ordinary problem of the table with an arbitrary, finite, number of legs".

30. Hacking, I. (1983), p.254.

31. See Chapter 1, p.9; Chapter 6, p.215.

32. Torgersen, T. (1994), "Chapman Conference on Hydrogeological Processes Promotes Interaction", in *EOS*, **75**:47, 22 Nov 1994, p.547.

33. Torgersen, T. (1994), pp.547, 554.

34. Max Planck (1933), *Where is Science Going?*, Humanities Press, pp.214-5; see also p.39.

35. Vonnegut, K. (1987), *Bluebeard*, Delacorte Press, p.156. This is the major faultline between the more "exact" sciences, such as physics and certain parts of chemistry, and the other empirical and social sciences; the former benefit from repetition.

36. See, *eg.*, Tarricone, P.: "The Politics of Nuclear-Waste Disposal", in *Civil Engineering*, March 1994, pp.56-8. The site chosen by the Illinois Department of Nuclear Safety (INDS) for a low level radioactive waste facility was later rejected by a special review commission. Technical shortcomings of the original decision included "suspect travel time estimates: The commission found that statistical modeling was of limited help in determining the travel time of contaminants through the [subsurface], because it was based on generalizations that sacrificed critical detail and hydrostratigraphic assumptions that were not sufficiently varied. This reduced the modeling's value to the commission".

37. McMullin indicates that in at least some cases, models improve and converge to an "approximately true" description of prototype structure; this is more likely the case in exact and mature sciences with less reliance on unique historical events. Within hydrology, applied groundwater models are more likely to be inconclusive than process models because 1) they apply sweeping conditions often only marginally supported by field evidence; and 2) because the results are rarely checked by the implied measurements (post-audits); 3) model calibration rests on correlations with datasets, and not on phenomenological insight; 4) delineating the domain within which the results are "reasonably good" is much more difficult; 5) the systems are diverse, and hydrologists do not benefit from the repeated and more detailed examination of nearly identical systems; and 6) the supporting experiments and measurements are projected into the future along with the prediction. As a result, the extension of modeling activities to the applied arena is beset by major problems.

38. Kuhn, T.S. (1962), p.101.

39. See also Pramer, D. (1985), "Terminal Science", in *Bioscience*, **35**, p.141: "Terminal science is an extreme form of insularity; it is science performed for its own sake or solely for the sake of scientists. Terminal science can be intellectually satisfying and profoundly important to the practitioners of a discipline and to the discipline itself. However, the vision of those who practice terminal science does not extend beyond those limits. Terminal science is not only antithetical to science for the people, it is also an indulgence that few today can afford".

40. Chapman, F. (1994), "Female Wine Lovers Tired of Less-Than-Rosé Treatment", Knight-Ridder Newspapers, printed in the *Albuquerque Journal*, 5 October 1994, p.B1.

8

Complete Bibliography

Abbott, M.B. (1992), "The Theory of the Hydrologic Model, or: The Struggle for the Soul of Hydrology", in O'Kane, P. (ed.) (1992), *Advances in Theoretical Hydrology: A Tribute to James Dooge*, Elsevier, pp.237-254.

Allchin, D. (1992), "How Do You Falsify a Question?: Crucial Tests Versus Crucial Demonstrations", in Hull, D., *et al.* (eds.) (1992), *PSA 92*, Proceedings of the Philosophy of Science Association, pp.74-88.

Anderholm, S.K. (1994), "Ground-Water Recharge Near Santa Fe, North-Central New Mexico", United States Geological Survey, Water-Resources Investigations Report 94-4078.

Anderson, M.P. (1983), "Ground-Water Modeling - The Emperor Has No Clothes", in *Ground Water*, **21**:6, pp.666-669.

Anderson, M.P. and Woessner, W.M. (1992a), *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*, Academic Press.

Anderson, M.P. and Woessner, W.M. (1992b), "The Role of the Postaudit in Model Validation", in *Advances in Water Resources*, **15**, pp.167-173.

Appleyard, B. (1993), *Understanding the Present: Science and the Soul of Modern Man*, Anchor Books, Doubleday.

Bacon, F. (1597), *Essays*, Everyman's Library (1986).

Bacon, F. (1620a), *The Great Instauration*, in Burtt, E.A. (ed.) (1939), *The English Philosophers From Bacon to Mill*, The Modern Library, pp.5-23.

Bacon, F. (1620b), *Novum Organum*, in Burtt, E.A. (ed.) (1939), *The English Philosophers From Bacon to Mill*, The Modern Library, pp.24-123.

Bacon, F. (1620b*), *Novum Organum*, in Urbach, P. and Gibson, J. (trans. and eds.), Open Court (1994).

Balasubramanian, K., Hayot, F. and Saam, W.F. (1987), "Darcy's Law From Lattice-Gas Hydrodynamics", in *Physical Review A*, **36**:5, pp.2248-2253.

Barber, B. (1961), "Resistance by Scientists to Scientific Discovery", in *Science*, **134**, 1 Sept 1961, pp.596-602.

Bardsley, E. (1993), "A Moribund Science?", in the *Journal of Hydrology (New Zealand)*, **31**:1, pp.1-4.

Bauer, H. (199x), "Darwin on Trial [review]", in *Journal of Scientific Exploration*, **6**:2, pp. 73-78, book review of Johnson, P. (1991), *Darwin on Trial*, Regnery Gateway.

Bear, J. (1972), *Dynamics of Fluids in Porous Media*. Elsevier / Dover Books (1988).

Bear, J., Beljin, M., and Ross, R. (1992), "Fundamentals of Ground-Water Modeling", in *EPA Ground Water Issue*, EPA/540/S-92/005.

Bear, J. and Veruijt, A. (1987), *Modeling Groundwater Flow and Pollution*. Dordrecht and Boston: D. Reidel Publishing Company.

Berkowitz, B., Naumann, C. and Smith, L. (1994), "Mass Transfer at Fracture Intersections: An Evaluation of Mixing Models", in *Water Resources Research*, **30**:6, pp.1765-1773.

Beven, K. (1993), "Prophecy, Reality and Uncertainty in Distributed Hydrological Modelling", in *Advances in Water Resources*, **16**:1, pp.41-51.

Blake, W.P. (1859), "Observations on the Mineral Resources of the Rocky Mountain Chain, Near Santa Fe, and the Probable Extent Southwards of the Rocky Mountain Gold Field", in *Boston Society of Natural History Proceedings*, **7**, pp.64-70

Bogen, J. and Woodward, J. (1988), "Saving the Phenomena", *The Philosophical Review*, **97**, pp.303-352.

Borton, R.L. (1968), "General Geology and Hydrology of North-Central Santa Fe County, New Mexico, New Mexico State Engineer Office open-file report.

Bredehoeft, J.D. and Konikow, L.F. (1993a), "Ground-Water Models: Validate or Invalidate?", in *Ground Water*, **31**:2, pp.178-9.

Bredehoeft, J.D. and Konikow, L.F. (1993b), "Reply to Comment", in *Advances in Water Resources*, **15**, pp.371-372.

Brush, S.G. (1989), "Prediction and Theory Evaluation: The Case of Light Bending", in *Science*, **246**, 1 Dec 1989, pp.1124-1129.

Burges, S. (1993), text of a UNNY Symposium address, University of Waterloo, Waterloo, Ontario, Canada, July 1994.

Burroughs, E.R. (1912), *A Princess of Mars*, Ballantine Books, 19th printing (1983).

Bury, J.B. (1932), *The Idea of Progress: An Inquiry Into Its Origin and Growth*, Dover Books (1955).

Carr, E.H. (1961), *What Is History?*, Vintage Books.

Cartwright, N. (1983), *How the Laws of Physics Lie*. New York: Oxford University Press.

Cartwright, N. (1989), "The Born-Einstein Debate: Where Application and Explanation Separate", in *Synthese*, **81**, pp.271-282.

Chapman, F. (1994), "Female Wine Lovers Tired of Less-Than-Rosé Treatment", Knight-Ridder Newspapers, printed in the *Albuquerque Journal*, 5 October 1994, p.B1.

Chamberlain, T.C. (1890), "The Method of Multiple Working Hypotheses", in *Science* (old series), **15**:92 (1890); reprinted in *Science*, **148**, 7 May 1965, pp.754-759.

Chamberlain, T.C. (1892), "The Method of Multiple Working Hypotheses", in *Journal of Geology*,**39** (1931), pp.155-165.

Cope, E.D. (1874), "Notes on the Santa Fe Marls and some of the contained Vertebrate Fossils", in *Philadelphia Academy of Natural Science Proceedings, Paleontology Bulletin*, **18**, pp.147-152.

Culhane, P.J. (ed.) (1987), *Forecasts and Environmental Decisionmaking: the Content and Predictive Accuracy of Environmental Impact Statements*, Westview Press.

Darcy, H. (1856), *Les Fontaines Publiques de la Ville de Dijon*. Paris: Victor Dalmont.

Davis, P.A., Olague, N.E., and Goodrich, M.T. (1992), "Application of a Validation Strategy to Darcy's Experiment", in *Advances in Water Resources*, **15**, pp.175-180.

de Marsily, G., Combes, P. and Goblet, P. (1992), "Comment on 'Groundwater Models cannot be Validated', by L.F. Konikow and J.D. Bredehoeft", in *Advances in Water Resources*, **15**:6, pp.367-369.

Dooge, J. (1983), "On the Study of Water", in *Hydrologic Science Journal*, **28**:1, pp.23-48.

Dooge, J. (1986), "Looking for Hydrologic Laws", in *Water Resources Research*, **22**:9, pp.46s-48s.

DuMars, C.T., O'Leary, M. and Utton, A.E. (1983), *Pueblo Indian Water Rights: Struggle for a Precious Resource*, University of Arizona Press.

Dunne, T. (1980), "Models of Runoff Processes and Their Significance", in National Academy Press (1982), *Studies in Geophysics: Scientific Basis of Water-Resource Management*, pp.17-29.

Dyson, F. (1979), *Disturbing the Universe*, Harper and Row.

Fanther, G. (1956), "Henry Darcy - Engineer and Benefactor of Mankind", in the *Journal of Petroleum Technology*, **8**:10, pp.12-14.

Fishel, V.C. (1935), "Further Tests of Permeability With Low Hydraulic Gradients", in *Transactions of the American Geophysical Union*, **16**, pp.409-503.

Flew, A. (1979), *A Dictionary of Philosophy*, St. Martin's Press, revised 2nd edition (1984).

Fourier, J. (1822), *The Analytical Theory of Heat*, Dover (1955).

Freeze, R.A. (1994), "Henry Darcy and the Fountains of Dijon", in *Ground Water*, **32**:1, pp.23-30.

Freeze, R. A. and Back, W. (eds.) (1983), *Physical Hydrogeology*. Stroudsburg: Hutchinson Ross.

Freeze, R. A. and Cherry, J. A. (1979), *Groundwater*. Englewood Cliffs: Prentice Hall.

Freeze, R.A. and Cherry, J.A. (1989), "Guest Editorial - What Has Gone Wrong?", in *Ground Water*, **27**:4, pp.458-464.

Gale, J.E. (1987), "Comparison of Coupled Fracture Deformation and Fluid Flow Models with Direct Measurement of Fracture Pore Structure and Stress Flow Properties", in *29th U.S.Symposium on Rock Mechanics*, Tucson AZ.

Galileo Galilei (1638), *Two New Sciences*, Drake, S. (trans., ed.), Wall & Thompson, Toronto, 2nd Edition (1989).

Galusha, T. and Blick, J.C. (1971), "Stratigraphy of the Santa Fe Group, New Mexico", in *American Museum of Natural History Bulletin*, **144**.

Glen, W. (1982), *The Road to Jaramillo*, Stanford University Press.

Glymour, C. (1980), *Theory and Evidence*, Princeton University Press.

Goode, D.J. and Konikow, L.F. (1990), "Reevaluation of Large-Scale Dispersivities for a Waste Chloride Plume: Effects of Transient Flow", in Kovar, K. (ed.), *ModelCARE 90: Calibration and Reliability in Groundwater Modeling*, The Hague, Netherlands, September 1990, IAHS Publication 195, pp.417-426.

Goodrich, D.R. and Woolhiser, D.A. (1994), "Comment on 'Physically Based Hydrologic Modeling, 1: A Terrain-Based Model for Investigative Purposes' by R.B. Grayson, I.D. Moore, and T.A. McMahon", in *Water Resources Research*, **30**:3, pp.845-847.

Goodwin, Irwin (1995), "Clinton's R&D Budget Defers Pain to Unkindest Cuts by Republicans", in *Physics*

Today, **48**:4, April 1995.

Gould, S.J. (1989), *Wonderful Life: The Burgess Shale and the Nature of History*, W.W.Norton and Co., Inc.

Grayson, R.B., Moore, I.D. and McMahon, T.A. (1994), "Physically Based Hydrologic Modeling, 1: A Terrain-Based Model for Investigative Purposes; and 2: Is the Concept Realistic?", in *Water Resources Research*, **26**:10, pp.2639-2666.

Grayson, R.B., Moore, I.D. and McMahon, T.A. (1994), "Reply", in *Water Resources Research*, **30**:3, pp.855-856.

Griggs, R.L. (1964), "Geology and Ground-Water Resources of the Los Alamos Area New Mexico", United States Geological Survey, Water-Supply Paper 1753.

Hacking, I. (ed.) (1981), *Scientific Revolutions*, Oxford University Press.

Hacking, I. (1983), *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*, Cambridge University Press.

Hamlin, C. (1990), *A Science of Impurity: Water Analysis in Nineteenth Century Britain*, University of California Press.

Hassanizadeh, S.M. and Carrera, J. (1992), "Editorial", in *Advances in Water Resources*, **15**:1, pp.1-3.

Hawking, S.W. (1988), *A Brief History of Time: From the Big Bang to Black Holes*, Bantam Books.

Hearne, G.A. (1985a), "Mathematical Model of the Tesuque Aquifer System Near Pojoaque, New Mexico", United States Geological Survey, Water-Supply Paper 2205.

Hearne, G.A. (1985b), "Simulation of an Aquifer Test on the Tesuque Pueblo Grant, New Mexico", United States Geological Survey Water-Supply Paper 2206.

Heitler, W. (1967), "Quantum Chemistry: The Early Period", in the *International Journal of Quantum Chemistry*, **1**:1, pp.13-36.

Hixson, J. (1976), *The Patchwork Mouse*, Anchor Press / Doubleday.

Hofmann, J. (1990), "How the Models of Chemistry Vie", in Fine, A., *et al.*, (eds.) (1990), *PSA 1990*, Vol.1, East Lansing: Philosophy of Science Association, pp.405-419.

Hofmann, J. and Hofmann, P. (1992), "Darcy's Law and Structural Explanation in Hydrology", in Hull, D., *et al.* (eds.) (1992), *PSA 92*, Proceedings of the Philosophy of Science Association, pp.23-35.

Hofmann, P. (1994), "The Case of the Immodest Modeler: Uncertainty and Realism in Groundwater

Modeling", in the Proceedings of the *New Mexico Conference on the Environment*, New Mexico Environment Department, Albuquerque, NM.

Hon, Giora (in press), "Is the Identification of Experimental Error Contextually Dependent? The Case of Kaufmann's Experiment and Its Varied Reception", in Buchwald, J. (ed.), *Scientific Practice: Theories and Stories of Doing Physics*, University of Chicago Press.

Hubbert, M. K. (1969), *The Theory of Groundwater Motion and Related Papers*. Hafner Publishing Company.

Ingraham, M.H. (1972), *Charles Sumner Slichter: The Golden Vector*, University of Wisconsin Press.

International Atomic Energy Agency (1988), *Radioactive Waste Managment Glossary*, 2nd edition, International Atomic Energy Agency Technical Document IAEA-TECDOC-447.

Keillor, G. (1993), *The Book of Guys*, Penguin Books.

Keller, E.F. (1983), *A Feeling for the Organism: The Life and Work of Barbara McClintock*, W.H. Freeman and Co., New York.

Kelly, V.C. (1978), "Geology of Espanola Basin, New Mexico", New Mexico Bureau of Mines and Mineral Resources, Geologic Map 48.

Kelly, V.C. (1952), Tectonics of the Rio Grande Depression of Central New Mexico, in *Guidebook to the Rio Grande Country*, New Mexico Geological Society, pp.93-105.

Kernodle, J.M., McAda, D.P. and Thorn, C.R. (1995), "Simulation of Ground-Water Flow in the Albuquerque Basin, Central New Mexico, 1901-1994, With Projections to 2020", United States Geological Survey, Water-Resources Investigations Report 94-4251.

Kitcher, P. and Salmon, W.C. (eds.) (1989), *Scientific Explanation*, in: Giere, R.N. (general editor), *Minnesota Studies in the Philosophy of Science*, Vol.13, University of Minnesota Press.

Klemes, V. (1982), "Empirical and Causal Models in Hydrology", in National Academy Press (1982), *Studies in Geophysics: Scientific Basis of Water-Resource Management*, pp.95-104.

Klemes, V. (1986), "Dilettantism in Hydrology: Transition or Destiny?", in *Water Resources Research*, **22**:9, pp.177s-188s.

Klemes, V. (1988), "A Hydrological Perspective", in the *Journal of Hydrology*, **100**, pp.3-28.

Konikow, L.F. (1986), "Predictive Accuracy of a Ground-Water Model - Lessons From a Postaudit", in *Ground Water*, **24**:2, pp.173-184.

Konikow, L.F. and Bredehoeft, J.D. (1992a), "Groundwater Models Cannot be Validated", in *Advances in*

Water Resources, **15**:1, pp.75-83.

Konikow, L.F. and Bredehoeft, J.D. (1992b), "Reply to Comment", in *Advances in Water Resources*, **15**:6, pp.371-372.

Konikow, L.F. (1986), "Predictive Accuracy of a Ground-Water Model - Lessons From a Postaudit", in *Ground Water*, **24**:2, pp.173-184.

Konikow, L.F. and Person, M. (1985), "Assessment of Long-Term Salinity Changes in an Irrigated Stream-Aquifer System", in *Water Resources Research*, **21**:11, pp.1611-1624.

Konikow, L.F. and Swain, L.A. (1990), "Assessment of Predictive Accuracy of a Model of Artificial Recharge Effects in the Upper Coachella Valley, California", in Simpson, E.S. and Sharp, J.M. (eds.), *Selected Papers on Hydrogeology*, 28th International Geological Congress, Washington D.C., 9-19 July 1989.

Koopman, F.C. (1975), "Estimated Ground-Water Flow, Volume of Water in Storage, and Potential Yield of Wells in the Pojoaque River Draina0ge Basin, Santa Fe County, New Mexico", United States Geological Survey, Open-File Report 74-159.

Kroes, P. A. and Sarlemijn (1989), "Fundamental Laws and Physical Reality", in Sarlemijn, A. and Sparnaay, M.J. (eds.), *Physics in the Making*, Elsevier Science Publishers, pp.303-328.

Kuhn, T. (1962), *The Structure of Scientific Revolutions*, University of Chicago Press, 2nd Edition (1970).

Kuhn, T. (1964), "A Function For Thought Experiments", in Hacking, I. (ed.) (1981), *Scientific Revolutions*, Oxford University Press, pp.6-27.

Lakatos, I. and Musgrave, A. (eds.) (1970), *Criticism and the Growth of Knowledge*, Cambridge University Press.

Lauden, L. (1977), *Progress and Its Problems: Towards a Theory of Scientific Growth*, University of California Press.

Lauden, L. (1984), *Science and Values: The Aims of Science and Their Role in Scientific Debate*, University of California Press.

Lauden, R., Lauden, L. and Donovan, A. (eds.) (1988), *Scrutinizing Science: Empirical Studies of Scientific Change*, Kluwer Academic Publishers.

Lehr, J.H. (1980), "To Model or Not to Model - That is the Question", in *Ground Water*, **18**:2, pp.106-107.

Li, C., Wilson, J.L. and Hofmann, P. (1993), "The Influence of Pore Size on the Diffusion Coefficient Inside Porous Media: A Pore Scale Numerical Experiment", Proceedings of the Society for Industrial and Applied Mathematics *(SIAM) Conference on Mathematical and Computational Issues* *in the Geosciences*. Houston, TX.

Li, C., Wilson, J.L. and Hofmann, P. (1993), "Fickian and Non-Fickian Phenomena in Porous Media: A Microscopic Numerical Experiment", in H.J.Morel-Seytoux (ed.): Proceedings of the 13th Annual American Geophysical Union, *AGU Hydrology Days*, Fort Collins, CO, pp.149-160.

Lohman, S.W. (1979), "United States Geological Survey Professional Paper 708", in *Ground-Water Hydraulics*, pp.61-67.

Lowdin, P-O. (1967), "Nature of Quantum Chemistry", in the *International Journal of Quantum Chemistry*, **1**:1, pp.7-12.

Maloszewski, P. and Zuber, A. (1992), "On the Calibration and Validation of Mathematical Models for the Interpretation of Tracer Tests in Groundwater", in *Advances in Water Resources*, **15**:1, pp.47-62.

Manley, K. (1978), "Structure and Stratigraphy of the Espanola Basin, Rio Grande Rift, New Mexico", United States Geological Survey, Open-File Report 78-667.

Marcus Aurelius (121-80 AD), *Meditations (Writings to Himself)*

Mayo, D.G. and Hollander, R.D. (eds.) (1991), *Acceptable Evidence: Science and Values in Risk Management*, Oxford University Press.

Mazur, A. (1981), *The Dynamics of Technical Controversy*, Communications Press.

McAda, D.P. (1990), "Simulation of the Effects of Ground-Water Withdrawal from a Well Field Adjacent to the Rio Grande, Santa Fe County, New Mexico", United States Geological Survey, Water-Resources Investigations Report 89-4184.

McAda, D.P. and Wasioliek, M. (1988), "Simulation of the Regional Geohydrology of the Tesuque Aquifer System Near Santa Fe, New Mexico", United States Geological Survey, Water-Resources Investigations Report 87-4056.

McMullin, E. (1978a), "Structural Explanation", *American Philosophical Quarterly*, **15**:2, pp.139-147.

McMullin, E. (1978b), "The Conception of Science in Galileo's Work", in Butts, R.E. and Pitt, J. (eds.) (1978), *New Perspectives on Galileo*, D. Reidel Publishing Co., pp.209-257.

McMullin, E. (1985), "Galilean Idealization", in *Studies in the History of the Philosophy of Science*, **16**:3, Pergamon Press Ltd., pp.247-273.

McMullin, E. (1987), "Explanatory Success and the Truth of Theory", in Rescher, N. (ed.), *Scientific Inquiry in Philosophical Perspective*. Lanham: University Press of America, pp.51-73.

Medawar, P.B. (1967), *The Art of the Soluble*, Methuen & Co. Ltd.

Medawar, P.B. (1969), *Induction and Intuition in Scientific Thought*, American Philosophical Society.

Medawar, P.B. (1991), *The Threat and the Glory*, Oxford University Press.

Meinzer, O.E. and Fishel, V.C. (1934), "Tests of Permeability With Low Hydraulic Gradients", in *Transactions of the American Geophysical Union*, **15**, pp.405-503.

Mercer, J. W. and Faust, C. R. (1981), *Ground-Water Modeling*. Reston: National Water Well Association.

Mourant, W.A. (1980), "Hydrologic Maps and Data for Santa Fe County, New Mexico, New Mexico State Engineer Office basic data report.

Mott, N. (1941), "Application of Atomic Theory to Solids", in *Nature*, **147**, pp.623-624.

Mulkay, M. and Gilbert, G.N. (1981), "Putting Philosophy to Work: Karl Popper's Influence on Scientific Practice", in *Philosophy of the Social Sciences*, **11**, pp.389-407.

Muskat, M. (1937), *Flow of Homogeneous Fluids Through Porous Media*, McGraw-Hill.

Nash, J.E. (1992), "Hydrology and Hydrologists - Reflections", in O'Kane, P. (ed.) (1992), *Advances in Theoretical Hydrology: A Tribute to James Dooge*, Elsevier, pp.191-199.

Nash, J.E., Eagleson, P.S., Philip, J.R. and Van Der Molen, W.H. (1990), "The Education of Hydrologists", in *Hydrological Sciences Journal*, **35**:6, pp.597-607.

National Academy Press (Fiering, M.B., chair) (1982), *Studies in Geophysics: Scientific Basis of Water-Resource Management*.

National Academy Press (Schwartz, F.W., chair) (1990a), *Ground Water Models: Scientific and Regulatory Applications*.

National Academy Press (O'Melia, C.R., chair) (1990b), *Keeping Pace with Science and Engineering: Case Studies in Environmental Regulation*.

National Academy Press (Parker, F.L., chair) (1990c), *Rethinking High-Level Radioactive Waste Disposal, A Position Statement of the Board on Radioactive Waste Management*, by the Commission on Geosciences, Environment, and Resources of the National Research Council.

National Academy Press (Eagleson, P.S., chair) (1991), *Opportunities in the Hydrologic Sciences*.

Nietzsche, F. (1886), *Beyond Good and Evil: Prelude to a Philosophy of the Future*, Vintage Books Edition (1989).

Ohm, G.S. (1827), *The Galvanic Circuit Investigated Mathematically*, Kraus (1969).

O'Kane, P. (ed.) (1992), *Advances in Theoretical Hydrology: A Tribute to James Dooge*, Elsevier.

Oreskes, N., Shrader-Frechette, K., and Belitz, K. (1994), "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences", in *Science*, **263**, 4 Feb 1994, pp.641-646.

Peters, R.H. (1991), *A Critique for Ecology*, Cambridge University Press.

Peters-Lidard, C.D. and Wood, E.F. (1994), "Estimating Storm Areal Average Rainfall Intensity in Field Experiments", in *Water Resources Research*, **30**:7, pp.2119-2131.

Philip, J.R. (1988), "The Fluid Mechanics of Fracture and Other Junctions", in *Water Resources Research*, **24**:2, pp.239-246.

Philip, J.R. (1992), "Hydrology in the Real World", in O'Kane, P. (ed.) (1992), *Advances in Theoretical Hydrology: A Tribute to James Dooge*, Elsevier, pp.201-207.

Pinder, G. (1986), Deposition, 10 Jan 1986, in United States District Court, District of Massachusetts, Civil Action 82-1672S: Anne Anderson *et al. v.* W.R. Grace *et al.*

Planck, M. (1933), *Where is Science Going?*, The Humanities Press, New York.

Poincaré, H. (1905), *Science and Hypothesis*, Dover Edition (1952).

Poincaré, H. (1905), *The Value of Science*, Dover Edition (1958).

Popper, K.R. (1959), *The Logic of Scientific Discovery*, Routledge (1992).

Popper, K.R. (1963), *Conjectures and Refutations: The Growth of Scientific Knowledge*, Routledge.

Popper, K.R. (1974), *Autobiography*, in Schillp, P.A. (ed.) (1974), *The Philosophy of Karl Popper*, Open Court, pp.3-181.

Posson, D.R., Hearne, G.A., Tracy, J.V. and Frenzel, P.F. (1980), "Computer Program for Simulating Geohydrologic Systems in Three Dimensions", United States Geological Survey, Open-File Report 80-421.

Pramer, D. (1985), "Terminal Science", in *Bioscience*, **35**, p.141.

Purtyman, W.D. and Adams, H. (1980), "Geohydrology of Bandelier National Monument, New Mexico", Los Alamos Scientific Laboratory Informal Report LA-8461-MS.

Purtymun, W.D. and Johansen, S. (1974), "General Geohydrology of the Pajarito Plateau", in *Guidebook to Ghost Ranch*, New Mexico Geological Society, 25th Field Conference, pp.347-349.

Putnam, H. (1969), "The 'Corroboration' of Theories", in Schilpp, P.A. (ed.) (1974), *The Philosophy of*

Karl Popper, Open Court, pp.221-240; and in Hacking, I. (ed.) (1981), *Scientific Revolutions*, Oxford University Press, pp.60-79 (pagination in notes is from the Hacking book).

Quinton, A. (1980), "Bacon", in *Renaissance Thinkers*, Oxford University Press (1993), pp.113-204.

Reiland, L.J. and Koopman, F.C. (1975), "Estimated Availability of Surface and Ground Water in the Pojoaque River Drainage Basin, Santa Fe County, New Mexico", United States Geological Survey, Open-File Report 74-151.

Resnick, D.B. (1992), "Convergent Realism and Approximate Truth", in Hull, D., *et al.*, (eds.), *PSA 92*, Proceedings of the Philosophy of Science Association, 1992, p.421-434.

Robertson, J.W. and Gale, J.E. (1990), "A Laboratory and Numerical Investigation of Solute Transport in Discontinuous Fracture Systems", in *Ground Water*, **28**:1, pp.25-36.

Rogers, P.P. (1983), "Analyzing Natural Systems" (book review), in *EOS, Transactions of the American Geophysical Union*, **64**:25, p.419.

Rothbart, D. and Slayden, S.W. (1994), "The Epistemology of a Spectrometer", in *Philosophy of Science*, **61**, pp.25-38.

Santa Fe Metropolitan Water Board (1984), "Santa Fe Regional Water Supply System: Report on Baseline Data, Goals, Policy and Studies".

Shapere, D. (1966), "Meaning and Scientific Change", in Hacking, I. (ed.) (1981), *Scientific Revolutions*, Oxford University Press, pp.28-59.

Shapiro, A.M. and Nicholas, J.R. (1989), "Assessing the Validity of the Channel Model of Fracture Aperture Under Field Conditions, in *Water Resources Research*, **25**:5, pp.817-828.

Schilpp, P.A. (ed.) (1974), *The Philosophy of Karl Popper*, Vols. 1-2, Open Court.

Schumm, S.A. (1991), *To Interpret the Earth: Ten Ways to be Wrong*, Cambridge University Press.

Shrader-Frechette, K. S. (1988): "Values and Hydrogeological Method: How Not to Site the World's Largest Nuclear Dump", in Byrne, J. and Rich, D. (eds.), *Planning for Changing Energy Conditions*, New Brunswick: Transaction Books, pp.101-137.

Shrader-Frechette, K. S. (1989), "Idealized Laws, Antirealism, and Applied Science: A Case in Hydrogeology", in *Synthese*, **81**, pp.329-352.

Shrader-Frechette, K. S. (1993), *Burying Uncertainty: Risk and the Case Against Geological Disposal of Nuclear Waste*, University of California Press.

Shrader-Frechette, K. S. and McCoy, E.D. (1994), "Applied Ecology and the Logic of Case Studies", in

Philosophy of Science, **61**, pp.228-249.

Simpson, E.S. and Sharp, J.M. (eds.) (1989), *Selected Papers on Hydrogeology*, 28th International Geological Congress, Washington, D.C., 9-19 July 1989.

Slater, J.C. (1967), "The Current State of Solid-State and Molecular Theory", in the *International Journal of Quantum Chemistry*, **1**:1, pp.37-102.

Smith, R.E., Goodrich, D.R., Woolhiser, D.A. and Simanton, J.R. (1994), "Comment on 'Physically Based Hydrologic Modeling, 2: Is the Concept Realistic?' by R.B. Grayson, I.D. Moore, and T.A. McMahon", in *Water Resources Research*, **30**:3, pp.851-854.

Spiegel, Z. and Baldwin, B. (1963), "Geology and Water Resources of the Santa Fe Area, New Mexico", United States Geological Survey, Water-Supply Paper 1525.

Stearns, N.D. (1927), "Laboratory Tests on Physical Properties of Water-Bearing Materials", United States Geological Survey, Water-Supply Paper 596, pp.144-159.

Stewart, J.A. (1986), "Drifting Continents and Colliding Interests: A Quantitative Application of the Interests Perspective", in *Social Studies of Science*, **16**, pp.261-279.

Tarricone, P. (1994), "The Politics of Nuclear-Waste Disposal", in *Civil Engineering*, March 1994, pp.56-58.

Thomas, L. (1980), *Late Night Thoughts on Listening to Mahler's Ninth Symphony*, Viking Press.

Torgersen, T. (1994), "Chapman Conference on Hydrogeological Processes Promotes Interaction", in *EOS*, 22 Nov 1994, pp.547,557.

Tsang, C.F. (1991), "The Modeling Process and Model Validation", in *Ground Water*, **29**:6, pp.825-831.

Tsang, Y.W. (1992), "Usage of 'Equivalent Apertures' for Rock Fractures as Derived From Hydraulic and Tracer Tests", in *Water Resources Research*, **28**:5, pp.1451-1455.

United States Department of Energy, Office of Civilian Radioactive Waste Management (1986), *DOE Environmental Assessment - Yucca Mountain Site, Nevada Research and Development Area*, DOE/RW-0073, Vol.2.

United States Nuclear Regulatory Commission, Office of Nuclear Regulatory Research (1984), *NRCA Revised Modelling Strategy Document for High Level Waste Performance Assessment.*

Vincenti, W.G. (1990), *What Engineers Know and How They Know It: Analytical Studies From Aeronautical History*, Johns Hopkins University Press.

Wang, H. F. and Anderson, M. (1982), *Introduction to Groundwater Modeling: Finite Difference and Finite*

Element Methods, San Francisco: W.H. Freeman and Company.

Wasiolek, M. (in press), "Subsurface recharge to the Western Side of the Sangre de Cristo Mountains Near Santa Fe, New Mexico", United States Geological Survey, Water-Resources Investigations Report 94-4072.

Watson, I. and Burnett, A.D. (1993), *Hydrology: An Environmental Approach*, Buchanan Books.

Weber, R.L. (ed.) (1973), *A Random Walk in Science*, The Institute of Physics, London and Bristol.

Weinberg, G.M. (1975), *An Introduction to General Systems Thinking*, Wiley-Interscience.

Wilson, J.L., Li, C. and Hofmann, P. (1993), "Mixing Rules of Pore and Fracture Junctions", in Proceedings of *3rd Annual WERC Technology Development Conference*, Waste Education and Research Consortium, Las Cruces, NM.

Wilson, J.L. and Hofmann, P. (1994), "On the Role of Groundwater Modeling in Hydrogeologic Practice", Keynote address to the groundwater session of the *Geological Association of Canada/Mineralogical Association of Canada Joint Annual Meeting*, Waterloo, Ontario, Canada.

Woodward J. (1989), "Data and Phenomena", in *Synthese*, **79**, pp.393-472.

This document is accepted on behalf of the faculty of the Institute by the following committee:

Adviser

April 12, 1995

Date

I release this document to New Mexico Institute of Mining and Technology.

Students Signature 12 April 1995

Date