Hydrol. Earth Syst. Sci., 14, 369–382, 2010
www.hydrol-earth-syst-sci.net/14/369/2010/

Hydrology and Earth System Sciences

HESS Opinions
On the use of laboratory experimentation: "Hydrologists, bring out shovels and garden hoses and hit the dirt"

M. G. Kleinhans, M. F. P. Bierkens, and M. van der Perk

Faculty of Geosciences, Universiteit Utrecht, P.O. Box 80115, 3508 TC Utrecht, The Netherlands

Received: 6 October 2009 – Published in Hydrol. Earth Syst. Sci. Discuss.: 29 October 2009
Revised: 11 February 2010 – Accepted: 13 February 2010 – Published: 23 February 2010

Abstract. From an outsider's perspective, hydrology combines field work with modelling, but mostly ignores the potential for gaining understanding and conceiving new hypotheses from controlled laboratory experiments. Sivapalan (2009) pleaded for a question- and hypothesis-driven hydrology where data analysis and top-down modelling approaches lead to general explanations and understanding of general trends and patterns. We discuss why and how such understanding is gained very effectively from controlled experimentation in comparison to field work and modelling. We argue that many major issues in hydrology are open to experimental investigations. Though experiments may have scale problems, these are of similar gravity as the well-known problems of fieldwork and modelling and have not impeded spectacular progress through experimentation in other geosciences.

1 Introduction

Viewed from the outsider's perspective of planetary science, or geomorphology, or meteorology, the science of hydrology uses but a subset of the tools for exploring nature as available to all geosciences. Much effort is put in field measurement and in physics-based modelling, wherein hydrological phenomena are reduced to the laws of physics following the optimistic agenda set by Freeze and Harlan (1969). Since their publication, fundamental problems of reductionism were encountered en route to a physics-based generally valid supermodel: underdetermination of model predictions by limited measurements of boundary conditions and initial conditions, underdetermination of model parameters, and underdetermination of predictions by ambiguity about the required level of simplification of physics-based relations in the model. The gravity of these problems, particularly the parameter problem, is attested by the fact that it was given a new name in hydrology: equifinality, but similar problems abound in the other geosciences. Yet there remained a societal need for hydrological predictions, so that much effort has been put into building and calibrating models for specific sites. It has been argued that this is a cul-de-sac for hydrology, because it does not lead to progress on big questions but leads to an unchecked growth of models applicable to one unique place only, which is exactly the opposite of what the reductionistic enterprise was about (Klemes, 1986; Beven, 2000, 2002; Sivapalan, 2003).

"What then remains for the hydrologist to do if we take away from him the curve fitting, model calibration, the chasing of systems responses, correlations, finite elements, kriging, etc.?", as Klemes (1986) asked. Scientists in the first place want to understand nature. This is not to deprecate the relevance of applications for human interventions and predictions with benefit for human society. But science that only provides facts and useful predictions is impoverished; we want to answer the "why" questions (Mayr, 1985; Kleinhans et al., 2005; McDonnell et al., 2007). Application may or may not follow.

From the perspective of other geo- and extraterrestrial sciences and philosophy of science, one tool for exploring and understanding nature is nearly entirely ignored in hydrology: controlled laboratory experimentation (also see Hopp et al., 2009). This paper explores why the potential for novel insights and hypotheses from experiments is tremendous (also see Hacking, 1984, chapters 9 and 13). From the same perspective, there may be more potential for fresh insights from an experimental approach to physics-based modelling

Correspondence to: M. G. Kleinhans
(m.kleinhans@geo.uu.nl)

(explanatory modelling rather than calibrated modelling) than currently acknowledged.

The objective of this paper is to discuss why and how controlled experimentation can lead to new insights and is complementary to fieldwork and modelling. First we discuss basic characteristics, benefits and problems of field data and modelling and, in more detail, experimentation. Limitations, such as due to scaling problems, are extensively discussed and we will argue how these problems are not more grave than basic problems of fieldwork and modelling as discussed extensively in hydrological literature. Then we will illustrate briefly how several other disciplines of the geosciences and planetary science employ observations, modelling and experimentation to interrogate the real world. A brief exploration of the basic logic of scientific explanation is provided as appendix, showing in more depth why experiments are ideally suited to hypothesis generation. Thus we will argue that more explanatory modelling and experimentation in hydrology will lead to better understanding of its major questions.

2 Three pillars of the geosciences

There are basically three ways in which geoscientists interrogate reality (Fig. 1):

- observations recorded in data,
- established laws of nature implemented in models, and
- intervention and manipulation in experiments.

Understanding is gained with all three, but in different ways and with different limitations which will be discussed below with emphasis on experimentation. That a combination of these approaches is more powerful than each alone needs no elaboration. When the results from all three epistemic approaches converge, we may foster hope that we possess good explanations for natural phenomena.

2.1 Observation: overwhelming reality in the field

Field data is as close as possible to reality. It contains variation in time and space that is of interest (pattern or signal) or is the result of processes under study in other sciences but left out here (noise). Direct derivation of understanding from data may be seriously hampered by incompleteness, inaccessibility or bias of the observer in the inference of hypotheses from field observation; i.e. due to the limited frame of reference of a mortal observer.

In general, geoscientific theories and hypotheses based on observations and data, ranging from mechanistical theories to explanatory reconstructions of past conditions, usually are underdetermined by the available data (Kleinhans et al., 2005, also see appendix). That is, there is insufficient available evidence to choose one theory over its rivals or select one set of model parameters over another. The term underdetermination is derived from the Duhem-Quine thesis, which states roughly that a theory is never testable in isolation from other theories and data (see Kleinhans et al., 2005, for discussion on weak and strong underdetermination; weak underdetermination is referred to here).

The underdetermination problems are so pervasive in all geosciences and other sciences that they determine its daily practice to a large extent. Typical examples of underdetermination problems in all geosciences are (Kleinhans et al., 2005):

- Measurement techniques may disturb the observed processes.
- The time scale of human observation is (much) shorter than that of the phenomenon under study.
- Many processes and phenomena cannot (yet) be observed directly or even indirectly. Erosional and sedimentary landforms of the past may have been obliterated by later erosion, and phenomena may not be accessible in practice.
- Many processes are intrinsically chaotic in that they are very sensitive to initial conditions (spatial variation) and boundary conditions (temporal variation) of the system or area under consideration, which are then difficult to specify in enough detail.

An obvious example in hydrology is the practical impossibility to map the considerable heterogeneity of hillslope properties, such as hydrogeology, soils and vegetation, in sufficient detail both within and between watersheds to explain the observed temporal and spatial variability in groundwater flow and surface runoff (Sivapalan, 2003). In fact, so many combinations are possible that every place becomes unique in an arbitrary sense that is nevertheless problematic for satisfactory explanation (Beven, 2000; McDonnell et al., 2007).

To conclude, data commonly support multiple explanatory hypotheses that are empirically equivalent due to the underdetermination problems, but contradict each other in meaning (Chamberlin, 1890). The clearest approach to reduce the number of hypotheses is by triangulation between many different parameters (e.g. Son and Sivapalan, 2007), such as done in geology (Kleinhans et al., 2005). Triangulation means selecting the explanation that fits observations of more than one type, in contrast to selecting an explanation that just fits observations of just one type. In the case of hydrology, that would mean not merely hydrograph fitting. Rather, it would involve simultaneous fitting to the other terms in the water balance such as groundwater dynamics, changes in soil moisture and evaporation, and fitting to other variables such as deuterium composition, which obviously requires data on these variables (Beven, 2000; Son and Sivapalan, 2007).

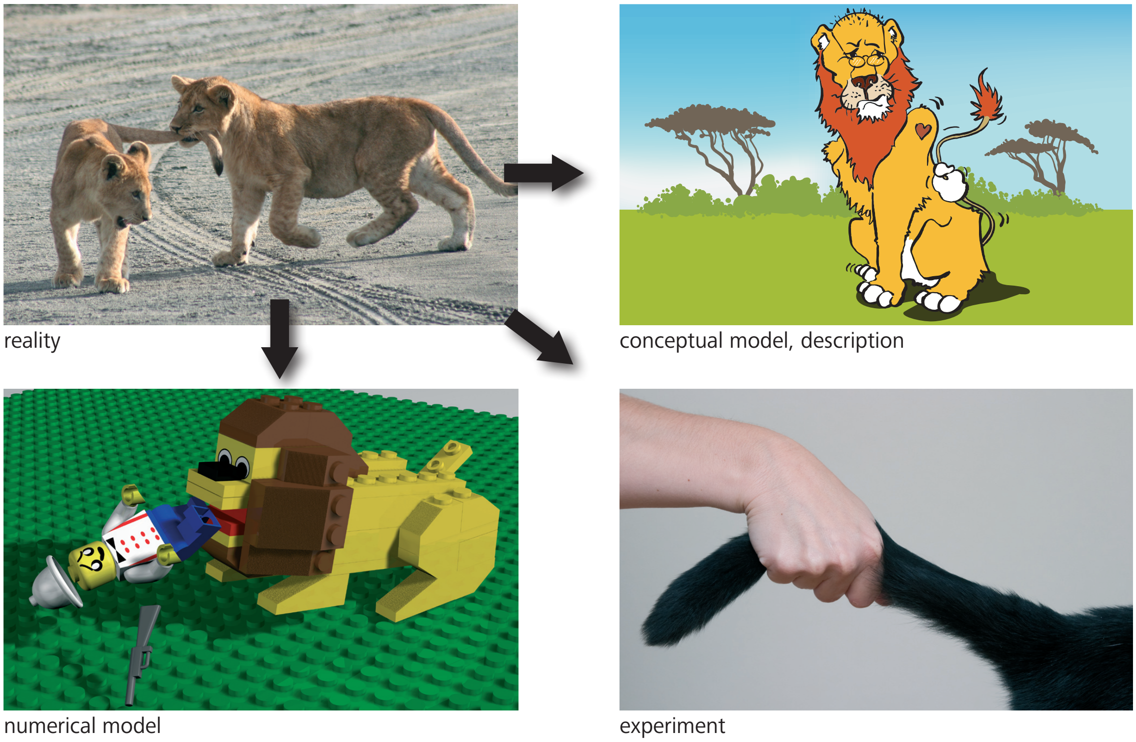

Fig. 1. Three geoscientific ways to interrogate reality. Firstly, description of reality by concepts and data (top right), which may be theory-laden, that is, biased by the describer's frame of reference (the lion is twisting its own tail). The baldness in the cartoon lion refers to the first author, the tattoo to the second and the beard to the third. Secondly, "to twist the lion's tail" and observe what would happen – Lord Bacon's view on doing experimental science – is not commonly possible with large watersheds or the weather system because it is dangerous. Instead, we twist tails of down-scaled representatives of lions: cats (bottom right), which may lead to scale problems. Thirdly, modelling based on established laws (bottom left) is limited in general representativeness of nature by the numerics and the choices of laws, parameters and initial and boundary conditions.

2.2 Models as parsimonious descriptions of reality

2.2.1 Model verification and validation

Numerical models describe reality in terms of mathematical equations, usually at least partly based on laws of natural sciences (see Appendix A). Modelling allows full control over specified boundary conditions and laws. Thus, a physics-based model may be used to test whether a hypothesis does not conflict with the laws of physics.

However, even the laws of physics implemented in models are usually simplified to allow numerical solutions. Furthermore, for many problems it is not obvious which laws apply and to what extent simplification is possible. Such simplifications include model parameters, for instance parameters in macroscopic laws (e.g. Darcy) that represent more expensive and difficult to model microscopic processes (e.g. pore scale flow governed by the Navier-Stokes equations).

Numerical issues, including numerical instability, numerical inaccuracy associated with the numerical integration, and numerical dispersion may also be problematic. For instance, models that solve the Navier-Stokes equations are inherently instable and highly sensitive to minor changes in boundary conditions. Furthermore, the different numerical solutions chosen in different flow routing codes provide different answers despite they are based on the same physics.

Particularly in hydrology, the incorporation of more physical processes in a model leads to inclusion of more parameters. The values of these parameters are usually poorly known, so that models need to be calibrated for each setting. But limited calibration sets allow a wide range of combinations of parameters that give the same results, well-known as equifinality. Furthermore, insensitive parameters are then very poorly constrained, which is known as the problem of parameter identification (van der Perk and Bierkens, 1997). These problems render model predictions inaccurate for different settings (Beven, 2000, 2002).

Thus one cannot be certain whether a mismatch between model results and observations is due to the simplifications and numerical techniques or the underdetermined initial and boundary conditions. As such, extensive physics-based models are not very useful for simulating the details of a concrete existing case (Klemes, 1986; Konikow and Bredehoeft, 1992; Oreskes et al., 1994).

Put more precisely, it is fundamentally impossible to verify and validate models (Oreskes et al., 1994), nor to falsify models. Verification here means to establish their truth content, i.e. the choice of laws to represent the natural phenomena of interest. Validation means establishment of legitimacy, i.e. the model is flawless and internally consistent (also see Klemes, 1986; Konikow and Bredehoeft, 1992). Confirmation or falsification refers to the (dis)agreement between observation and prediction, which merely support the probability of veracity, but do not verify a model. Note that this terminology is closer to the meaning of the words but differs considerably from the way it is commonly used (e.g. Klemes, 1986; van der Perk and Bierkens, 1997).

The world, as geosciences study it, is a mess of abundant intertwined processes, rich history and complexity. There are patterns, but there is also noise and accidents. Scientists have to limit themselves, but by building simplified representations and models of reality necessarily leave out much. The problems of underdetermination partly explain why the construction of a single, generally valid reductionistic model (see Freeze and Harlan, 1969) is fundamentally impossible. In the case of hydrology, much of what is specified in (underdetermined) model parameters, initial and boundary conditions is the result of excluded processes, such as tectonics, landscape evolution, soil formation, climate change, and life. These processes and phenomena have their own disequilibrium dynamics, nonlinearity, thresholds and length scales. These and many other processes and phenomena together formed the Earth and left their imprint and their history, a snap-shot of which is the excruciatingly difficult-to-map of spatial variation. Then there are also accidents, such as landslides and other disasters waiting to happen, the exact course, initiation and timing of which depends on coincidental rainstorms or droughts, earthquakes and so on, which are hard to predict or even hindcast. Small wonder there is uniqueness of place (Beven, 2000).

However, the underlying physics, chemistry and biology of all these phenomena, including the hydrologic, are not unique; merely of varying importance. Models have difficulty fitting the data because the choice of relevant temporal and spatial scales and relevant physics throw out much of the rich history, which then has to be brought back into the model as (underdetermined) initial and boundary conditions. Small wonder that our physics-based models do not fit the data exactly. In geology and biology an explanation for a phenomenon is not complete without reference to both physical factors and history (Mayr, 1985; Kleinhans et al., 2005, and references therein).

2.2.2 The seductive Siren of parsimony

Parsimony is often mentioned as a guide to the question what physics (not) to include. Yet parsimony, also known as Occam's Razor (sometimes Ockham's Razor), is related to one of the most complicated issues in philosophy of science: simplicity. Indeed, scientist's understanding of Occam's razor and simplicity, its usefulness for science and what it actually says about the world or about our possibility to understand it varies greatly (Riesch, 2007). It is quite revealing that scientists pay lip service, or more, to the principle because they think that philosophers of science endorse it, whereas the latter study the principle and its use by scientists because scientists use it.

Loosely put, a parsimonious explanation or model includes just enough elements to explain: not more, but not less either. Overparameterisation leads to underdetermination as is well known (e.g. van der Perk, 1997). Oversimplified models may beg the question, and are sometimes harder to apply because much needs to be specified. For instance, a one-dimensional flow model for meandering rivers needs much more parameterisation than a three-dimensional model because the latter solves the secondary flow pattern whereas the former has to parameterise it. Moreover, using different combinations of laws of physics can result in exactly the same outcome given uncertainty in parameters and input, as frequently occurs in hydrology and morphology. Unfortunately, it is hard to decide then which model is better and more parsimonious without being grounded again on the discussion of statistical measures for goodness-of-fit and uncertainty. In short, Occam's Razor is as helpful for geoscience as a rusty bread-knife for shaving (Kleinhans et al., 2005).

2.2.3 "Experimental" approach to explanatory modelling

Models are very useful to present results of complicated sets of equations that the unaided human mind cannot comprehend. They serve the same purpose for laws as data reduction does for data. In contrast to predictive (site-specific) modelling, explanatory modelling merely attempts to explain the general phenomenon under study rather than predict a case as accurately as possible. Explanatory modelling also covers the downward approach described in Sivapalan et al. (2003), where model complexity is increased step by step until the simplest possible model parsimoniously explains the general trends in the data of a specific case.

Furthermore, models are extremely useful to study sensitivity of results to certain parameters and to explore the viability of hypotheses given certain laws of nature (Oreskes et al., 1994; Kleinhans et al., 2005). Given the chosen laws in the model, it can be studied what result these laws in this model predict if a certain set of initial and boundary conditions were the case, or whether an emergent (statistical) relation exists between initial and boundary conditions and model outcomes, or what initial and boundary conditions are required to yield a certain result. For specific cases, the downward approach can be turned around to do a diagnostic analysis: by adding and calibrating a single process description, the hypothesis that this process explains a certain aspect of the data can be tested (Samuel et al., 2008).

Moreover, models can be used to twist a lion's tail as if they are an experimental facility in which certain effects are included, excluded or modified at will. This second way of explanatory modelling is experimental, without point-by-point comparison to data of a specific case but directed at explaining general trends and creating hypotheses.

For instance, an unexpected phenomenon was identified by combining a reductionistic biophysical model of plant growth and a saturated-unsaturated hillslope hydrology model. The plant growth model demonstrated that carbon assimilation rate was larger under slight water stress than under unstressed conditions. This was caused by reduced evaporation, which led to higher leaf temperature that in turn caused higher carboxylation rates (Brolsma et al., 2010a). When coupled to the hillslope model, a reduced precipitation scenario resulted in a groundwater level rise. This surprising effect was caused by reduced biomass upslope due to the water stress, which reduces interception evaporation, which in turn increased groundwater recharge (Brolsma et al., 2010b). Uncertainties in model parameters, initial and boundary conditions were hardly relevant in these cases because the model was applied to a hypothetical case. Rather, model-derived hypotheses such as these can be used in the analysis of field data and for the setup of new measurements and even for experiments dedicated to testing such a hypothesis.

To summarise: models can be used as tools to mediate between nature and theory based on physics, chemistry and biology (Morgan and Morrison, 1999); i.e. to gain understanding. Particularly explanatory modelling, including diagnostic analysis, hypothesis testing and experimental modelling to generate hypotheses, is useful for such understanding.

2.3 Experiments: controlled, material, and yet serendipitous

2.3.1 Materiality and serendipity

Experimentation by definition allows good control over initial and/or boundary conditions (Hacking, 1984), and involves to some extent the same materials as nature (Morgan, 2003) but with much better accessibility. Experiments also produce serendipitic results.

Materiality is maintained in experiments, contrary to modelling (Morgan, 2003). This is quite important, because the behaviour of the material (water, soil, plants) is unlimited by a simplifying description in terms of laws as in models. One could loosely say that it is the ultimate reductionistic approach, for the material must obey all relevant laws of physics even if we do not yet understand which and how. For instance, all terms in the three-dimensional Navier-Stokes equations are 'retained' in experimental flows; in fact, experimental setups have to be devised specifically to exclude certain effects or dimensions. Due to materiality, experiments may confound us like field observations, whereas models can only surprise us because we can go back to the underlying equations for understanding (Morgan, 2003).

Materiality also allows us to obtain a different sort of knowledge: "a feeling of what happens". This feeling for the behaviour of water and sediment is acquired by fiddling and tinkering with the experimental setup, materials and instruments, and much more so than in the field where unknown variability may overwhelm the pattern. When this embodied knowledge is added to data reduction and to description in terms of mathematically posed laws, it conveys a deeper understanding of what the latter describe and mean. We believe this is what happened, for instance, in the laboratory experiments of Abdul and Gillham (1984) where the relation between groundwater flow and streamflow generation was demonstrated unambiguously for the first time. This combination is also very powerful in training the intuition of students at all levels.

2.3.2 A continuum of fieldwork, experiments and modelling

To clarify what exactly distinguishes experiment from field observation, we provided a strict definition of experiment based on control and materiality. Science in practice has hybrids between field observation and experiment, and between numerical model and experiment. To do justice to this practice while keeping the benefits of experimentation clear, we discuss these hybrids below.

Hybrids of experiments and models represent reality in a variety of ways that are insightful to compare, such as simplification and degree of materiality (Morgan, 2003). Consider experiments and models on flow in pores. A volume of soil can be transferred to the laboratory and subjected to a variety of boundary conditions (de Rooij, 1996). Nondestructive three-dimensional mapping with Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) could be used to construct a digital representation (Lehmann et al., 2006; Kleinhans et al., 2008) for use in a flow model. In this case the experiment and the model are *representative of* soils with pores in reality. Alternatively, the pores could be incorporated in an experiment by artificial structures and sediments, such as glass beads and small porous pipes, and incorporated in a flow model by a network of channels with sizes according to a certain distribution (Lehmann et al., 1998; Joekar Niasar et al., 2009). In this case the experiment and model are *representative for* real soils. Such a variety of setups could test different hypotheses about the effect of pore structure and network on groundwater flow.

Field "experiments" – a terminology often found in hydrology, biochemistry, coastal science and other geosciences – is a contradiction in terms, because barely anything is changed and controlled in nature. The word experiment means nothing in that context as it merely refers to data collection. This is not to deprecate field analyses, which is very valuable. In particular, comparison of similar catchments in similar conditions but one significant difference such as climate or human interference may give a strong signal on the effect of such a difference (e.g. Bosch and Hewlett, 1982; Likens, 2004; Samuel et al., 2008). Yet these field studies are hardly controlled if just one of many variables is controlled.

To clarify, we can evaluate laboratory and field "experiments" along a scale from full to no control. Simple laboratory experiments allow full control over initial and/or boundary conditions (e.g. Black, 1970; Ghodrati et al., 1999). On the other end of the spectrum there is field observation, which allow no control at all. In between these two extremes there are hybrids. One such hybrid is a laboratory experiment on a large block of material from a field site (e.g. Heppell et al., 2000; Holden and Burt, 2002). This allows full control over the boundary conditions but hardly any control over, e.g. initial moisture content or soil characteristics such as the presence of macro-pores. This case is somewhere halfway between observation and experiment (also see Hopp et al., 2009). Yet another hybrid is the field "experiment", which usually allows but one or two controls such as logging (e.g. Bosch and Hewlett, 1982). Such experiments are very close to observation and nearly as far as possible from experiment on the scale of control.

2.3.3 Scale problems in experiments

Of course there are scale problems in experiments, unless the prototype system in nature is so small that it will fit in the laboratory. But this is as much an argument not to experiment as equifinality is an argument not to model. On the contrary, given underdetermination problems in field data and modelling, any addition to our toolbox to explore nature is welcome. Experiments may lead to new understanding if they are allowed to differ quantitatively from a prototype as much as an uncalibrated physics-based model would do. It also matters which aspects are under scrutiny whether scale effects are really problematic.

For instance, it could be argued that vegetation cannot be scaled to the laboratory. Against experiments with vegetation on braided rivers it can be argued that the stems of the plants scale like Sequoia Gigantea to the Rubicon river. In terms of size this may be correct, but size does not matter here. The relevant properties of the plants are their hydraulic resistance as well as the strength their roots provide to the sediment (Tal and Paola, 2007). This added strength can additionally be quantified systematically with geotechnical and other experiments (Kleinhans, 2010b). Surely there are some species among the millions on Earth of which the sprouts under some controlled conditions produce the required vegetation effects for hydrological experiments.

In geomorphology, experiments have yielded many new insights, new empirical relations and new physics-based models ranging on spatial and temporal scales from sand particle interaction and turbulence in milliseconds to landscape (largest catchments) and sedimentary basin formation (entire seas) over millions of years. Typical scale problems, also relevant for potential hydrological experiments, are that a sand particle in the experiment represents a much larger volume of sediment in reality (Postma et al., 2008), or that very thin experimental surface flows have relatively low Reynolds numbers and relatively high Froude numbers and surface tension is important (Peakall et al., 1996) (although no-one really knows how much) compared to nature. Yet these experiments recreated natural phenomena at a small scale and led to new hypotheses that explain natural phenomena.

Likewise for hydrology, it could be argued that it is at presently unknown how to scale from small catchments to large catchments, let alone from microscopic experimental catchments to natural catchments. Yet, whilst the same scale factors applied to the geomorphological examples, and for larger time scales too, this did not stop geomorphologists from doing such experiments with good results. A frequently applied safeguard against severe scale problems is applying basic laws of relevant processes for order-of-magnitude predictions at the experimental scale to help design experiments. For instance, if a certain ratio of subsurface flow and surface runoff is required in the experiment, a good experimental sediment with a certain conductivity and hydraulic roughness can be chosen from calculations with Darcy and Manning. Necessarily, this involves simplifications. In this particular experiment macropores and strong channel turbulence are ignored. It will not do, and it is unnecessary, to recreate a microcosm in which as many processes and details are included as possible. As in explanatory modelling, experiments are simpler than reality.

Scale problems can lead to quantitative bias and even different processes unrepresentative of nature, but apparently these do not fatally preclude the derivation of explanations for natural geomorphological phenomena (Peakall et al., 1996; Paola et al., 2001; Postma et al., 2008). Many phenomena are in fact nearly without scale and it is difficult to distinguish between erosive landscapes and alluvial fans in nature and in the laboratory if no scale is provided (Rodriguez-Iturbe and Rinaldo, 1997; Paola et al., 2001, e.g.). This scaling has in fact been proposed as another alley to explanation of field data (Rodriguez-Iturbe and Rinaldo, 1997; Sivapalan, 2003; McDonnell et al., 2007) and could very well include experiments. In fact, it could be argued that a phenomenon is truly understood when it has been subsumed under the laws of nature *and* has been reproduced experimentally.

In summary, experiments have great potential to gain understanding in hydrology. The problems of scaling down to laboratory size are no reason not to try, as such problems have not impeded progress in geomorphology on similar spatial scales and longer time scales. Furthermore physics-based modelling on both the real and laboratory scale could mediate between reality and experiment, as can hybrids between models, experiments and fieldwork.

3 Experimentation in hydrology?

To be sure, hydrology has its experiments, albeit few. In this chapter we provide examples, suggest explanations for the scarcity of experimentation in hydrology, and elaborate on the opportunities that experimentation would provide for some grand challenges for hydrology.

3.1 Experiments in hydrology

Here we give a number of examples are given from high to low control on parameters and boundary conditions. Note that most of these appeared in soil and geomorphology journals or mixed journals such as Water Resources Research, but not in specialist hydrological journals.

Salehin et al. (2004) studied the basic effects of sediment structure on hyporheic exchange with only eight experiments on two experimental heterogeneous sediment beds in a very small laboratory flume. They found that solute penetration was confined to a shallower region and led to faster near-surface transport compared to homogeneous beds. This was confirmed with modelling applied to heterogeneous and homogeneous beds. The artificially emplaced sand layers could easily be made more naturally in larger flumes by having them formed by the flow and this is fruitful to explore. But the point is that this small but careful experimentation revealed an explanation for a natural phenomenon by focussing on the contrast between homogeneous and heterogeneous beds. A similar approach was taken by Ghodrati et al. (1999) with a well-controlled artificial sediment and a mechanically produced macropore of varying size to characterise macropore flow and interaction with matrix flow in soils.

Black (1970) created laboratory catchment models of various idealised morphologies measuring less than half a meter in size. The models were constructed of styrofoam with a 1 cm soil blanket of polyurethane sponge. He argued that laboratory models of watersheds can be used to study the effect of various watershed parameters on runoff behaviour, especially on peak flows and timing, indicating that watershed similitude can be attained. The control on shape, uniformity of material and the boundary condition (rainfall simulation) was very large in these experiments.

Heppell et al. (2000) studied material response and solute leaching of an artificial soil under laboratory rainfall. They found a trend of less solute release under intense rainfall compared to less intense rainfall which was explained by a transport non-equilibrium effect. This study was done with a clay from the topsoil of a catchment of interest in order to obtain good control over the material properties, including composition, aggregate size distribution and macropore distribution. A similar approach was taken by Holden and Burt (2002) on a peat sample to assess hydrological effects of global warming on upland blanket peats.

A vertical two-dimensional sand slope in the laboratory was used to identify the significant effect of groundwater flow on streamflow generation (Abdul and Gillham, 1984). Despite the simple setup and use of uniform sand, basic phenomena were identified that remain significant. In particular, if the groundwater table is near the surface then a change in head leads to a large groundwater contribution to streamflow. In other cases, the change in head may result in overland flow. Which mechanism prevails depends primarily on precipitation intensity, surface slope and hydraulic conductivity.

Michaelides and Wainwright (2008) compared hillslope-channel flow coupling in a Froude-scaled experiment to a numerical model applied at the experiment scale. Hillslope angle, channel angle, hillslope discharge and channel discharge were systematically varied in a setup with side-slopes to a straight channel. Effects of hillslope flow on the channel routing were identified by using more than one statistical measure for model performance, which is significant because concurrent changes in variables in the channel and on the hillslope propagate errors in the model resulting from process representation and/or model structure. In short, not only obvious parameters such as channel discharge and gradient determine flow velocity and depth, but also hillslope gradient and discharge. A process relevant in nature was identified and understood by comparing results between experiments and between some experiments and a model.

A large hillslope experiment is being designed in a Biosphere 2 dome (Hopp et al., 2009), where the temporal and spatial scale of the facility allows for interactions between vegetation, soil, water chemistry, subsurface and overland flow. The experimental scale (30×15 m) is large enough to allow natural evolution of spatial patterns and variability of these properties. Hopp et al. (2009) discuss the intricacies and limitations of the design of this large-scale experiment. Following the detailed study and modelling of this microcosm, hypotheses are expected that certain phenomena in the real world may have the same explanations as in the experiment.

Data of a well-instrumented artificial grassland catchment of 490 m^2 was used by Kendall et al. (2001) to examine hydrograph separation from hydrometric, isotopic and geochemical approaches. The “Hydrohill” catchment was created by excavating a natural catchment down to bedrock, surrounded by impermeable concrete and refilled to restore he natural soil profile. Even in this “highly controlled artificial catchment”, as the authors call it, the boundary conditions and spatial heterogeneity of the parameters were not known in enough detail. Consequently the data underdetermined the distinction of unique subsurface and surface flow paths.

These examples illustrate how the three pillars of the geosciences, fieldwork, experimentation and modelling, are complementary and can be combined in various ways (see Kleinhans, 2010b, and engineering literature for examples where actual field cases and experiments are compared and see hydrological literature for combinations of field data and modelling).

3.2 Reasons for scarcity of hydrological experiments

We suggest three possible reasons why experimentation is relatively rare in hydrology:

- Historically there is an emphasis on good prediction for direct use by the society. In addition to statistical methods based on observation, modelling is the logical tool for operational prediction, so it is not surprising that there has been a lot of effort in hydrological modelling, particularly highly detailed reductionistic modelling and data-model integration.
- There may be a cultural aspect to the emphasis on observation and modelling in hydrology. The scarcity of experiments in literature and in presentations at conferences does perhaps not stimulate the use of experiments to address the grand challenges of hydrology.
- There is a mismatch between the time scale of interest and the spatial scale of interest in hydrological prediction in large basins. In general it takes more time to erode an entire mountain range than it takes to carve a minor gully. In catchment hydrology short-term changes (floods) are of interest over large spatial scales. Spatial variation in parameters (called initial conditions in other contexts) such as permeability etc. is caused by long-term landscape evolution and are therefore not coevolved in the model.

To elaborate on the last point: for the large-scale phenomenon the short-duration changes in forcing are irrelevant whilst an average forcing produces the phenomenon well. This means that, contrary to geomorphology, much more must be specified (from observations) for good predictions of the phenomena of interest. In agreement with this suggestion, experiments aimed at understanding entire catchments or aquifers seem to be much more rare than small-scale laboratory experiments on elements of the system or particular pore-scale processes such as wetting phenomena. However, landscape evolution takes place over much longer time scales than most hydrological events of interest. Experiments could show how hydrology, soil development, vegetation evolution and morphology change when a different is forced on it. Realistic experimental fluvial or erosional landscapes have been created using a highly simplified hydrological regime consisting of one endless flood or a single-size flood interrupted by a single-size low-flow period. In fact, the formative discharge is a geomorphologically meaningful concept. It would be very interesting to force a more realistic upstream discharge or precipitation pattern on such an experiment to study how response, particularly the hydrological response, differs from that in the constant forcing experiments. The next logical step is to apply changing boundary conditions. Such experimentation could show how important heterogeneity is for the long-term hydrodynamics and how that affects long-term landscape dynamics.

3.3 Growing opportunities for experimentation in hydrology

At present and in the near future, there is a growing demand for longer-term predictions in view of global change issues and growing anthropogenic pressure on the environment. The initial and boundary conditions cannot be measured in sufficient detail for accurate long-term prediction of effects of changing climate, vegetation and so on. Part of the reason is that boundary conditions and properties described by parameters are in fact co-evolving phenomena such as soil and vegetation. So, to understand the dynamics of large basins, it is necessary to understand how the landscape and its structure and heterogeneity came about. For such longer-term predictions the feedbacks between hydrology, vegetation, soil, morphology and so on become more important, so there is a move towards more interdisciplinary approaches. Therefore the focus is also shifting from operational prediction to understanding.

For explanatory modelling, of which results cannot directly be compared to data for verification, the question then arises whether particular outcomes are model artifacts or real phenomena, which can be tested by creative and controlled experimentation. Furthermore ungauged basins receive more attention, for which the detailed initial and boundary conditions required for detailed reductionistic modelling may not be available. These are not grand challenges of hydrology alone, but of all earth system sciences. To address these grand challenges, controlled laboratory experiments and explanatory modelling will become more important tools.

4 Telescopic comparison with some other geosciences

Many different geosciences tend to focus mostly on one or two of the three pillars of the earth sciences. Hydrology needs no introduction to this readership, and the underemployed potential of experimentation has been remarked upon. By comparing some geosciences to hydrology we will illustrate why and how we can fruitfully combine fieldwork, experimentation and modelling, and where ignoring one of the pillars is damaging to science. A study on the reasons for differences between the approaches of geosciences – is it the nature of the subject? Is there a socio-historical reason? Is the practice of these approaches so different that they diverged into different disciplines? – is much beyond the scope of this paper.

4.1 Geomorphology

Geomorphology has combined field data analysis and experiments since before the term was coined. To be fair, the emphasis has always been on fieldwork while experiments usually were small-scale. Modelling was introduced with the advent of computers as in all natural sciences, but remained at a relatively low level of complexity compared to the models of meteorology and hydrology. Large-scale Froude-scaled experiments in fluvial engineering have largely been replaced by numerical modelling for cost reasons and because of high faith in models outside the modelling community. Within the community there are model comparisons, not to determine which is the best, but to learn about nature from differences between the models and their outcomes (e.g. Davies et al., 2002, in coastal science, e.g. Nearing et al., 2005, in soil erosion).

The relation between fieldwork, experiment and modelling in geomorphology is not an easy one. Although geomorphology is identifiable as a discipline by a number of conferences and journals, it lies on the overlap between more field-oriented geology (Quaternary geology, sedimentology) and more model-oriented sciences (civil engineering, geophysics). Apart from the contrasts in quantitative and qualitative approaches there is also the usual misunderstanding between their fundamentally different questions: "what was the cause in the past" of the geologists versus "what were the laws involved; how does it work" of the process-oriented sedimentologists and morphologists (Baker, 1996).

There is much scope for fast progress in geomorphology by combination of experiments in large-scale facilities as well as more sophisticated modelling on supercomputers as is common in meteorology (see Kleinhans, 2010b, for review on fluvial morphology).

4.2 Meteorology

Meteorology has a long tradition of observation, and has probably the largest and most diverse data collection system on Earth based on a high density network of sensors as well as remote sensing.

Several sophisticated models are run continuously on supercomputers for ensemble forecasting, while their initial and boundary conditions are continuously updated by data. The models have not and could not have been developed by individuals; they are community models that are often compared against each other to learn from the differences. Also climate modelling proceeds in this manner, and explanatory modelling as well as scenario modelling is a key activity. The obvious societal relevance is one reason why this science has much more resources than other geosciences.

Weather prediction by models did not improve gradually over the past decades. Some improvements of the models, such as the replacement of parameterisations by more physics, actually led to deterioration of the predictive capacity. This apparent paradox led to the search for other model elements that needed improvement, and in the longer term the models improved considerably (G. Komen, KNMI, personal communication, 2004). An experimental approach to modelling is common; model improvement was not obtained by extensive calculations of model uncertainty but by twisting the model's tail to "experimentally" determine whether a certain model element is responsible for a certain phenomenon.

Experiments are less common in meteorology and related sciences. One reason perhaps is that the intrinsic scales of weather are large (shower, cyclone) and hard to scale back to the laboratory. Another reason is that theory derived from experiments is borrowed from other sciences, for instance in wind tunnel experiments some boundary layer descriptions and turbulence closures are directly based on fluid dynamics. The same can be said of oceanography and glaciology, which collaborate with meteorology in climatology. The border between such sciences is obviously arbitrary but results from the disciplinary boundaries defined by tradition. So, if we forget these for a moment, then meteorology does indeed use experiments for the study of fundamental processes in controlled conditions. Furthermore, we note that meteorology is interdisciplinary, as is hydrology (Klemes, 1986).

Future progress at the border between hydrology and climatology can perhaps be made in combining experimental work on microclimate, vegetation and landuse with the large-scale forcing by vegetation of precipitation and evapotranspiration forcing (Pielke Sr., 2008).

4.3 Planetary science: comparing Mars and Earth

Planet Mars has been studied mostly by photogeologic interpretation. This is related to the fact that most questions on Mars are about a distant past, for instance how the planet developed tectonically, or how wet and warm the climate was billions of years ago. These questions are most directly addressed with interpretation of the surface features (Kleinhans, 2010a). Spectral remote sensing and ground-penetrating radar led to the first geochemical analyses and subsoil mapping in the past decade. Both terrestrial analogues and concepts are borrowed from geology, geomorphology and so on, but are sometimes over-interpreted and unconfined by available data and physics-based modelling.

Physics-based modelling is rarely applied except for impact cratering; tectonic modelling has been applied as well as global aquifer modelling, climate modelling and some landscape evolution modelling. All these model exercises were explanatory, in part because most of the required input data are unavailable and in part because the focus is on major questions rather than detailed quantitative hindcasts.

Several studies on Mars helped to understand Earth better. For instance, tremendous floods left clear marks on Mars and helped to interpret less clear marks on Earth of the Missoula Flood events and similar events in Siberia at the end of the last glacial (Baker and Milton, 1974).

Experiments applied to Mars are rare: near-surface atmospheric conditions and soil properties have been simulated to assess possibilities for the emergence and survival of amino acids (with implications for primitive life forms) (Ten Kate et al., 2005), box canyon were formed experimentally by groundwater sapping (Howard et al., 1988) and deltas formed in impact crater lakes have been recreated experimentally at a small scale (Kraal et al., 2008).

There is considerable scope for rapid progress in planetary science by modelling and experimentation borrowed from other sciences, including hydrology (see Kleinhans, 2010a, for review). For hypotheses and ideas established disciplinary boundaries are already fruitfully ignored.

5 Conclusions: a bright future for hydrology

Based on a comparison between different geosciences we submit that much progress can be made where one of the pillars of the geosciences (fieldwork, modelling and experimentation) was less frequently employed. Hence, hydrology (and teaching in hydrology) would benefit from more experimental work:

– Experimentation leads to novel ideas and hypotheses for major questions of hydrology, mostly through abduction.

– Combinations and hybrids of experiments, models and field observation allow compensation for drawbacks of each approach.

– Feeling and manipulating the material in experiments adds insight to observation and modelling like playing music adds insight to listening and studying mathematics of music.

Experimentation is an art that needs as much work as modelling or fieldwork to master (like learning to play a musical instrument). Furthermore, investment is required: first in small and cheap experimental facilities for fast exploration and hopefully at some point in large facilities to overcome (or prove negligible) certain scale effects and allow larger systems and more detailed measurements. These practical problems are surmountable through collaboration with the experimenters of, e.g. civil engineering and geomorphology in existing facilities.

It is well known that a simpler calibrated model may be more accurate and much cheaper to deduce predictions for a unique location than a complicated reductionistic model given computational cost, required level of detail of initial and boundary conditions and the fundamental problem of model verification and validation. However, reductionistic models are extremely useful for gaining understanding and testing viability of hypotheses – in short, by "experimentation". An experimental attitude to modelling complements the proposal by Sivapalan (2009) for pooling of data on large watersheds. Physics-based models can be applied explanatorily to theoretical and more practical cases to learn what general trends can be found from the incorporated laws of physics. Physical, chemical and biological processes can be introduced step by step to assess their effect on the trend.

There is much scope for experiment design and explanatory modelling that do not focus solely on hydrology but also on the coeval morphology, soil, ecology, microclimate and so on. The landscape evolution then includes the spatiotemporal variation and its effects on the hydrology. Also the results of landscape models, or ecological models, or network, pore and connectivity models could be used as input for hydrological models to compare general hydrological trends for contrasting inputs and compare these to analytical solutions and experimental results. The aim of such exercises is not to fit the hydrograph, so to speak, but to infer and test hypotheses and gain understanding of major hydrologic issues. Feedbacks between different domains such as hydrology, meteorology, soil, morphology, vegetation and so on are the grand challenges for many earth science disciplines including the discipline hydrology. This is especially valuable for understanding and prediction in ungauged basins. Handling water and dirt, even in small experiments, will enhance this understanding considerably.

Appendix A

How we explain: geo-logic and hypothesis conception

In the main text the reasoning why hydrology could benefit from experiments proceeded by analogy with other geosciences and qualitative arguments such as materiality in experiments. In the following section we explore in more depth how earth scientists arrive at hypotheses and explanations through three elementary pieces of logic, and why and how experimentation plays such an important role in the generation of hypotheses (Fig. A1). All three forms of logic are employed in various combined ways in fieldwork, modelling and experimentation.

A1 Deduction, induction, *and abduction*

When scientists are asked how their science works, they commonly and rightly refer to induction and deduction. Statistical generalisations such as Hack's Law were obtained by induction, whereas physics-based prediction is a typical deductive exercise.

But the practice of science hardly involves following the recipes of deduction and induction. In fact, there is not agreement yet among philosophers of science as to what amounts to an explanation and what exactly is understanding (see, e.g. Lipton, 1991; de Regt et al., 2009). Three applicable types of logical reasoning are based on causes, effects and laws, two of which are necessary to arrive at the third (Fig. A1). With deduction and induction, the third possibility is *abduction*. They are not merely alternatives but answer different types of questions. Neither are different sciences limited to one of them, but all three are employed in all sciences, including hydrology (Kleinhans et al., 2010).

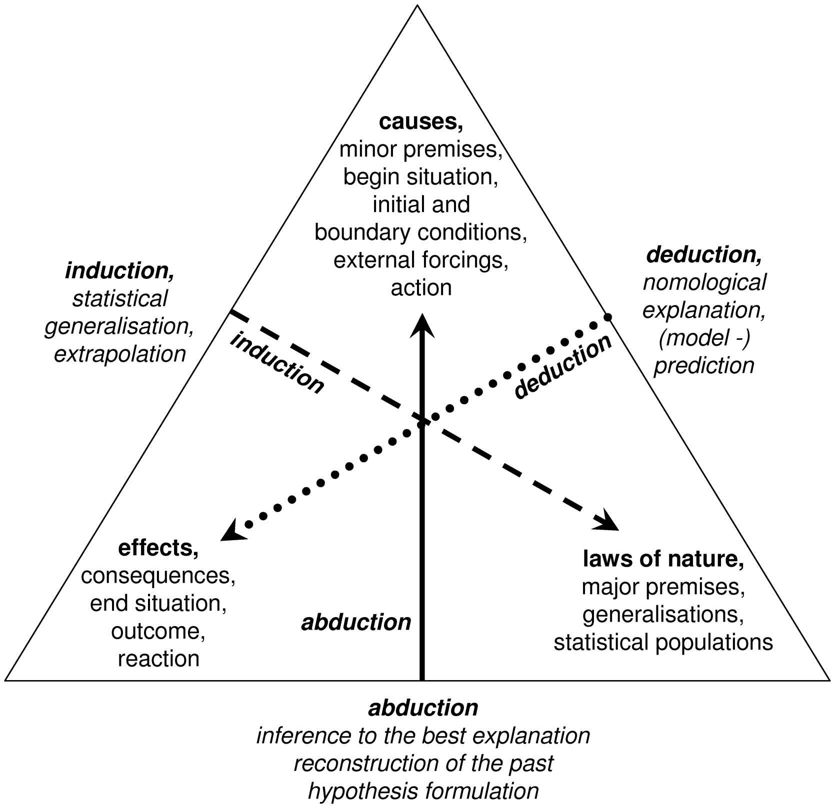

Fig. A1. The relation between abduction, deduction and induction. Several alternative terms encountered in literature are given. Each has its own weakness (see text) (from Kleinhans et al., 2010).

New hypotheses are conceived through abduction (this is not an entirely complete and correct account but sufficient for now; for an authoritative account see Lipton, 1991). The term abduction was coined by the american philosopher C. S. Peirce more than a century ago but has surprisingly remained unknown by most scientists (except Baker, 1996). Some are clearly ill at ease with it or ill at ease with the fact that something is missing from our vocabulary to describe what we do as scientist; for instance Savenije (2009) called it the "art" of science. We will explain that it is a form of logic in its own right and we will argue that experiments are ideally suited for abduction, which is why experiments may yield many new hypotheses about the world.

A2 Deduction

For deduction, the initial conditions (causes) are combined with laws of nature to explain or predict the effects (Fig. A1). This is what happens in analytical solutions for linear stability analyses and physics- or chemistry-based modelling to solve boundary value problems (e.g. Freeze and Harlan, 1969, in hydrology). For specific sites it has obvious relevance such as flood forecasting.

Deduction is a solid form of logic compared to induction and abduction. Its Achilles heel is in the choice of relevant laws and the common use of generalisations rather than laws, and the initial and boundary conditions which must be based on measurements that may be incomplete or contain errors (Oreskes et al., 1994) (also see Sect. 2.2).

A3 Induction

Induction leads to (statistical) generalisations based on both causes and effects (Fig. A1). Interpolation and extrapolation are induction too (see examples in Klemes, 1986). Induction yielded useful generalisations in the geosciences, such as Hack's "law", and also in hydrology such as an empirical predictor for vegetal effects on water yield and evapotranspiration, identified from 94 catchments in Bosch and Hewlett (1982).

The problems of induction are well known: the validity range of empirical relations is determined by the range and bias of the data included, and the amount of data is obviously never large enough to create a universally valid generalisation, that is, law. Nevertheless empirical relations somehow contain information about reality and have shown the way to underlying mechanisms in the past, but not infallibly so.

A4 Abduction

In abductive inference, final conditions, facts and so on are (often implicitly) combined with laws or generalisations of nature, to arrive at the best of a limited number of hypotheses that explain the observations. It starts with a surprising observation, followed by an insight how the phenomenon may have come about. Thus abduction leads to hypotheses, including hypotheses about conditions or events in the past that led to the present phenomenon under observation. Such hypotheses can then be tested by modelling as in the downward approach, in diagnostic analysis, in explanatory modelling and in an experimental approach to modelling. The hypotheses can also be tested in experiments, including generalised simplified setups and scaled experiments dedicated to a unique place.

The major limitation of abduction is that one cannot be certain that all possible hypotheses, including the correct one, have been conceived. The right hypothesis might be one that no-one thought of. For example, several clues from geological investigations, combined with a law that iridium must come from outside the Earth, led to the hypothesis that dinosaurs became extinct after a comet impacted. This is an example of abduction, which earth scientists and also detectives commonly employ (Baker, 1996; Kleinhans et al., 2010). Thus abduction led to one process-oriented narrative of what happened, but alternative hypotheses have also been formulated and the jury is still out. The inference of a perceptual model, a perspective view of the watershed's functioning (Sivapalan, 2003, 2009), is also an example of abduction from end results (observations) to relevant laws, generalisations and boundary conditions.

Abduction is the most interesting for geoscientists even though it is perhaps the most fallible of the three, precisely because it leads to new ideas and hypotheses. In some postwar philosophy of science the practice of science was divided between the context of justification or falsification of theories, that is, the science, and the context of discovery, that is, the magic. The former was considered open to philosophical and logical analysis, whereas the latter was considered the realm of psychologists.

The account of abductive inference (C.S. Peirce and later work, see Lipton, 1991; Baker, 1996; Kleinhans et al., 2010) puts the conception of ideas back into science where it belongs. New ideas are often inspired by observation in the field, by experimental playing with models and, importantly, by experimentation. In the cases of modelling or experimentation, abduction follows: a result may resemble a natural phenomenon of interest, from which it can then be abduced that the natural phenomenon is perhaps also explained by the same mechanism and initial and boundary conditions as in the model or experiment.

At the lower level of the experiment itself, abduction also plays a major role. Experiments may yield many surprising results. Given the mechanisms in play and the results, the causes must then be abduced, so that hypotheses are generated how the surprising result came about. Contrary to field observations, the initial and boundary conditions are relatively well known, which makes the abduction easier and less speculative. Contrary to models, there is materiality in experiments, which enhances understanding, the likelihood of surprises and the relation to reality. Hence, if a hypothesis is derived from experiments, it is not unlikely that the same hypothesis applies to a real-world situation. Thus experiments are instrumental in the generation of hypotheses and understanding.

Acknowledgements. The ultimate cause of this paper is the stimulating discussion with Murugesu Sivapalan. Earlier discussions with Thom Bogaard and Victor Baker were appreciated. Erwin Zehe, Majid Hassanizadeh and two anonymous reviewers helped to improve the paper. Fred Trappenburg of Geomedia, Utrecht University, painted and virtually modelled two representative lions and seduced a live scale lion to have its tail twisted (gently) on camera; Kimberlee Kessler Design granted permission for the use of the real lion photograph.

Edited by: H. H. G. Savenije

References

Abdul, A. and Gillham, R.: Laboratory studies on the effects of the capillary fringe on streamflow generation, Water Resour. Res., 20, 691–698, 1984.

Baker, V.: The pragmatic roots of American quaternary geology and geomorphology, Geomorphology, 16, 197–215, 1996.

Baker, V. and Milton, D.: Erosion by catastrophic floods on Mars and Earth, Icarus, 23, 27–41, 1974.

Beven, K. J.: Uniqueness of place and process representations in hydrological modelling, Hydrol. Earth Syst. Sci., 4, 203–213, 2000, http://www.hydrol-earth-syst-sci.net/4/203/2000/.

Beven, K.: Towards a coherent philosophy for modelling the environment, P. R. Soc. Lond. A-Conta, 458, 1–20, 2002.

Black, P.: Runoff from watershed models, Water Resour. Res., 6, 465–477, 1970.

Bosch, J. and Hewlett, J.: A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration, J. Hydrol., 55, 3–23, 1982.

Brolsma, R., Karssenberg, D., and Bierkens, M.: Vegetation competition model for water and light competition I: model description, 1-dimensional competition and the influence of groundwater, Ecol. Model., accepted, 2010a.

Brolsma, R., van Vliet, M., and Bierkens, M.: Climate change impact on a groundwater-controlled ecohydrological hillslope system, Water Resourc. Res., in review, 2010b.

Chamberlin, T.: The method of multiple working hypotheses, Science, 15, 92–96, 1890.

Davies, A. G., Rijn, L. C. V., Damgaard, J. S., de Graaff, J. V., and Ribberink, J. S.: Intercomparison of research and practical sand transport models, Coast Eng., 46, 1–23, 2002.

de Regt, H., Leonelli, S., and Eigner, K.: Understanding, geen idee, University of Pittsburgh, Pittsburgh, USA, chapter 1, 1–17, 2009.

de Rooij, G.: Preferential flow in water-repellent sandy soils, Wageningen University, PhD thesis, 1996.

Freeze, R. and Harlan, R.: Blueprint for a physically-based, digitally simulated hydrologic response model, J. Hydrol, 9, 237–258, 1969.

Ghodrati, M., Chendorain, M., and Chang, Y.: Characterization of macropore flow mechanisms in soil by means of a split macropore column, Soil Sci. Soc. Am. J., 63, 1093–1101, 1999.

Hacking, I.: Representing and intervening, Cambridge University Press, Cambridge, UK, 1984.

Heppell, C., Burt, T., Williams, R., and Johnson, A.: A laboratory investigation of the release of a conservative tracer and herbicide from topsoil aggregates under varying rainfall intensities, Soil Use Manage., 16, 175–182, 2000.

Holden, J. and Burt, T.: Laboratory experiments on drought and runoff in blanket peat, Eur. J. Soil Sci., 53, 675–689, 2002.

Hopp, L., Harman, C., Desilets, S. L. E., Graham, C. B., McDonnell, J. J., and Troch, P. A.: Hillslope hydrology under glass: confronting fundamental questions of soil-water-biota co-evolution at Biosphere 2, Hydrol. Earth Syst. Sci., 13, 2105–2118, 2009, http://www.hydrol-earth-syst-sci.net/13/2105/2009/.

Howard, A., Kochel, R., and Holt, H.: Sapping features of the Colorado Plateau, no. SP-491 in NASA Special Publications, NASA, Scientific and Technical Information Division, Washington DC, USA, 1988.

Joekar Niasar, V., Hassanizadeh, S., Pyrak-Nolte, L., and Berentsen, C.: Simulating drainage and imbibition experiments in a high-porosity micromodel using an unstructured pore network model, Water Resour. Res., 45, W02430, doi:10.1029/2007WR006641, 2009.

Kendall, C., McDonnell, J., and Gu, W.: A look inside "black box" hydrograph separation models: a study at the Hydrohill catchment, Hydrol. Process., 15, 1877–1902, 2001.

Kleinhans, M. G.: A tale of two planets: Geomorphology applied to Mars surface, fluvio-deltaic processes and landforms, Earth Surf. Proc. Land., 35, 102–117, doi:10.1002/esp.1895, 2010a.

Kleinhans, M. G.: Sorting out river channel patterns, Prog. Phys. Geogr., accepted, 2010b.

Kleinhans, M. G., Buskes, C., and de Regt, H.: Terra Incognita: Explanation and Reduction in Earth Science, Int. Studies in the Philosophy of Science, 19, 289–317, doi:10.1080/02698590500462356, 2005.

Kleinhans, M. G., Jeukens, C., Bakker, C., and Frings, R.: Magnetic Resonance Imaging of coarse sediment, Sediment. Geol., 208, 69–78, doi:j.sedgeo.2008.07.002, 2008.

Kleinhans, M. G., Buskes, C., and de Regt, H.: Philosophy of Earth Science, in: Philosophies of the sciences, edited by: Allhoff, F., Wiley-Blackwell, Chichester, UK, chapter 9, 213–236, 2010.

Klemes, V.: Dilettantism in hydrology: transition or destiny?, Water Resour. Res., 22, 177S–188S, 1986.

Konikow, L. and Bredehoeft, J.: Ground-water models cannot be validated, Adv. Water Resour., 15, 75–83, 1992.

Kraal, E., van Dijk, M., Postma, G., and Kleinhans, M. G.: Martian stepped-delta formation by rapid water release, Nature, 451, 973–976, doi:10.1038/nature06615, 2008.

Lehmann, P., Stauffer, F., Hinz, C., Dury, O., and Flühler, H.: Effect of hysteresis on water flow in a sand column with a fluctuating capillary fringe, J. Contam. Hydrol., 33, 81–100, 1998.

Lehmann, P., Wyss, P., Flisch, A., Lehmann, E., Vontobel, P., Krafczyk, M., Kaestner, A., Beckmann, F., Gygi, A., and Flühler, H.: Tomographical imaging and mathematical description of porous media used for the prediction of fluid distribution, Vadose Zone J., 5, 80–97, doi:10.2136/vzj2004.0177, 2006.

Likens, G.: Some perspectives on long-term biochemical research from the Hubbard Brook ecosystem study, Ecology, 85, 2355–2362, 2004.

Lipton, P.: Inference to the best explanation, Routledge, Oxon, UK, 219 pp., 1991.

Mayr, E.: How biology differs from the physical sciences, MIT-press, Cambridge, Massachusets, 43–63, 1985.

McDonnell, J., Sivapalan, M., Vaché, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M., Selker, J., and Weiler, M.: Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology, Water Resour. Res., 43, W07301, doi:10.1029/2006WR005467, 2007.

Michaelides, K. and Wainwright, J.: Internal testing of a numerical model of hillslopechannel coupling using laboratory flume experiments, Hydrol. Process., 22, 2274–2291, doi:10.1002/hyp.6823, 2008.

Morgan, M.: Experiments without material intervention: model experiments, virtual experiments, and virtually experiments, in: The philosophy of scientific experimentation, edited by: Radder, H., University of Pittsburgh Press, Pittsburgh, USA, chapter 11, 216–235, 2003.

Morgan, M. and Morrison, M.: Models as mediators; perspectives on natural and social science, Cambridge University Press, Cambridge, UK, chapter 2, 10–37, 1999.

Nearing, M., Jetten, V., Baffaut, C., Cerdan, O., Couturier, A., Hernandez, M., Le Bissonnais, Y., Nichols, M., Nunes, J., Renschler, C., Souchère, V., and van Oost, K.: Modeling response of soil erosion and runoff to changes in precipitation and cover, Catena, 61, 131–154, doi:10.1016/j.catena.2005.03.007, 2005.

Oreskes, N., Shrader-Frechette, K., and Belitz, K.: Verification, validation and confirmation of numerical models in the earth sciences, Science, 263, 641–642, 1994.

Paola, C., Mullin, J., Ellis, C., Mohrig, D., Swenson, J., Parker, G., Hickson, T., Heller, P., Pratson, L., Syvitski, J., Sheets, B., and Strong, N.: Experimental Stratigraphy, GSA Today, 11, 4–9, 2001.

Peakall, J., Ashworth, P., and Best, J.: Physical modelling in fluvial geomorphology: principles, applications and unresolved issues, in: The scientific nature of geomorphology, edited by: Rhoads, B. and Thorn, C., Wiley, Chichester, UK, 222–253, 1996.

Pielke Sr., R.: A broader view of the role of humans in the climate system, Phys. Today, 11, 54–55, 2008.

Postma, G., Kleinhans, M. G., Meijer, P., and Eggenhuisen, J.: Sediment transport in analogue flume models compared with real-world sedimentary systems: a new look at scaling evolution of sedimentary systems in a flume, Sedimentology, 55(6), 1–17, doi:10.1111/j.1365-3091.2008.00956.x, 2008.

Riesch, H.: Simple or simplistic? Scientists views on Occam's razor, in: Proc. of the First Biennial Conf. of the Soc. for Philosophy of Science in Practice, edited by: Boon, M. and Waelbers, K., University of Twente, The Netherlands, 2007.

Rodriguez-Iturbe, I. and Rinaldo, A.: Fractal river basins. Chance and self-organisation, Cambridge University Press, New York, 1997.

Salehin, M., Packman, A., and Paradis, M.: Hyporheic exchange with heterogeneous streambeds: Laboratory experiments and modeling, Water Resour. Res., 40, W11504, doi:10.1029/2003WR002567, 2004.

Samuel, J., Sivapalan, M., and Struthers, I.: Diagnostic analysis of water balance variability: A comparative modeling study of catchments in Perth, Newcastle, and Darwin, Australia, Water Res. Resour., 44, W06403,, doi:10.1029/2007WR006694, 2008.

Savenije, H. H. G.: *HESS Opinions* "The art of hydrology"*, Hydrol. Earth Syst. Sci., 13, 157–161, 2009, http://www.hydrol-earth-syst-sci.net/13/157/2009/.

Sivapalan, M.: Process complexity at hillslope scale, process simplicity at the watershed scale: is there a connection?, Hydrol. Process., 17, 1037–1041, doi:10.1002/hyp.5109, 2003.

Sivapalan, M.: The secret to "doing better hydrological science": change the question!, Hydrol. Process., 23, 1391–1396, doi:10.1002/hyp.7242, 2009.

Sivapalan, M., Bloeschl, G., Zhang, L., and Vertessy, R.: Downward approach to hydrological prediction, Hydrol. Process., 17, 2101–2111, doi:10.1002/hyp.1425, 2003.

Son, K. and Sivapalan, M.: Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data, Water Resour. Res., 43, W01415, doi:10.1029/2006WR005032, 2007.

Tal, M. and Paola, C.: Dynamic single-thread channels maintained by the interaction of flow and vegetation, Geology, 35, 347–350, doi:10.1130/G23260A.1, 2007.

Ten Kate, I. L., Garry, J. R. C., Peeters, Z., Quinn, R., Foing, B., and Ehrenfreund, P.: Amino acid photostability on the Martian surface, Meteorit. Planet. Sci., 40, 1185–1193, 2005.

van der Perk, M.: Effect of model structure on the accuracy and uncertainty of results from water quality models, Hydrol. Process., 11, 227–239, 1997.

van der Perk, M. and Bierkens, M.: The identifiability of parameters in a water quality model of the Biebrza River, Poland, J. Hydrol., 200, 307–322, 1997.