AGU PUBLICATIONS

Water Resources Research

EDITORIAL

10.1002/2017WR020584

Correspondence to:
G. Blöschl,
bloeschl@hydro.tuwien.ac.at

Citation:
Blöschl, G. (2017), Debates—Hypothesis testing in hydrology: Introduction, *Water Resour. Res.*, *53*, 1767–1769, doi:10.1002/2017WR020584.

Received 20 FEB 2017
Accepted 20 FEB 2017
Accepted article online 27 FEB 2017
Published online 29 MAR 2017

Debates—Hypothesis testing in hydrology: Introduction

Günter Blöschl[1]

[1]Institute of Hydraulic and Water Resources Engineering, Vienna University of Technology, Vienna, Austria

Abstract This paper introduces the papers in the "Debates—Hypothesis testing in hydrology" series. The four articles in the series discuss whether and how the process of testing hypotheses leads to progress in hydrology. Repeated experiments with controlled boundary conditions are rarely feasible in hydrology. Research is therefore not easily aligned with the classical scientific method of testing hypotheses. Hypotheses in hydrology are often enshrined in computer models which are tested against observed data. Testability may be limited due to model complexity and data uncertainty. All four articles suggest that hypothesis testing has contributed to progress in hydrology and is needed in the future. However, the procedure is usually not as systematic as the philosophy of science suggests. A greater emphasis on a creative reasoning process on the basis of clues and explorative analyses is therefore needed.

> What we really cannot do is deal with actual, wet water [..]. That is the central problem which we ought to solve some day, and we have not.
>
> Richard Feynman, *The Feynman Lectures on Physics* (1963)

Hydrology is a hard subject. In his celebrated lectures, Richard Feynman refers to turbulence as the main unresolved problem of classical physics that cannot be analyzed from first principles. Indeed, turbulence complicates matters for hydrologists both directly—in the study of overland and open channel flow—and indirectly—in the study of subsurface flow, which is often controlled by the legacy of turbulent flow through the erosion and sedimentation processes that formed the soils. Additionally, chemical and biological processes play a role, as well as human activities. Hydrologists study these processes at the landscape scale where observations are difficult, in particular those of the subsurface (alas, if only we had ground penetrating TV!), and where boundary conditions cannot be controlled. This implies that we can rarely do our own experiments and have to, instead, content ourselves with whatever boundary conditions Nature provides.

These specific characteristics of our study objects have implications for the process of acquiring new knowledge, or making progress, in the field of hydrology. The standard approach to advancing natural sciences, known as the scientific method, mainly developed in the 17th century, involves making hypotheses about how nature works, deriving predictions from them as logical consequences, and then carrying out experiments based on those predictions to determine whether the original hypothesis was correct. New hypotheses are motivated when past hypotheses prove inadequate and are often inspired by experiments or observations. One of the most famous experiments in scientific history is the Michelson-Morley experiment on the speed of light which eventually led to special relativity. It did have controlled boundary conditions and it was repeated again and again with increasing accuracy to test alternative hypotheses. This kind of experimentation is rarely feasible in hydrology. Research in hydrology is therefore not easily aligned with the scientific method.

So just how do we make progress in hydrology? This question prompted me to invite four distinguished hydrologists and coauthors with a wide range of perspectives to share their views in a debate series: "Debates—Hypothesis testing in hydrology." The debaters were asked to comment on how hydrologic questions can be framed in a way that allows the testing of hypotheses and, ideally, generalization; how experiments should be designed to facilitate hypothesis testing; and whether there are alternatives to the hypothesis driven approach as a way to organize our research and make tangible scientific progress. The four debate pieces [*Baker*, 2017; *McKnight*, 2017; *Neuweiler and Helmig*, 2017; *Pfister and Kirchner*, 2017] indeed reflect a wide range of views. I commend the authors for their thoughtfulness and broad perspectives.

Anecdotal evidence, from discussions in the corridors, suggests this typical hydrologist's opinion on the subject: "Yeah, we need hypothesis testing, we do not do enough of it, we should do more." Most research funding agencies are adamant about requiring testable hypotheses (as noted by *Baker* [2017]; *McKnight* [2017]; *Pfister and Kirchner* [2017]), so if you want to get funded you better come up with a hypothesis in your proposal. Discussions at the Vienna Catchment Science Symposium on hypothesis testing in April 2016, where the idea of this debate series originated, however, were a little more subtle, and the articles in this debate series clearly take a more nuanced view on how and why we should test hypotheses.

The articles first discuss what can be considered a hypothesis in hydrology. Given the complexity of hydrological processes, hypotheses are often not single-line statements (such as "the speed of light changes with direction" in the Michelson-Morley experiment) but are enshrined in computer models [*Baker*, 2017; *Neuweiler and Helmig*, 2017; *Pfister and Kirchner*, 2017]. The models involve parameters that relate to one particular site, and a model structure that provides more generalizable understanding of how hydrological processes work. *Neuweiler and Helmig* [2017] emphasize that it is the model structure aspect where the knowledge gap is largest and detail specific processes that need closer scrutiny. However, *McKnight* [2017] and *Pfister and Kirchner* [2017] contend that hypotheses on cause and effect can be formulated quite separately from models and give examples from their own research.

One of the prerequisites of a useful hypothesis is its testability with a given experiment—or as is more often the case in hydrology—with a given data set. Negative outcomes of a hypothesis test (inconsistency of data and model) should be logically possible to avoid self-fulfilling prophesies. The tests should also be sufficiently powerful to detect inadequate or incorrect hypotheses; however, *Baker* [2017], *Neuweiler and Helmig* [2017], and *Pfister and Kirchner* [2017] note that testability may be difficult to achieve due to many reasons: too many degrees of freedom (overfitting, i.e., multiple model structures and multiple parameter combinations may fit the data), interactions of multiple processes (i.e., compensating processes leading to a lack of model identifiability), auxiliary assumptions (some of which may be implicit) that may or may not hold, and estimates of data accuracy that may themselves be flawed. These issues imply that the hypothesis testing procedure is often not as clear-cut in hydrology as some philosophers of science would suggest. Karl Popper proposed that hypotheses can only be falsified but not validated, and this view has become surprisingly popular in hydrology. In contrast, *Baker* [2017] and *Pfister and Kirchner* [2017] suggest that an approach based on rejection through singular falsifications may be doubtful, or at least not particularly useful, because of the above issues. A more symmetric view may therefore be more appropriate, where both consistency and inconsistency of the model with the data tell us something about the world, involving a gradual process of building confidence in a particular hypothesis as tests are repeated. Incidentally, this more symmetric balance between falsification and validation is consistent with both type I and type II errors being considered in formal statistical hypothesis testing.

In the scientific method, the process of testing hypotheses against experiments has the main goal of acquiring new knowledge. This is the science-as-knowledge perspective referred to by *Baker* [2017]. Notwithstanding the need for framing research proposals in terms of hypotheses, is this what we are actually doing in hydrology, and has it been useful? All four debate articles do see value in the hypothesis testing paradigm, although with different degrees of enthusiasm. *McKnight* [2017] suggests that hypotheses can help to focus long-term collaborative studies, assist with designing field monitoring strategies and in planning the overall research approach. She gives numerous examples of how the hypothesis-testing paradigm has helped in her own research progress, and provides guidance on how to obtain and test such hypotheses from a field work perspective. *Neuweiler and Helmig* [2017] address the hypothesis testing paradigm from a modeling perspective. They highlight the importance of model hypothesis testing but emphasize the many issues with the paradigm in practice.

Pfister and Kirchner [2017] also recognize advantages and disadvantages. On the one hand, they relate a number of success stories where hypothesis testing has led to substantial advances. One such example is the finding that event runoff response of catchments mainly consists of preevent water, although they also give examples of misuse of hypothesis testing where only those data are reported that are consistent with expectations (cherry picking). On the other hand, they emphasize that much of the important research in hydrology does not follow the hypothesis paradigm but is exploratory in character which allows for a broader, more creative perspective. They advocate for exploratory research to be treated on equal footing with the prevailing hypothesis paradigm. They argue that it would have been difficult for Galileo to obtain funding for his telescope, given he had no idea that he would discover the moons of Jupiter. Finally, *Baker* [2017] is the most critical of the debaters regarding the usefulness of the hypothesis paradigm. Drawing from ideas of Charles Peirce and other philosophers of science, he argues that the thought process of scientific discovery is complex and should take center stage, rather than the knowledge itself. While hypothesis tests have spurred major advances in hydrology, the hypothesis formulation is highly idealized and rarely conforms to how hydrologists perform their research. Nor is it conducive to leading to discovery, if the emphasis is on the systematic testing rather than on the generation of hypotheses which requires creativity.

Although the language of the four debate articles differs, there is the common theme that the process of hypothesis formulation, while being useful to structure research, may hamper creativity and thus scientific discovery in hydrology. *Baker* [2017] argues that the reasoning process is more important, similar to a medical doctor interpreting the symptoms of a disease, or Sherlock Holmes combining the clues of a crime scene to form a consistent narrative. It is the overall consistency of the narrative that makes the investigations fruitful and adds confidence to the process of inquiry. *McKnight* [2017] provides compelling examples of how surprises during field studies have provided new insights, e.g., on the role of photochemical processes in controlling stream chemistry. She concludes that creativity is a key element in successfully applying the hypothesis approach. *Neuweiler and Helmig* [2017] note that a creative mind is needed to come up with new formulations of model structure, as it is not usually feasible to include all possible combinations of the variables of interest in the model equations. *Pfister and Kirchner* [2017] note that exploratory research aimed at generating hypotheses has played an important role in hydrological research progress and argue that funding of exploratory research is essential for advancing hydrology. Broadly speaking, there is therefore agreement among the debaters that a creative reasoning process based on clues and explorative analyses should complement a more formal hypothesis testing framework.

When looking back over the decades, it is clear that there has been tremendous progress in hydrology. Formal hypothesis testing did play an important role, but so did creativity. Much of the progress has been facilitated by advances in observational and computational methods, and stimulated by a tighter exchange within the research community. Progress in hydrology thus requires a number of ingredients. A transparent scientific method is one of them. In practice, analysis steps may be more convoluted than the process of hypothesis formulation suggested by philosophers of science because we can rarely perform experiments with controlled boundary conditions. Perhaps we should acknowledge this situation by giving greater emphasis to a creative reasoning process based on clues, to parallel Sherlock Holmes's proverbial successes, as opposed to calibrating preconceived models to data. Making the reasoning process more presentable will foster our scientific discourse. As noted by *Pfister and Kirchner* [2017], these efforts need to go hand in hand with designing new measurement methods specifically geared at testing hypotheses, and conversely with generating hypotheses with an eye to the data that will actually be available to test them. Finally, hydrology is a community effort, so an open exchange of data, models and ideas is essential. Feedback and further comments from the *Water Resources Research* readership are most welcome.

References

Baker, V. R. (2017), Debates—Hypothesis testing in hydrology: Pursuing certainty versus pursuing uberty, *Water Resour. Res.*, *53*, doi:10.1002/2016WR020078.

McKnight, D. M. (2017), Debates—Hypothesis testing in hydrology: A view from the field: The value of hydrologic hypotheses in designing field studies and interpreting the results to advance hydrology, *Water Resour. Res.*, *53*, doi:10.1002/2016WR020050.

Neuweiler, I. and R. Helmig (2017), Debates—Hypothesis testing in hydrology: A subsurface perspective, *Water Resour. Res.*, *53*, doi:10.1002/2016WR020047.

Pfister, L. and J. W. Kirchner (2017), Debates—Hypothesis testing in hydrology: Theory and practice, *Water Resour. Res.*, *53*, doi:10.1002/2016WR020116.