The [Geo]Scientific Method; hypothesis testing and geoscience proposal writing for students

Michelle J. Markley[1]

ABSTRACT

Most undergraduate-level geoscience texts offer a paltry introduction to the nuanced approach to hypothesis testing that geoscientists use when conducting research and writing proposals. Fortunately, there are a handful of excellent papers that are accessible to geoscience undergraduates. Two historical papers by the eminent American geologists G. K. Gilbert and T. C. Chamberlin (Gilbert, 1886; Chamberlin, 1897) were the first to fully articulate and explore the method of multiple working hypotheses. Both papers still make for inspirational reading. A long essay on the scientific method by Johnson (1933) presents both a recipe for rigorous scientific thinking and a traditional but detailed articulation of linear hypothesis testing using geologic examples. More recently, papers by Frodeman (1995) about the fundamentally non-linear nature of interpretation and reasoning in the geosciences and Cleland (2001) about a "smoking gun" approach to validating hypotheses are helpful articulations of the geoscientific method, i.e. a shared understanding of how geoscientists articulate, frame, and tackle research questions.

INTRODUCTION

What first-year undergraduates know about scientific methodology likely comes from a high school science class in which they learned a linear scientific method: identify a problem, make a hypothesis, gather data, and test the hypothesis. Yet geologists commonly follow a nonlinear path towards testing hypotheses, and we tend to work on many hypotheses at a time, rarely fully accepting or rejecting any of them. For undergraduate geology majors (and even graduate students), learning the nuances of testing hypotheses like a practicing geologist is a transition, and few college-level geology textbooks adequately support that transition. For example, the textbook I use to teach Physical Geology, Marshak's Essentials of Earth, articulates a linear version of the scientific method. By way of more detailed explanation, Essentials of Earth carefully distinguishes between a hypothesis and a theory, using as a case study the development of plate tectonic theory from the hypothesis of continental drift. Comparable descriptions of the nature of hypothesis testing, the distinction between a hypothesis and a theory, and the case study of the history of ideas about drift and tectonics occur early in both Monroe, Wicander, and Hazlett's Physical Geology and Chernikoff and Whitney's Geology.

It seems natural for introductory geology textbooks to use plate tectonic theory to illustrate elements of the scientific method. I joke with my Physical Geology class that the story of the plate tectonic revolution is the American geologist's Passover story; we delight in its telling and retelling. The story offers opportunities to elaborate upon any number of subdisciplines in geology, from paleontology to rock magnetism. Wegener makes for a wonderful hero – dead in the field before his time. And the embrace of plate tectonics in America is certainly an excellent case study of the development of a theory: plate tectonics only gained widespread embrace after support developed within many different subdisciplines; it was a hypothesis that has now survived repeated challenges; and it has predictive power. Perhaps introductory geology textbooks focus on this careful explanation of the scientific usage of the word theory because of the ongoing culture wars over evolution (which is, after all, "only a theory"). We hope that our undergraduates become scientifically literate, and literacy involves a clear understanding of what a scientific theory is.

Yet we also hope that our undergraduate geoscience majors become more than literate. We hope that they become skilled researchers and sophisticated thinkers. Specifically, we want them to learn a nuanced and realistic approach to designing hypotheses and developing tests for them. I have found surprisingly little support for this learning in college-level geoscience textbooks. For example, the two texts I use to teach upper-level courses each year, Davis and Reynolds' Structural Geology of Rocks and Regions and Prothero and Dott's Evolution of the Earth (books I love for many other reasons), offer next to nothing. Among the textbooks I regularly consult for teaching, the only book to offer detailed and thoughtful description of the way that geologists think about designing and testing hypotheses is Tectonics by Twiss and Moores (in their "Interlude"). Fortunately, more resources are out there in the form of a handful of excellent papers in widely available journals. Below, I discuss five classic papers written for American geologists (and mostly by American geologists) about methodology and the nature of hypothesis testing in the geosciences. Each paper offers an example other than plate tectonic theory, and each is accessible enough to use as supplementary reading in an undergraduate geology class (I hope, too, that this piece is accessible and succinct enough to use as supplementary reading in an undergraduate geology class). Finally, in the conclusions, I share some of my experiences using proposal writing as a way to teach the geoscientific method to undergraduate geology majors.

MULTIPLE WORKING HYPOTHESES AND HYPOTHESIS TESTING

The methodological concept that is most likely a part of the everyday vocabulary of a geologist is "the method of multiple working hypotheses," espoused by the eminent geologist T.C. Chamberlin (1897). Essentially, Chamberlin warns against focusing one's research on testing a single hypothesis at a time, because one is likely to favor it to the detriment of understanding the problem.

---

[1] Department of Geology and Geography, Mount Holyoke College, 50 College Street, South Hadley, MA 01075; mmarkley@mtholyoke.edu

Instead, he argues for developing multiple hypotheses in the early stages of a study. Chamberlin uses a family metaphor for his argument:

> In developing the method of multiple working hypotheses, the effort is to bring up and into view every rational explanation of the phenomenon in hand and to develop every tenable hypothesis relative to its nature, cause or origin, and to give all of these as impartially as possible a working form and due place in the investigation. The investigator thus becomes the parent of a family of hypotheses; and by his parental relations to all is morally forbidden to fasten his affections unduly upon any one.

At its simplest, Chamberlin's essay is an admonition to keep an open mind. His idea of multiple working hypotheses permeates the geologic literature on methodology (some papers refer to it as "The Method"), and it is a touchstone for most other writing on earth science methods.

A closely related essay by G. K. Gilbert (1886) also stresses the importance of multiple working hypotheses:

> The great investigator is primarily and preeminently the man [sic] who is rich in hypotheses. In the plenitude of his wealth he can spare the weaklings without regret; and having many from which to select, his mind maintains a judicial attitude. The man who can produce but one, cherishes and champions that one as his own, and is blind to its faults. With such men, the testing of alternative hypotheses is accomplished only through controversy. Crucial observations are warped by prejudice, and the triumph of the truth is delayed.

Unlike Chamberlin’s essay, in which he outlines reasons for using multiple working hypotheses, Gilbert’s essay teaches by example. Gilbert details a field-oriented geomorphology problem and three hypotheses to explain it. The problem he cites is the variable elevation of the paleo-shoreline of ancient Lake Bonneville (a larger, Pleistocene manifestation of Great Salt Lake in Utah). Assuming that the shoreline was level when the lake was full, Gilbert reasons that the shoreline is no longer level because of subsequent crustal warping. Gilbert focuses on the difference between the highest and lowest shoreline elevations. He presents calculations by himself and a colleague, based on assumptions about the physical properties of the earth's crust, on the maximum amount of uplift that could result from each of three hypotheses. Only one of his hypotheses appears to be capable of producing as much uplift as the observed elevation difference, so Gilbert tentatively discards the other two hypotheses and recommends further tests of the survivor.

Gilbert’s example is astonishingly current for two reasons. Firstly, rather than validating any one hypothesis (as the linear scientific method would dictate), he selects among his hypotheses by elimination. In his own words, he casts aside the weaklings. This method of hypothesis testing, also called falsification, responds to the rub that scientists may never really prove that a hypothesis is true, because there is always the possibility that new data will show that it is false. Instead, the only way to stop working with a hypothesis is to show that it is false. Secondly, Gilbert tests his three hypotheses by running numbers in about the same manner as a modern geologist might design a handful of simple computer models. His modeling adds to the strength of the surviving hypothesis. Not only is it the only hypothesis left, it is also reasonable, given the best quantitative knowledge of earth processes and materials. A similar use of modeling to strengthen hypotheses is currently common in the geoscience literature.

A 1933 essay on the scientific method by Douglas Johnson offers a different approach to hypothesis testing. Johnson also uses an example of a geomorphic problem: the seaside association of rocky cliffs and erosional benches located above high tide line. In playing with this example, Johnson stresses the importance of analysis, which he defines as "the process of separating observations, arguments, and conclusions into their constituent parts, tracing each part back to its source and testing its validity, for the purpose of clarifying and perfecting knowledge." In other words, be careful with your assumptions. Johnson encourages multiple hypotheses. To test these hypotheses, he recommends a deductive process of analyzing each hypothesis in order to determine its consequences: for example, details that might characterize the benches in the case that the hypothesis is valid. These details are useful for validating or falsifying (i.e. eliminating) individual hypotheses, the basic process demonstrated by Gilbert. Some geologists call this method prediction because one predicts, based on the various hypotheses, what one might observe in the field (note that this use of the word prediction does not really involve what processes will occur in the future).

Johnson's example contrasts with Gilbert's because his tests involve gathering or reconsidering field data rather than constructing simple models. For example, Johnson selects a hypothesis regarding storm waves, and he conjectures that this hypothesis predicts that the benches have loose debris on them. Is there debris on the benches? Unfortunately, Johnson's example proves more confusing than illuminating at this point because he goes on to invalidate the assumptions underlying the discriminatory character of this prediction. Nonetheless, Johnson makes a useful point:

> There will usually be found, however, some one or more consequences peculiar to hypothesis A, while certain others are peculiar to hypothesis B, and so on. It is these unlike consequences which have the highest critical value in discriminating between valid and invalid hypotheses, and it is on these that the investigator will most depend on drawing conclusions.

In other words, thoughtful geologists will gather the data that will allow them to choose between competing hypotheses.

Another excellent essay on hypothesis testing in the geosciences is Carol Cleland's 2001 essay. She argues that geoscientists rarely eliminate hypotheses, for the same reason that Johnson steps away from eliminating his hypothesis about benches: the realization (or fear) that some of the assumptions underlying the usefulness of the test are invalid. In addition, Cleland argues that geoscience hypotheses are unusually difficult to falsify. Events are too complicated and time frames too long to set up experimental tests. Critical evidence might be lost or beyond our ability to identify. Instead, she posits:

> A look at the actual practices of historical researchers, however, reveals that the main emphasis is on finding positive evidence- a smoking gun. A smoking gun is a trace [i.e. data gathered from the rock record] that picks out one of the competing hypotheses as providing a better causal explanation for the currently available traces than the others.

Cleland's "smoking gun" idea bears some resemblance to Johnson's "unlike consequence," but Cleland's essay contains a much clearer, more detailed, and more modern exposition of the idea. In addition, she gives an example that works: the cause of extinction of the dinosaurs. High iridium concentrations in sediments deposited at the same time as the extinction are the smoking gun. Because iridium is much more common in space than on Earth, researchers interpret this observation as good evidence for a meteorite impact. The impact and the extinctions occurred at the same time, supporting a cause-and-effect relationship.

Cleland argues that field data, rather than laboratory experiments or computer models, provide us with smoking guns:

> This brings us to a crucial point: although computer-aided models may suggest what to look for in nature, and traces and some auxiliary assumptions may be investigated in the laboratory, one cannot experimentally test a historical hypothesis per se; to recapitulate, the time frame is too long and the test conditions too complex to be replicated in the lab.

Certainly, most geoscience research starts with field observations, which lead to questions based on these observations. Ultimately, hypotheses must agree with field data. This necessity is one of Johnson's central points. Cleland is arguing, however, that only field data allow the geoscientist to select one hypothesis among many, if the hypothesis concerns the causes or nature of past events. In contrast, Gilbert's example from Lake Bonneville uses simple calculations, quite like computer-aided modeling, to falsify two of his hypotheses, allowing him to tentatively select the only remaining hypothesis. These two approaches differ significantly, and both are useful. Even if experimental or model results are not smoking guns, they allow geoscientists to explore which hypotheses are possible, given the current state of quantitative knowledge of the pertinent physical processes.

Ultimately, our knowledge of what is possible is constantly evolving. As a result, models and experiments do not provide a final test. Nor do field data. For example, some geologists continue to favor different causes for the extinction of the dinosaurs, even though the field data are clear that a giant meteorite impact occurred at that time. Perhaps the meteorite didn't cause the extinction, or perhaps the story is more complicated. Why did mammals survive? The geoscience community is diverse in its thinking, and geologic events and phenomena are complex. Can we ever consider any specific geologic hypotheses to be fully tested? Having posed this question, it is illuminating to turn again to Chamberlin's 1897 essay.

COMPLEXITY AND CIRCULAR REASONING

Although Chamberlin's essay focuses mainly on arguments for using multiple working hypotheses, he also provides the example of the origin of the Great Lakes:

> The mooted question of the origin of the Great Lake basins may serve as an illustration. Several hypotheses have been urged by as many students of the problem of the cause of these great excavations. All of these have been pressed with great force and with an admirable array of facts. Up to a certain point we are compelled to go with each advocate. It is practically demonstrable that these basins were river valleys antecedent to the glacial incursion. It is equally demonstrable that there was a blocking up of outlets. We must conclude then that the present basins owe their origin in part to the preexistence of river valleys and to the blocking up of outlets by drift. That there is a temptation to rest here, the history of the question shows. But on the other hand it is demonstrable that these basins were occupied by great lobes of ice and were important channels of glacial movement. The leeward drift shows much material derived from their bottoms. We cannot therefore refuse assent to the doctrine that basins owe something to glacial excavation. Still again it has been urged that the earth's crust beneath these basins has flexed downward by the weight of the ice load and contracted by its low temperature and the basins owe something to crustal deformation. This third cause tallies with certain features not readily explained by the others. And still it is doubtful whether all of these combined constitute an adequate explanation of the phenomena. Certain it is, at least, that the measure of participation of each must be determined before a satisfactory elucidation can be reached. The full solution therefore involves not only the recognition of multiple participation but an estimate of the measure and mode of each participation. For this the simultaneous use of a full staff of working hypotheses is demanded.

Chamberlin's example is surprising because it gives no sense of hypothesis testing. He presents multiple hypotheses, but he does not eliminate any hypotheses, nor does he shore up support for others with a smoking gun, an experiment, or a realistic quantitative calculation. Rather than appeal to any further testing of these hypotheses, Chamberlin indicates that future research might weigh their relative roles or identify even more hypotheses to add to the list of factors that contributed to the development of the Great Lakes. Chamberlin is fully aware of the significance of his example. The final paragraph of his essay begins: "the studies of the geologist are peculiarly complex. It is rare that his problem is a simple unitary phenomenon explicable by a single simple cause."

Chamberlin's articulation of the complexity of geoscience explanations rings true. But none of the essays discussed above fully addresses the significance of this complexity for hypothesis testing. A fascinating 1995 essay by Robert Frodeman fleshes out some of these concerns from a different perspective. Rather than understanding hypothesis testing as the only goal of research, Frodeman understands it as one of the many tools that geologists use to understand past events. Another tool, for example, is analogies between modern and ancient processes and events. As a description of the way geologists reason, Frodeman uses an example of a geologist looking at an outcrop:

> More to the point, our understanding of an outcrop is based on our understanding of the individual beds, which are in turn made sense of in terms of their relationship to the entire outcrop. This back-and-forth process of reasoning operates at all levels; wholes at one level of analysis become parts at another. Thus, our understanding of a region is based on our interpretation of the individual outcrops in that region, and vice versa; and our interpretation of an individual bed within an outcrop is based on our understanding of the sediments and structures that make up that bed, and vice versa. On a still more complex level, our overall comprehension of the Cenomanian-Turonian boundary event is determined through an intricate weighing of the various types of evidence (e.g. lithology, macro- and micro-paleontology, and geochemistry). This overall interpretation is then used to evaluate the status of the individual pieces of evidence. Such circular reasoning is usually viewed as a vice, a logical fallacy to be avoided at all costs. But Heidegger argued that this type of circularity is not only unavoidable, it is actually, if properly handled, the means by which understanding progresses.

Anyone who has inspected an outcrop or participated in a panel discussion with a group of geologists can see that Frodeman's description is accurate. This circularity is different from the more linear process we invoke when we talk about hypothesis testing. In addition, it is easy to see how both field data and experiments or quantitative models fit into the circular reasoning process. The ease of this fit is clear in the geologic literature, where geologists often present both new data and model results in the same publications.

What are the implications of Frodeman's description for the role of hypothesis testing in the geoscientific method? Do geologists even test hypotheses? Yes, hypothesis testing is an intuitive and ubiquitous way of thinking and asking questions, even though it is not the only way that research progresses. Hypothesis testing is a fundamental part of our common methodology, even though this common methodology encompasses the diversity of approaches outlined above.

IDEAS FOR TEACHING

When I first started to think about how I might teach undergraduates a more nuanced and realistic sense of how working geoscientists articulate and test hypotheses, I decided to start assigning proposals as writing assignments in my Historical Geology class. Proposal writing requires a very clear statement of hypotheses, an understanding of how to test them, and a set of arguments showing that the hypotheses are worth testing. I wanted to give students experience writing a few different proposals over the course of a semester so that they could practice the format. Focusing these writing assignments was difficult for me at first because most undergraduate proposal-writing experience relates to independent research projects outside of the classroom (often a proposal for senior thesis research). Furthermore, most undergraduates, and even many graduate students, have not seen written arguments involving hypothesis testing because they have never read a research proposal. Proposals are almost always confidential; usually the only people who read them are the handful of scientists who review them.

How could I incorporate proposals into coursework? At first, I asked students to pick examples of controversies from their text and write proposals about competing hypotheses and ideas for how to move the thinking forward. Most students worked with the controversies that were best articulated in the text: causes of the Ordovician trilobite and the Pleistocene megafauna extinctions. After two years of reading papers on these two debates, I became frustrated with my own assignment. By reviewing stalled debates, my students may have been learning how to articulate scientific questions as hypotheses, but their papers were repetitive and relied too heavily on the text.

Over the past two years, I have used recent research in the journals Science and Nature as springboards for student proposals. Each year, I make a list of the past year's historical geology papers in these two journals. Each student picks two papers from the list and presents them to the class. Presentations occur each week, and other students can use any presentation (and associated paper) as a starting point for a written proposal to move the research forward. Students have a lot of choice: they choose which papers to present, and they choose which papers to explore more fully in their proposals. Consistently, the most significant challenge of the proposal is finding an appropriate scale for hypotheses--somewhere between testing the hypothesis of uniformitarianism and testing the hypothesis that someone might eventually find a more complete fossil of a poorly understood species. I remind students that most scientific research occurs in small steps, and many researchers start new projects by first considering what kinds of data they are able to gather or model. Then they consider what questions these data or models might address. Articulating and contextualizing these questions as hypotheses is often the last step in putting together a proposal. Even though writing proposals is challenging, students generally like the assignment: it brings cutting edge research into the classroom each week, it allows students to write about the topics and techniques that most interest them, and it encourages original and critical thinking. What's not to like about that?

REFERENCES: SUGGESTED READING

Chamberlin, T. C., 1897. The Method of Multiple Working Hypotheses. Journal of Geology v. 5, no. 8, p. 837-848.

Chernikoff, S. and D. Whitney, 2006. Geology, 4th edition, Prentice Hall.

Cleland, C. E., 2001. Historical science, experimental science, and the scientific method. Geology v. 29, no. 11, p. 987-990.

Davis, G. H. and S. J. Reynolds, 1996. Structural Geology of Rocks and Regions, 2nd edition, Wiley.

Frodeman, R., 1995. Geological Reasoning: Geology as an interpretive and historical science. Geological Society of America Bulletin v. 107, no. 8, p. 960-968.

Gilbert, G. K., 1886. The Inculcation of the Scientific Method by Example, with an illustration drawn from the Quaternary Geology of Utah. American Journal of Science v. 31, p. 284-299.

Johnson, D., 1933. Role of Analysis in Scientific Investigation. Bulletin of the Geological Society of America v. 44, p. 461-493.

Marshak, S., 2009. Essentials of Geology, 3rd edition, W. W. Norton and Company.

Monroe, J. S., R. Wicander, and R. Hazlett. 2006. Exploring the Earth, 6th edition, Brooks Cole.

Prothero, D. and R. Dott, 2009. Evolution of the Earth, 8th edition, McGraw Hill Science.

Twiss, E. M. and R. J. Moores, 1995. Tectonics, W. H. Freeman.

APPENDIX: TIPS FOR WRITING A GEOSCIENCE PROPOSAL

In the simplest terms, a proposal explains the nuts and bolts of the research you propose to do and why you propose to do it. A good proposal also positions this research in the context of pressing problems and questions. In terms of organization, a traditional proposal has four parts: (1) a statement of the hypothesis or hypotheses, (2) arguments that the hypotheses merit testing, (3) explanation of proposed tests, and (4) demonstration that you can successfully complete the tests. But before you sit down and write that proposal, you probably have some thinking to do.

Step one: Consider what has already been done and what could be done. What have other folks accomplished? What are some outstanding questions or problems? Which techniques and/or field areas have already been explored, and which show promise? You might find some ideas in the form of authors' suggestions for further research. You might think of some on your own. Identify the outstanding questions that interest you. Do not be afraid to think small. Although some research questions are ambitious (for example, the origin of the Great Lakes or the extinction of the dinosaurs), much of scientific research occurs in small steps. In fact, many researchers start new projects by first considering what kinds of data they are able to gather or model. Then they consider what questions these data or models might address.

Step two: make the transition from questions to testable hypotheses. Articulate as many hypotheses as you can. Remember that a hypothesis is not a question or a research goal; a hypothesis is a proposition that guides further research. You can distinguish a hypothesis from a question or a goal because a hypothesis is phrased so that it could be proven true or false. The biggest challenges of proposal writing are twofold: (1) developing enough sophistication to conceive (or even recognize) interesting hypotheses, and (2) exercising the wisdom to identify which of these hypotheses are testable. From your long list of hypotheses that you have brainstormed, choose two or a few interesting hypotheses that you could test. Beware - the most common flaw of an interesting hypothesis is that it is exceedingly difficult to test. Balance what interests you with what you might reasonably hope to accomplish.

Step three: write that proposal. Articulate your hypotheses as clearly and quickly as possible. Then flesh out the hypotheses. Explain how these hypotheses relate to the current state of research. Rather than simply outlining the work that has already been accomplished in the field, use arguments to establish that your hypotheses are plausible enough to merit the attention of researchers and explain why your hypotheses interest geologists (or some other constituency). Then explain the research that you might do to test these hypotheses. Do you propose to collect more field data? What kind? Where? Do you propose to design computer models? What variables and knowledge would you incorporate into these models? Why? Do you propose to conduct laboratory experiments? Where? What materials would you use and why? Explain exactly how these data, models, or experiments would test your hypotheses. In addition, explain your strategy. Do you propose to confirm or falsify your hypotheses? Do you have predictions for your results? Would you look for a smoking gun? If you have multiple working hypotheses, explain how the hypotheses relate to each other. Are they compatible or mutually exclusive? Would confirmation or falsification of one of the hypotheses affect the other hypotheses? As you write, remember that a research proposal does a lot of explaining, yet it is entirely different from a research paper or a lab report. You are convincing your reader that you can do research that should be done.