Bokulich, A. and N. Oreskes (2017), "Models in Geosciences"
in L. Magnani and T. Bertolotti (eds.)
Springer Handbook of Model-Based Science
Springer: 891-911.

41. Models in Geosciences

Alisa Bokulich, Naomi Oreskes

The geosciences include a wide spectrum of disciplines ranging from paleontology to climate science, and involve studies of a vast range of spatial and temporal scales, from the deep-time history of microbial life to the future of a system no less immense and complex than the entire Earth. Modeling is thus a central and indispensable tool across the geosciences. Here, we review both the history and current state of model-based inquiry in the geosciences. Research in these fields makes use of a wide variety of models, such as conceptual, physical, and numerical models, and more specifically cellular automata, artificial neural networks, agent-based models, coupled models, and hierarchical models. We note the increasing demands to incorporate biological and social systems into geoscience modeling, challenging the traditional boundaries of these fields. Understanding and articulating the many different sources of scientific uncertainty – and finding tools and methods to address them – has been at the forefront of most research in geoscience modeling. We discuss not only structural model uncertainties, parameter uncertainties, and solution uncertainties, but also the diverse sources of uncertainty arising from the complex nature of geoscience systems themselves. Without an examination of the geosciences, our philosophies of science and our understanding of the nature of model-based science are incomplete.

41.1 What Are Geosciences?

The geosciences (sometimes also referred to as the Earth sciences) cover a very broad spectrum of disciplines including geology, paleontology, hydrology (the distribution and movement of water, on the surface and underground), glaciology (the study of ice and glaciers), climate science, oceanography, geophysics (the internal structure of the Earth, its gravitational and magnetic fields, plate tectonics, and volcanology), and geomorphology (how surface landscapes change over time). There is significant overlap between these different subfields because the various subsystems of the Earth are not isolated from one another and are often interacting in complex ways. Usually, the geosciences are understood as ending where biological systems begin, but given, for example, the great relevance of plants for the hydrological cycle (e.g., ecohydrology) and erosion phenomena (e.g., biogeomorphology), as well as the great relevance of human activity in altering the climate, landscapes, and oceans, this division is becoming increasingly difficult to maintain [41.1].

Although the geosciences have traditionally focused on the Earth, the conceptual and disciplinary divides between studies of the Earth and studies of other planets are also breaking down. For example, the wealth of new information coming from the space program (e.g., the Mars Rovers, HiRise images from the Mars Reconnaissance Orbiter, and images of various planets and moons from the Cassini-Huygens spacecraft and the New Horizons Pluto mission) has helped to generate the field of *planetary* geomorphology in addition to terrestrial (Earth) geomorphology. Planetary geomorphology includes the study of landscapes on not only planets, but also on moons (such as Saturn's moon Titan, which has the largest dune field in our Solar System) and other large celestial bodies (such as the Comet 67P which was determined by the Rosetta-Philae lander module to have water).

The phenomena that geoscientists investigate are extremely complex and can span a vast range of spatial and temporal scales. Hence, idealized models play a central role in all of the geosciences. These models are used for a variety of purposes, including both prediction and explanation. They are used not only for basic scientific research (theoretical tools for advancing insight and understanding) but also for planning purposes, policy, and hazard mitigation. Models are used to forecast a wide range of phenomena of human interest, such as earthquakes, volcanic eruptions, landslides, flooding, the movement of groundwater and spread of contaminants, and coastal erosion.

The geosciences are one of the most rapidly growing areas of interest in scientific modeling. This is led, in large part, by the tremendous amount of attention and resources that have been invested recently in climate modeling. Climate science is not unique, however, and many of the methodological issues found there are in fact widespread among the Earth sciences. Although, traditionally, philosophers of science have largely neglected the geosciences, leaving it a philosophical terra incognita [41.2], it is increasingly being recognized that our picture of the nature of science is inadequate if we do not take this research in the geosciences into account.

A complete review of all the relevant work in the diverse domains of the geosciences – and all the conceptual and methodological issues in modeling that arise within these different fields – is not possible in a single chapter. We provide here an overview of philosophical perspectives on this research that we hope will encourage more scholars to explore these topics further. The sections are organized primarily by the relevant conceptual and methodological issues.

41.2 Conceptual Models in the Geosciences

Conceptual models are the first step one takes before creating a more formal model (i. e., either a physical or numerical model). It is conceptualization of the key processes operating in the system of interest and the interactions between the components in the system. A conceptual model can simply take a narrative form or it can be an elaborate diagram. Typically, however, conceptual models can yield only qualitative predictions.

Some of the earliest models in geomorphology were conceptual models. Two historically important examples of conceptual models are Grove Karl Gilbert's (1843–1918) *balance of forces* conceptual model and William Morris Davis's (1850–1934) *cycle of erosion* conceptual model. In 1877, Gilbert introduced a conceptual model of a stream that appealed to physical concepts such as equilibrium, balance of forces, and work to explain the tendency of a stream to produce a uniform-grade bed. *Gilbert* describes his conceptual model as follows [41.3, p. 112]:

> "Let us suppose that a stream endowed with a constant volume of water is at some point continuously supplied with as great a load as it is capable of carrying. For so great a distance as its velocity remains the same, it will neither corrade (downward) nor deposit, but will leave the grade of its bed unchanged. But if in its progress it reaches a place where a less declivity of bed gives a diminished velocity, its capacity for transportation will become less than the load and part of the load will be deposited. Or if in its progress it reaches a place where a greater declivity of bed gives an increased velocity, the capacity for transportation will become greater than the load and there will be corrasion of the bed. In this way a stream which has a supply of *débris* equal to its capacity, tends to build up the gentler slopes of its bed and cut away the steeper. It tends to establish a single, uniform grade."

As *Grant* et al. note [41.4, p. 9]:

> "Gilbert's greatest and most enduring contribution to conceptual models in geomorphology [...] was the application of basic principles of energy and thermodynamics to the behavior of rivers. He did so with clarity of expression and an absence of math-

ematics that appeals directly to intuition, logic, and analog reasoning."

Note that Gilbert's model provides the conceptual foundation on which a numerical model, giving an equation to describe the balance of these forces, could be constructed, though he himself does not take this further step.

Another seminal conceptual model in the history of geomorphology is Davis's cycle of erosion [41.5]. *Davis* was a professor of geology at Harvard University; in an 1899 article entitled *The geographical cycle*, he established a framework for thinking about modeling in geomorphology [41.6, p. 481]:

> "All the varied forms of the lands are dependent upon – or, as the mathematician would say, are functions of – three variable quantities, which may be called structure, process, and time."

The evolution of a landscape may be understood as a cycle, which begins with a *peneplain* (a low relief plain) near a base (e.g., sea) level, is followed by rapid uplift leading to a youthful stage of rugged topograph, in which streams become established, and then a mature stage of tectonic stability in which those streams widen and gradually erode the landscape back down toward the base level. Finally, there will be an *old age* stage, involving low-relief landscapes with hills where mountains used to be. This then becomes eroded back to the peneplain stage until tectonic activity resumes and the cycle begins again. He used this idea to explain, for example, the features of the Appalachian mountains. It was a qualitative and explanatory conceptual model: it sought to explain and provide qualitative predictions for various features of a landscape.

For Davis, the conceptual model was the end point of his research; in recent years, most geoscientists have sought to quantify these sorts of processes. Thus, conceptual models can be seen as either the final product (an end it itself), or as a preliminary step in the process of creating a physical or mathematical model. In the case of mathematical models, there are two levels of modeling at which questions can be raised: Is the fundamental conceptual model adequate? And has that conceptual model been adequately represented or captured by that particular choice of mathematical equations?

41.3 Physical Models in the Geosciences

Until the mid-twentieth century, most conceptual models in the geosciences were realized as physical models. Physical models, also sometimes referred to as *hardware* or *table top models*, are a (usually, but not always) scaled-down version of the physical system of interest. In the geosciences, the systems of interest are typically large-scale, complex, open systems that are not amenable to experimental manipulation. A physical model allows a geoscientist to bring a version of the landscape into the laboratory, manipulate various variables in a controlled way, and explore hypothetical scenarios.

One of the central questions for geologists in the late nineteenth century was the origin of mountains (a subject known as orogenesis, from the Greek word *oros* meaning mountain). A popular orogenic theory in the nineteenth century was that mountains resulted from an overall contraction of the Earth, which was thought to be a consequence of the nebular hypothesis, first proposed by *Immanuel Kant* [41.7] and *Pièrre-Simone Laplace* [41.8]. To explore this hypothesis, the Swiss geologist Alphonse Favre (1815–1890) built a physical model involving layers of clay on a piece of stretched rubber, which was then released and the resulting structures were observed. The ability of this model to successfully reproduce some of the features of mountains led Favre to conclude that it supported the *plausibility* of the hypothesis [41.9, p. 96]. It was, what we would today call, a "how-possibly model explanation" (see Chap. 4, Sect. 4.4).

One of the great challenges for physical modeling in the geosciences, however, is that the relevant pressures, temperatures, durations, etc., of geological processes are largely beyond our reach. This limitation was recognized in the nineteenth century by the French geologist and director of the École Nationale des Mines, Auguste Daubrée (1814–1896), who notes (*Daubrée* [41.10, p. 5], [41.11] quoted in *Oreskes* [41.9, p. 99]),

> "[T]he equipment and forces that we can set to work are always circumscribed, and they can only imitate geological phenomena at the scale [. . .] of our own actions. "

In order to make further advances in physical modeling in the geosciences, it was realized that the relevant forces and processes would have to be appropriately scaled in the model. The quantitative mathematical theory by which such scaling could be achieved, however, would not be developed until the work of M. King Hubbert (1903–1989), an American oil company geologist, in the late 1930s and 1940s.

Hubbert's work provided [41.9, p. 110]:

> "the first fully quantitative treatment of the question how to choose the physical properties of materials in a model to account for the much smaller scale and time frame as compared with nature."

Hubbert's 1945 paper begins by noting the paradox that has long perplexed the geologic sciences: How could an Earth whose surface is composed of hard, rigid rock have undergone repeated deformations as if it were composed of a plastic material, as field observations of mountains and strata suggest? He notes that this paradox is a result of failing to adequately consider the concept of physical similarity, which like geometric similarity in a map, requires that *all* the relevant physical quantities (not just lengths, but densities, forces, stresses, strengths, viscosities, etc.) bear *constant ratios* to one another [41.12, p. 1638]. He notes that when the strengths are appropriately scaled, the resulting strength of the rock on a human scale is "comparable with that of very soft mud or pancake batter" [41.12, p. 1651]. So, for example, since the elastic properties of solids depend on the strain rate, scale models that operate orders of magnitude faster than terrestrial processes need to use materials that are orders of magnitude weaker than terrestrial rocks [41.9, p. 113]. Hubbert's work on scaling not only helped explain the puzzling field observations, but also provided the key to more adequate physical modeling.

Physical models can be classified by how they do or do not scale down. At one extreme there are *life size* (1 : 1) replica models of the system of interest. Sometimes such 1 : 1 physical models are a localized study of a particular process, such as *Bagnold*'s [41.13] use of a wind tunnel to study how grains of sand saltate form ripples. However, a full-scale physical model can also be an entire complex system, such as the Outdoor Streamlab at the University of Minnesota. In this full-scale model of a river segment, water and sediment flow down an artificial river system where the sediment is collected, measured, and recirculated to a sediment feeder. Although such replica models are able to avoid some of the problems arising from scaling issues (discussed below), they still involve simplifications and *laboratory effects* that can affect the reliability of the conclusions drawn for their real-world counterparts. More generally, however, many of the systems that geoscientists are interested in (e.g., mountain ranges and coastlines) are simply too large to be recreated on a 1 : 1 scale; hence, this type of physical model is typically not feasible.

Scale models are physical models that have been shrunk down according to some scale ratio (scale models can in principle be enlarged versions of their real-world counterparts, though this is not typical in the geosciences). For example, a 500 m-wide real river may be represented by a 5 m-wide scaled physical model, in which case the scale is 1 : 100. As Hubbert realized, simply shrinking a system down by some factor, however, will rarely preserve the necessary dynamical relations [41.14, p. 4]:

> "A true scaled model requires perfect geometric, kinematic, and dynamic similitude, something that cannot be achieved when using the same fluid as in the real world system due to equivalent gravitational and fluid motion forces."

Further complicating accurate scale modeling is the fact that different hydrodynamic processes are occurring at different spatial scales, and different physical effects can become dominant at those different scales too. For example, when scaling down one might substitute a fine sand for a pebbly gravel, but then cohesive forces can become dominant in the model when they are negligible in the target. These are examples of what are known as *scale effects*, when the force ratios are incomparable between the model and target. In such cases, one might need to substitute a liquid with a different viscosity or a different bed material into the model to try to overcome these scaling limitations – an example of how modelers sometimes deliberately get things more wrong in the model in order to get the conclusions to come out more right.

More often, the physical models are *distorted scale models*, where not all factors are scaled by the same ratio. The San Francisco Bay model, which is a table-top working hydraulic model of the San Francisco bay and Sacramento–San Joaquin River Delta system built by the US Army Corps of engineers, is an example of a geometrically distorted scale model, with the horizontal scale ratio being 1 : 1000, while the vertical scale ratio is only 1 : 100, and the time scale being 15 min to one day (for a philosophical discussion of this model see *Weisberg* [41.15]). Relaxing scale requirements further get what are sometimes referred to as *analog physical models*, where one reproduces certain features of a target system without satisfying the scale requirements. These are typically seen as physical systems to be investigated in their own right for what they can teach us about certain physical processes, rather than miniature versions of some specific real system [41.14, p. 5].

Physical models have their own strengths and weaknesses. The strengths, as mentioned, involve bringing a version of the system of interest into the laboratory as a closed system that is amenable to experimental manipulation and control. One does not need to have a mathematical representation of the system in order to explore its behavior. The weaknesses, or limitations, of physical models predominantly fall into two classes: *laboratory effects* and *scale effects*. Laboratory effects are those that occur in the laboratory system but not in the real-world counterpart. These can be related to model boundary conditions (sometimes literally the wall or edge of the table) where the behavior can drastically change, unrealistic forcing conditions, or the omission of causally relevant factors in the model. Scale effects refer to problems in maintaining the correct relations between variables when they are scaled down. This can lead to certain forces (e.g., cohesive forces) becoming dominant in the model that are not dominant in nature. More generally, these laboratory and scale effects are yet another example of the problem of external validity: Does the model accurately reflect the behavior of the system in the real world? This problem is pervasive among the sciences, and physical models are no more immune to it, despite dealing with the same physical stuff as their target.

41.4 Numerical Models in the Geosciences

Numerical models are mathematical models that represent natural systems and their interactions by means of a system of equations. These equations are typically so complex that they cannot be solved analytically, and so they have to be solved by numerical methods (such as finite difference or finite volume methods) that provide an approximate solution, or the equations need to be substituted with alternative algorithms, such as cellular automaton models. Numerical models are often implemented on a computer in a simulation that shows how the model will behave over an extended period of time, with some sort of graphical output to visualize that behavior (for a review of some of the philosophical issues in computer simulations see *Winsberg* [41.16]). This has enabled geoscientists to do something that they were generally unable (and often unwilling) to do in the past: to expand the goals of the geosciences to include forecasting and prediction as well as explanation.

In the context of the geosciences, there are many different kinds of numerical models, which can be categorized in different ways. The British geomorphologist *Kirkby* et al. [41.17], for example, distinguish the following four broad types of numerical models:

1. Black-box models
2. Process models
3. Mass–energy balance models
4. Stochastic models.

As *Kirkby* et al. explain, *black-box models* are models where "the system is treated as a single unit without any attempt to unravel its internal structure" [41.17, p. 16]. *Tucker* [41.18] gives as an example of a black-box model what is known as *Horton's laws* of river network topology. The law predicts the average number of branching stream segments of a certain order (roughly size or width). It was discovered by Robert Horton in 1945 from purely empirical analyses of stream basins, but gives no insight into why this so-called law would hold (it is not a law in the traditional sense, in that it does not hold universally). Black-box models are phenomenological models that involve a brute fitting to the empirical data. Although such models give no insight or understanding of the internal processes, they can be useful for making predictions.

At the other extreme of numerical modeling are *process models*, which try to describe the internal mechanisms giving rise to the empirical relations. *Tucker* explains, while [41.18, p. 687]:

> "a black-box model of soil erosion would be based on regression equations obtained directly from data [...] a process model would attempt to represent the mechanics of overland flow and particle detachment."

In between these two extremes are what *Kirkby* et al. [41.17] have called *grey-box models*, where some mechanisms may be known and included, but the rest is filled by empirical relations.

An important class of process models are landscape evolution models (LEMs). LEMs are numerical models in which the evolution of the landscape is related to the key underlying physical processes. These include, for example, the physical and chemical processes of rock weathering leading to rock disintegration and regolith production (*regolith* is a generic term referring to loose rock material, such as dust, soil, and broken rock, that covers solid rock), gravity-driven mass movement/landsliding, and water flow/run off processes (e.g., represented by the St. Venant shallow-water equations, which are a vertically integrated form of the Navier–Stokes equations). Each of these processes is represented mathematically by a *geomorphic transport function* (GTF), which get linked together to form the LEM. LEMs are often constructed as a software framework within which a variety of different component processes (represented by a particular choice of GTFs or equations), arranged in a particular configuration, can be implemented. Examples of such LEMs include the channel-hillslope integrated landscape development (CHILD) model, developed by *Tucker* et al. [41.19], and the cellular automaton evolutionary slope and river (CAESAR) model developed by *Coulthard* et al. [41.20]. These LEMs can simulate the evolution of landscapes on scales ranging from 1 to 500 km^2 and temporal scales ranging from days to millennia.

Often a component of LEMs, but sometimes presented as a model on their own, are *mass-balance models* (or *energy-balance models*). Mass-balance models use the fact that mass–energy is conserved to develop a continuity equation to describe the movement of mass (or energy) between different *stores*. A store could be anything ranging from water in lake, the population of a species in ecosystem, the energy stored as latent heat in an atmospheric column, the carbon mass in a tree, to the depth of soil at a point on a hillslope [41.18, p. 688]. An example of a mass-balance numerical model is a glacier model that describes the relation between ice accumulation and ablation (by melting and sublimation) at a given point of time under certain climate conditions [41.21]. Similarly, an energy-balance model in glaciology would be one that calculates the energy (heat) fluxes at the surface of the glacier that control melting and affect mass balance.

Climate science is a field of the geosciences in which both energy-balance and process numerical models have been developed to a high level of sophistication. Energy-balance models represent the climate of the Earth as a whole, without detailed information about processes or geographical variation. General circulation models (GCM) go a step further in explicitly representing atmospheric and oceanic processes. The most recent generation of climate models are Earth system models (ESM), which additionally include information about the carbon cycle and relevant biogeochemical processes. More specifically, ESMs are a composite of a large number of coupled models or *modules*, including an atmospheric general circulation model, an oceanic general circulation model, an ice dynamics model, biogeochemistry modules for land and ocean (e.g., for tracking the carbon cycle), and a software architecture or framework in which all these modules are integrated and able to communicate with each other. Developing and running GCMs and ESMs require a large number of collaborating scientists (scores to hundreds), significant supercomputing time, and millions of dollars. Because of the resource-intensive nature of such modeling projects, there are currently only a few dozen of them, and their outputs are periodically compared in intercomparison projects (e.g., coupled model intercomparison project (CMIP5) [41.22]). (For more on coupled models and intermodel comparison projects, see Sect. 41.9 below.) At present, GCMs and ESMs typically have a spatial resolution of 100–300 km; to fill this gap at the finer level of resolution, regional climate models (RCMs) have been developed for various locations.

While the trend in climate modeling has been toward increasing the complexity of these models with ever more process modules being added, there has recently been an interesting debate about whether a fundamentally new approach to climate modeling is required (for an excellent review and assessment of the leading proposals see *Katzav* and *Parker* [41.23]). More generally the trend toward ever more complex models in the geosciences has led to what *Naomi Oreskes* calls the *model-complexity paradox* [41.24, p. 13]:

> "The attempt to make models capture the complexities of natural systems leads to a paradox: the more we strive for realism by incorporating as many as possible of the different processes and parameters that we believe to be operating in the system, the more difficult it is for us to know if our tests of the model are meaningful."

In opposition to this trend, many geoscience modelers have started developing what are known as *reduced complexity models*, which are motivated by the idea that complex phenomena do not always need complex models, and simpler models may be easier to understand and test. A simpler model may also be run more often, and with more different parameters, making it more amenable to sensitivity analysis (see Sect. 41.6.3 below).

In the context of geomorphology, reduced complexity modeling is often defined in contrast with what is termed *simulation* modeling (*simulation* here refers not to models that are run as a computer simulation, but rather models that try to simulate or mimic all the details of nature as closely as possible). While simulation models try to remain grounded in the fundamental laws of classical mechanics and try to represent as many of the processes operating, and in as much detail, as is computationally feasible, reduced complexity models represent a complex system with just a few simple rules formulated at a higher level of description. As physical geographers *Nicholas* and *Quine* note, emphasis added [41.25, p. 319]:

> "In one sense, the classification of a model as a *reduced complexity* approach appears unnecessary since, by definition, all models represent simplifications of reality. However, in the context of fluvial geomorphology, such terminology says much about

both the central position of classical mechanics within theoretical and numerical modeling, and the role of the individual modeler in defining what constitutes an acceptable representation of the natural environment."

One of the first successful reduced complexity models in geomorphology was a cellular automata-type model of a braided river (i. e., a river with a number of interwoven channels that shift over time) that used just two key rules [41.26]. This model was heralded as a paradigm shift in geomorphic modeling [41.27, p. 194]. As Brad Murray, one of the proponents of this approach, argues, knowing how the many small-scale processes give rise to the large-scale variables in the phenomenon of interest is a *separate* scientific endeavor from modeling that large-scale phenomenon (*Murray* [41.28]; see also *Werner* [41.29]). Although reduced complexity models may seem like caricatures of their target systems, they can be surprisingly successful in generating realistic behaviors and providing explanatory insight (for further philosophical discussion of reduced complexity models and this case, see *Bokulich* [41.30] and *Murray* [41.31]).

41.5 Bringing the Social Sciences Into Geoscience Modeling

The geosciences are considered a branch of the physical sciences, being concerned with the chemistry and physics of the Earth, its history, and (more recently) its future. As such, the geosciences are typically thought of as excluding the domains of both the biological sciences and social sciences. Maintaining these artificial divisions, however, has increasingly become difficult. As *Oreskes* argues [41.1, p. 247]:

> "Many, perhaps, most, significant topics in Earth science research today address matters that involve not only the functioning of physical systems, but the interaction of physical and social systems. Information and assumptions about human behavior, human institutions, and infrastructures, and human reactions and responses are now built into various domains of Earth scientific research, including hydrology, climate research, seismology and volcanology."

For example, hydrological models that attempt to predict groundwater levels on the basis of physical considerations alone, can be inadequate for failing to include possible changes in human groundwater pumping activity, an external forcing function that can have dramatic effects on the physical system.

Climate science is another domain of the geosciences in which the need to incorporate the social sciences (specifically patterns and projections of human behavior involving, e.g., emission scenarios and deforestation practices) is evident. The Intergovernmental Panel on Climate Change (IPCC) has attempted to incorporate these social factors by three separate working groups, the first on the physical basis and the others on the social and policy dimensions, each issuing separate reports, released at different times. But, as Oreskes notes, the social variables are not just relevant to the social–policy questions, but to "the work that provides the (allegedly) physical science basis as well" [41.1, p. 253].

Increasingly geoscientists are being called upon to not only use their models to predict geoscience phenomena, but also to perform risk assessments and to communicate those risks to the public. Given that geoscientists are typically not trained in risk assessment, risk policy, or public communication, the results can be troubling. Oreskes recounts the high-profile case of the 2009 earthquake in central Italy that killed 309 people, and for which six geophysicists were sentenced to six years in prison for involuntary manslaughter in connection with those deaths. Although the international scientific community expressed outrage that these seismologists were being charged with failing to predict the unpredictable, the prosecutor, as reported in *Nature* painted a different picture (*Hall* [41.32, p. 266]; quoted in *Oreskes* [41.1, p. 257]):

> "'I'm not crazy', Picuti says. 'I know they can't predict earthquakes. The basis of the charges is not that they didn't predict the earthquake. As functionaries of the state, they had certain duties imposed by law: to evaluate and characterize the risks that were present in L'Aquila.' Part of that risk assessment, he says, should have included the density of the urban population and the known fragility of many ancient buildings in the city centre. 'They were obligated to evaluate the degree of risk given all these factors', he says, and they did not."

Oreskes concludes from this case [41.1, p. 257]:

> "[s]eismology in the twenty-first century, it would seem, is not just a matter of learning about earthquakes, it is also about adequately communicating what we have (and have not) learned."

Whether it is communicating the risks revealed by geoscience models or incorporating social variables directly into geoscience models, geoscientists are under increasing pressure to find ways to model these hybrid geosocial systems.

In some areas, such as geomorphology, agent-based models (ABMs) (which are common in fields such as economics) are starting to be used. ABMs consist of a set of agents with certain characteristics, following certain rules of self-directed behavior, a set of relationships describing how agents can interact with each other, and an environment both within which, and on which, the agents can act. As *Wainwright* and *Millington* note [41.33, p. 842]:

> "Despite an increasing recognition that human activity is currently the dominant force modifying landscapes, and that this activity has been increasing through the Holocene, there has been little integrative work to evaluate human interactions with geomorphic processes. We argue that ABMs are a useful tool for overcoming limitations of existing [. . .] approaches."

These ABM models, with their simplistic representation of human behavior, however, face many challenges, including not only difficulties in integrating the different disciplinary perspectives required to model these hybrid geosocial systems, but also issues of model evaluation.

41.6 Testing Models: From Calibration to Validation

41.6.1 Data and Models

Empirical data was long assumed to be the objective and unimpeachable ground against which theories or theoretical models are judged; when theory and data clashed, it was the theory or model that was expected to bend. Beginning in the early 1960s, however, philosophers of science including *Kuhn* [41.34, pp. 133–134], *Suppes* [41.35], and *Lakatos* [41.36, pp. 128–130] began to realize that this is not always the case: sometimes it is reasonable to view the theory as correct and use it to interpret data as either reliable or faulty. In a 1962 paper called *Models of Data*, Suppes argued that theories or theoretical models are not compared with raw empirical data, but rather with models of the data, which are a cleaned up, organized, and processed version of the data of experience. The production of a data model can involve, among other things, *data reduction* (any data points that are due to error or noise, or what are otherwise artifacts of the experimental conditions are eliminated from consideration) and *curve fitting* (a decision about which of several possible curves compatible with the data will be drawn).

This same insight has been recognized by scientists as well. The ecological modeler *Rykiel*, for example, writes, "Data are not an infallible standard for judging model performance. Rather the model and data are two moving targets that we try to overlay one upon the other" [41.37, p. 235]. Similarly *Wainwright* and *Mulligan* argue that the data of measurements are an abstraction from reality depending on timing, technique, spatial distribution, scale, and density of sampling. They continue [41.33, p. 13]:

> "If a model under-performs in terms of predictive or explanatory power, this can be the result of inappropriate sampling for parametrization or validation as much as model performance itself. It is often assumed implicitly that data represents reality better than a model does (or indeed that data is reality). Both are models and it is important to be critical of both."

A similar point has been made by the historian *Paul Edwards* [41.38] in his book on the development of climate modeling. There he traces in detail the changing meaning of *data* in meteorology and atmospheric science, noting how the existing incomplete, inconsistent, and heterogeneous data had to be transformed into a complete and coherent global dataset, with large numbers of missing gridpoint values interpolated from computer models in a process known as "objective analysis" [41.38, p. 252]. *Edwards* further argues that even the data obtained from measuring instruments is model-laden. He notes, for example, that [41.38, pp. 282-283]:

> "meteorology's arsenal of instrumentation grew to include devices, from Doppler radar to satellites, whose raw signals could not be understood as meteorological information. Until converted – through modeling – into quantities such as temperature, pressure, and precipitation, these signals did not count as data at all."

The importance of recognizing this model-ladenness of data is vividly illustrated in *Elizabeth Lloyd*'s [41.39] recounting of the high-profile case in which it was claimed in a US congressional hearing that data from satellites and weather balloons contradicted climate model evidence that greenhouse warming was occurring. In the end, the climate models were vindicated as more reliable than the data. Lloyd concludes from this case that we need to move towards a more complex empiricist understanding of the nature of data.

The data from measurements can, for example, be skewed by the fact that measurements are local, and yet the model might require a more global value (especially when there is significant heterogeneity), or more generally that measurements can only be made at one scale, and yet have to be extrapolated to another scale. Hence, when using data to parameterize, calibrate, or validate a model (see below) it is important to be aware of the limitations of the *data model* as well, and pay attention to any biases or errors that may have been introduced during the collection and processing of that data.

In some areas of the geosciences, such as paleontology, models have even been used to correct biases in available data. For example, one aim of paleontology is to gather information about the deep-time history of biodiversity (ranging from the Cambrian explosion to the various mass extinctions) on the basis of the observed fossil record. The conditions under which fossils are formed, preserved, and revealed are not only rare, but highly contingent and uneven with respect to space, time, and type of organism. Hence, there is arguably a strong detection (or sampling) bias in the observations. While some have taken the paleodiversity curves constructed from these fossil observations as a literal description of ancient biodiversity, others have argued that observed paleodiversity is a composite pattern, representing a biological signal that is overprinted by variation in sampling effort and geological drivers that have created a nonuniform fossil record [41.40, 41]. Before any evolutionary theories can be tested against the data of the fossil record, these data need to be corrected to extract the relevant biological signal from other confounding factors. Thus, for example, "many vertebrate paleodiversity studies have relied on modeling approaches (e.g., multivariate regression models) to 'correct' data for uneven sampling" [41.40, p. 127]. Of course, how the data are to be properly corrected, including which models of possible drivers and sources of bias are included in the multivariate analysis yielding the corrected data, involves substantial theoretical assumptions. As *Kuhn* noted years ago, observations are not "given of experience", but are "collected with difficulty" [41.34, p. 126].

The model-ladenness of data has led philosophers such as *Giere* to claim that "it is models almost all the way down" [41.42, p. 55] – a conclusion *Edwards* [41.38] argues is strongly supported by his historical analysis of the nature of data in meteorology and atmospheric science. Others, such as *Norton* and *Suppe*, have taken this conclusion even further, arguing that it is models *all* the way down. They write [41.43, p. 73]:

> "Whether physically or computationally realized, all data collection from instruments involves modeling. Thus raw data also are models of data. Therefore, there is no important epistemological difference between raw and reduced data. The distinction is relative."

However, saying that all data is model-laden to some degree does not imply that there is no epistemological difference, nor that all models are epistemically on par [41.44, pp. 103–104]. One of the most underdeveloped issues in this literature on data models is an analysis of what makes some data models better than others, and under what sorts of conditions data models should – or should not – be taken as more reliable than more theoretical models.

41.6.2 Parametrization, Calibration, and Validation

In mathematical modeling, one can distinguish variables, which are quantities that can vary and are to be calculated as part of the modeling solution, and parameters, which are quantities used to represent intrinsic characteristics of the system and are specified external to the model by the modeler. Also specified external to the model are the boundary conditions and the initial conditions (the latter describe the values of the variables at the beginning of a model run). Whether something is a variable or parameter depends on how it is treated in a particular model. Parameters need not be constant and can also vary across space, for example, but how they vary is specified external to the model. One can further distinguish two general types of parameters: those related to characteristics of the dynamics of a process and those related to the characteristics of a specific system or location where the model is being applied [41.45, p. 7].

Sometimes parameters can be universal constants (e.g., gravitational acceleration or the latent heat of water), in which case specifying their values is relatively unproblematic (though the process by which the values of constants are initially determined is nontrivial, and as *Chang* [41.46] cogently argues, challenges arise even in so-called basic measurements, such as temperature). More typically in the geosciences, however, the value of a parameter has to be determined on the basis of complex measurements, and even an idealization or averaging of those measurements (such as in the case of the parameter for bed roughness of a stream bed). The process by which input parameters are initially chosen has not been well studied, and is greatly in need of a better understanding. What has been the subject of considerable attention is the problem of calibration: the adjustment of model parameters in response to inadequate model performance.

In an ideal world, modelers would build a model based on physical principles and the equations that represent them, and then, with the use of appropriate input parameters for physical variables (like temperature, pressure, permeability, equilibrium constants, etc.), build a numerical simulation that accurately reflects the system under analysis. But most models do not do this: for a variety of reasons the match between the model output and available empirical information is often quite poor [41.47]. Therefore, modelers *calibrate* their models: they adjust the input parameters until the fit of the model to available information is improved to a level that they consider acceptable.

There are several concerns that can be raised about this process. One is that parameterized models are nonunique, and there is no way to know which particular set of parameterizations (if any) is the so-called right one; many different parameterizations may produce a given output. (This may be understood as a variation on the theme of underdetermination, discussed further below.) As hydrologist *Beven* notes [41.45, p. 7]:

> "parameters are usually calibrated on the basis of very limited measurements, by extrapolation from applications at other sites, or by inference from a comparison of model outputs and observed responses at the site of interest."

Moreover, because of the variability and uniqueness of many complex systems, parameter values extrapolated from one site may not be appropriate for another. Even if one restricts oneself to a given site, a model calibrated for one purpose (e.g., predicting peak runoff) may be predictively useless for another purpose (e.g., predicting total runoff) [41.33, p. 15]. Indeed, if the chosen parameterization is not an accurate representation of the physical system under consideration, it is likely that the model will not perform reliably when used for other purposes. This helps to explain the observation that many calibrated models fail, not only when used for purposes other than that for which they were calibrated, but sometimes even when used for their intended purposes [41.48].

Once a model has been built and calibrated, many modelers engage in an activity they call model validation, by which they normally mean the testing of the model against available data to determine whether the model is adequate for the purpose in question. Many geoscientists acknowledge that the use of the term *validation* should not be taken to imply that the model is true or correct, but rather only that "a model is acceptable for its intended use because it meets specified performance requirements" [41.37, p. 229]. *Rykiel* thus argues that before validation can be undertaken, the following must be specified:

a) The purpose of the model
b) The performance criteria
c) The context of the model.

However, many so-called validated models have failed even in their intended use. For example, in a 2001 study, Oreskes and Belitz showed that many hydrological models fail because of unanticipated changes in the forcing functions of the systems they represent. More broadly, validated models may fail for the following reasons [41.9, p. 119]:

1. Systems may have emergent properties not evident on smaller scales.
2. Small errors that do not impact the fit of the model with the observed data may nonetheless accumulate over time and space to compromise the fit of the model in the long run.
3. Models that predict long-term behavior may not anticipate changes in boundary conditions or forcing functions that can radically alter the system's behavior.

The idea that a model can be validated has been critiqued on both semantic and epistemic grounds. Semantically, *Oreskes* et al. have noted that the terminology of *validation* implies that the model is *valid* – and thus serves as a claim about the legitimacy or accuracy – a claim that, as already suggested above, cannot be sustained philosophically and is often disproved in practice [41.47, 49]. Hence, a better term than model validation might be *model evaluation*. Even with this change in terminology, however, epistemological challenges remain. In many cases, the available empirical data (e.g., historic temperature records) have already been used to build the model, and therefore cannot also be used to test it without invoking circular reasoning. Some modelers attempt to avoid this circularity by calibrating and validating the model against different historical time periods, with respect to different variables, or even different entities and organisms.

Paleontologists, for example, use biomechanical models to try to answer functional questions about extinct animals based on the structures found in the fossil record (which is a subtle and difficult process, see e.g., [41.50]). These biomechanical models, which are used to make predictions about paleospecies, are validated or tested against data for present-day species. More specifically, *Hutchinson* et al. have used such models to determine how fast large theropod dinosaurs, such as Tyrannosaurus rex, could run. They write [41.51, p. 1018]:

> "The model's predictions are validated for living alligators and chickens [...]. [m]odels show that in order to run quickly, an adult Tyrannosaurus would have needed an unreasonably large mass of extensor muscle."

Such an approach may work in cases where very large amounts of data are available, or where there are clearly distinct domains that may be enlisted. In many areas of the geosciences, however, data is scant and all available data need to be used in the initial construction of the model.

41.6.3 Sensitivity Analysis and Other Model Tests

Irrespective of the difficulties of model construction and calibration, models can be highly effective in helping to identify the *relative* importance of variables, through techniques such as sensitivity analyses. Sensitivity analysis – also known (inversely) as robustness analysis – is the process of determining how changes in model input parameters affect the magnitude of changes in the output (for philosophical discussions of robustness analyses see, e.g., *Weisberg* [41.52] or *Calcott* [41.53]; for a comprehensive, technical introduction to sensitivity analysis in a variety of domains see *Saltelli* et al. [41.54]). For example, in the context of the paleontology research on models of Tyrannosaurus rex introduced above, *Hutchinson* writes [41.55, p. 116]:

> "Because any model incorporates assumptions about unknown parameters, those assumptions need to be explicitly stated and their influences on model predictions need to be quantified by sensitivity analysis [...]. In many models this can be determined by varying one parameter at a time between minimal and maximal values (e.g., crouched and columnar limb poses) and evaluating the changes in model output (e.g., the required leg muscle mass)."

Varying one parameter at a time is known as a *local* sensitivity analysis. However, for some sorts of systems (especially systems in which nonlinearities and thresholds operate), a complicating factor is that model sensitivity to a parameter can also depend on the values of the other model parameters [41.56, p. 141] and [41.33, p. 18]. Hence, in these latter cases, one needs to perform what is known as a *global* sensitivity analysis, where all the parameters are varied simultaneously to assess how their interactions might affect model output [41.57].

Sensitivity analysis is used in nearly all domains of modeling, and it can be an important guide to data collection: alerting the scientific community to where additional or better empirical information is most likely to make a difference. That is to say, sensitivity analyses can reveal which parameters are most important in a model (and hence should be targeted for additional data collection) and which parameters are relatively unimportant or even negligible. It may thus suggest parameters that should be omitted, which can save on computational time. Sensitivity analyses can also help determine whether a model might be overparameterized, which involves a kind of overfitting to the data that occurs when too many parameters are included and fixed.

Model testing can involve a wide spectrum of different techniques, ranging from subjective expert judgments to sophisticated statistical techniques. *Rykiel* [41.37] has assembled a list of 13 different procedures, which he calls *validation procedures*. However, given the concerns raised above about the term validation and the heterogeneity of the procedures collected in his list, the broader rubric of *model tests* is arguably more appropriate. *Rykiel*'s list is as follows [41.37, pp. 235–237]:

1. Face validity, where experts are asked if the model and its behavior are reasonable.
2. Turing-test validity, where experts assess whether they can distinguish between system and model outputs.
3. Visual validation, where visual outputs of model are (subjectively) assessed for visual goodness of fit.
4. Inter-model comparisons.
5. Internal validity of model.
6. Qualitative validation: the ability to produce proper relationships among model variables and their dynamic behavior (not quantitative values).
7. Historical data validation, where a part of the historical data is used to build the model and a part is used to validate it.
8. Extreme conditions tests, where model behavior is checked for unlikely conditions.
9. Traces: the behavior of certain variables is traced through the model to see if it remains reasonable at intermediate stages.
10. Sensitivity analyses: the parameters to which the model is sensitive are assessed against the parameters to which the system is or is not sensitive.
11. Multistage validation: validation at certain critical stages throughout the model-building process.
12. Predictive validation: model predictions are compared to system behavior.
13. Statistical validation: statistical properties of model output are evaluated and errors are statistically analyzed.

Although, as noted before, the term validation is inappropriate and this heterogeneous list could be usefully organized into different categories, it nonetheless provides a good sense of the broad spectrum of techniques that modelers deploy in testing and evaluating their models. Each of the procedure on this list can play an important role in the modeling process and is arguably worthy of further philosophical and methodological reflection.

41.7 Inverse Problem Modeling

One of the central tasks of geophysics is to determine the properties of the interior structure of the Earth on the basis of measurements made at the surface. The primary method by which this is done is known as *inverse problem modeling*. Most broadly, an inverse problem is defined as that of reconstructing the parameters of a system or model based on the data it produces; in other words, one starts with a set of observational data and then tries to reason back to the causal structure that might have produced it. The inverse problem is contrasted with the *forward problem*, which involves starting with a known model and then calculating what observations or data that model structure will produce. Inverse problems are found across the sciences, such as in finding the quantum potential in the Schrödinger equation on the basis of scattering experiments, diagnostic imaging in medicine using X-ray computer assisted tomography, or, most relevantly here, determining information about the interior structure of the Earth on the basis of travel-time data of waves (e.g., earthquakes). Indeed, the first methods for solving inverse problems were developed in the context of seismology by a German mathematical physicist Gustav Herglotz (1881–1953) and the geophysicist Emil Wiechert (1861–1928).

A fundamental challenge for inverse modeling methods is the problem of underdetermination [41.58, p. 120]:

> "[T]he model one aims to determine is a continuous function of the space variables. This means the model has infinitely many degrees of freedom. However, in a realistic experiment the amount of data that can be used for the determination of the model is usually finite. A simple count of variables shows that the data cannot carry sufficient information to determine the model uniquely."

In other words, the solution to the inverse problem is not unique: there are many different models that can account for any given set of data equally well. This is true for both linear and nonlinear inverse problems [41.59].

One method for trying to constrain this underdetermination is known as the model-based inversion approach, which involves introducing a second, intermediary model known as the *estimated* or *assumed* model [41.60, p. 626]. The estimated model is used in the forward direction to generate *synthetic* data, which is then compared with the observational data. On the basis of the discrepancy between the two datasets, the estimated model is modified and the synthetic data it produces is again compared in an iterative optimization process. As *Snieder* and *Trampert* note, however [41.58, p. 121]:

> "There are two reasons why the estimated model differs from the true model. The first reason is the nonuniqueness of the inverse problem that causes several (usually infinitely many) models to fit the data [...] The second reason is that real data [...] are always contaminated with errors and the estimated model is therefore affected by these errors as well."

In other words, one must also be aware of errors arising from the data model (as discussed earlier). Different modeling approaches for dealing with inverse problems in geophysics have been developed, such as the use of artificial neural network (ANN) models (see, e.g., *Sandham* and *Hamilton* [41.61] for a brief review).

Recently, a number of philosophers of science have highlighted the philosophical implications of the underdetermination one finds in geophysical inverse problems. *Belot* [41.62], for example, argues that this "down to earth underdetermination" shifts the burden of proof in the realism–antirealism debate by showing that a radical underdetermination of theory by (all possible) data is not just possible, but actual, and likely widespread in the geosciences (and elsewhere). *Miyake* similarly calls attention to the problem of underdetermination in these Earth models and notes that there are additional sources of uncertainty that are not even considered in the setting up of the inverse problem [41.63]. He argues that thinking of these Earth models as a case of what philosophers [41.64] call *model-based measurement* is important for understanding the epistemology of seismology.

41.8 Uncertainty in Geoscience Modeling

Geoscientists have paid considerable attention to the problem of model uncertainty and sources of error, but many (if not all) of the sources of uncertainty they identify are not unique to the geosciences. There are different ways in which one can construct a taxonomy of the sources of uncertainty in modeling. One can, for example, organize the sources of uncertainty by the relevant stage in the modeling process. Here, one can group the various uncertainties into the following three categories:

1. Structural model uncertainties
2. Parameter uncertainties
3. Solution uncertainties.

Alternatively, one can also organize the sources of uncertainty in modeling on the basis of various complexities that arise for the sort of systems one is trying to model. This latter approach is taken by geomorphologist *Stanley Schumm* [41.65], who organizes the sources of uncertainty into the following three categories:

1. Problems of scale and place
2. Problems of cause and process
3. Problems of system response.

Each of these ways of thinking about sources of uncertainty in modeling serves to highlight a different set of philosophical and methodological issues.

Uncertainties can be identified at each step of the modeling process. During the construction phase of the model there are a number of uncertainties that can be grouped together under the broad rubric of *structural model uncertainties*. In this category, there are what are termed *closure uncertainties*, which involve uncertainties about which processes are to be included or not included in the model [41.66, p. 291]. There can be uncertainties regarding both which processes are in fact operating in the target system (some processes might be unknown) and which of the processes known to be operating are in fact important to include (we may know that a process is operating, but not think it is relevant). Sometimes whether a process is important, however, depends on what other processes are included in the model, as well as on other factors, such as the relevant spatiotemporal scale over which the model will be applied. As an example of this type of structural model (closure) uncertainty, *O'Reilly* et al. [41.67] discuss the case of early attempts to model stratospheric ozone depletion (that resulted in the unexpected *ozone hole* in the Antarctic, which was discovered in 1985). They write [41.67, p. 731]:

> "[B]efore the ozone hole discovery led scientists to rethink their conceptual models, ozone assessments had not considered such multiphase reactions [i. e., heterogeneous chemical reactions] to be important. At the time, gas-phase atmospheric chemistry was much better understood than multiphase chemistry, and heterogeneous reactions were seen as arcane and generally unimportant in atmospheric processes."

Because these chemical processes were not well understood scientifically and were not recognized as important to this phenomenon, they were left out of the model, leading to a drastic underprediction of the rate at which ozone depletion would take place. More generally, as *Oreskes* and *Belitz* have noted, when modelers lack reliable information about known or suspected processes, they may simply leave out those processes entirely, which effectively amounts to assigning them a value of zero [41.48, 67]. Such closure uncertainties in modeling can thus lead to significant errors.

Second, there are *process uncertainties*, which are concerned with how those processes should be represented mathematically in the model. For many processes in the geosciences, there is no consensus on the right way to represent a given process mathematically, and different representations may be more or less appropriate for different applications. For example, there are different ways that turbulence can be represented in models of river flow, from the greatly simplified to the highly complex [41.66, p. 291].

Third, there are what are more narrowly called *structural uncertainties*; these are uncertainties in the various ways the processes can be linked together and represented in the model. Included in this category are uncertainties associated with whether a component is taken to be active (allowed to evolve as dictated by the model) or passive (e.g., treated as a fixed boundary condition). *Lane* [41.66, p. 291] gives the example of the different ways the ocean can be treated in global climate models: because of water's high specific heat capacity, the ocean responds slowly to atmospheric changes; hence, if used on short enough time scales, the modeler can represent the ocean as a passive contributor to atmospheric processes (as a source of heat and moisture, but not one that in turn responds to atmospheric processes). *Parker* [41.68] also discusses structural uncertainty in climate modeling, with regard to the choice of model equations.

Structural model uncertainties can give rise to *structural model error*, which *Frigg* et al. [41.69] define broadly as a discrepancy between the model dynamics and target system dynamics. They demonstrate that in a nonlinear model, even a small structural model error can lead to divergent outcomes more drastic than those due to the sensitive dependence on initial conditions characteristic of chaotic systems. In analogy with the well-known butterfly effect, they (following *Thompson* [41.70]) call this the *hawkmoth effect*. They conclude that the structural model error in a nonlinear model "is a poison pill ... operational probability forecasts are therefore unreliable as a guide to rational action if interpreted as providing the probability of various outcomes" [41.69, p. 57]. Nonetheless, they note that such models may still be useful for generating insight and understanding.

In addition to these three types of structural model uncertainty (closure, process, and structure uncertainties), another significant source of uncertainty is *parameter uncertainty*. As discussed earlier, models contain both variables (whose values are determined by the model itself) and parameters (whose values must be specified externally by the modeler). In the global circulation or ESMs of climate science, parameters are used, for example, in representations of unresolved processes (such as cloud systems or ocean eddies) that are on a finer-grained scale than that on which the model operates. Ideally, the value of a parameter is determined directly by field measurements, but often this is not possible. In many cases, the parameter is either prohibitively difficult to measure or has no simple field equivalent. The parameters then need to be estimated or calculated on the basis of other models (e.g., as detailed by *Edwards* [41.38] in his discussion of parameters in meteorology and atmospheric science). *Beven* [41.45, p. 8] gives the example of the parameter representing soil hydraulic conductivity in hydrology. Measurements of soil hydraulic conductivity are typically made on soil samples in a small area, but are known to exhibit order of magnitude variability over even short distances. Often, however, the model will require a value of hydraulic conductivity over a much larger spatial scale (e.g., the whole catchment area). Hence, substantial uncertainties can arise as one tries to determine an *effective* value for the parameter.

Parameters can also take on different values than their real-world counterparts during the process of calibration or optimization. An example is the bed roughness parameter, which is used to represent the grain size of a river bed affecting the friction and turbulence of the flow. As *Odoni* and *Lane* note [41.71, p. 169]:

> "it is common to have to increase this quite significantly at tributary junctions, to values much greater than might be suggested by [...] the bed grain size. In this case there is a good justification for it, as one-dimensional models represent not only bed roughness effects but also two- and three-dimensional flow processes and turbulence."

In other words, the bed roughness parameter in the model is used to capture not just the bed roughness, but other effects that act like bed roughness on the behavior of the flow. This is another example of what was earlier called *getting things more wrong in order to get them more right*. More generally, parameter values determined for one model may be calibrated for that particular model structure, and hence not be independent of that model structure or even different discretizations or numerical algorithms of that model structure, and therefore are not transferable to other models without additional error [41.45, p. 8]. Hence, one must be aware of the problem of *parameter incommensurability*, where parameters that share the same name might in fact "mean" different things [41.45, p. 8].

Although they are not strictly speaking parameters, one can also include under this umbrella category uncertainties in the initial conditions and the boundary conditions, which also need to be specified externally by the modeler in order to operate the model. Examples include [41.71, p. 169]:

> "the geometry of the problem (e.g., the morphology of the river and floodplain system that is being used to drive the model) or boundary conditions (e.g., the flux of nutrients to a lake in a eutrophication model)."

In order to integrate a model forward in time, one needs to first input the current state of the system as initial conditions. Not only can there be uncertainties in the current state of the system, but also some chaotic models will be very sensitive to such errors in the initial conditions.

The final category of model uncertainties is *solution uncertainties*. Once the model equations are set up, the parameters fixed, and the initial and boundary conditions are specified, the next step is to solve or run the model. Often in geoscience modeling, the governing equations are nonlinear partial differential equations that do not have general analytic solutions. In such cases, one must resort to various discretization or numerical approximation algorithms (e.g., finite difference methods, finite element methods, boundary element methods, etc.) to obtain solutions, which will not be exact (though they can often be benchmarked against analytic solutions). There can also be uncertainties introduced by the way the algorithm is implemented on a computer for a simulation model. *Beven* notes [41.45, p. 6]:

> "[D]ifferent implementations will, of course, give different predictions depending on the coding and degree of approximation. [The] computer code [...] represents a further level of approximation to the processes of the real system."

In implementing a model on a computer, decisions must be made about the appropriate choice of time steps and spatial discretizations, and these and other solution uncertainties can lead to further sources of error.

In his book *To interpret the Earth: Ten ways to be wrong*, *Schumm* identifies 10 sources of uncertainty, which he organizes into the three categories of problems of scale and place, problems of cause and process, and problems of system response [41.65]. The first source of uncertainty concerns *time*. Compared to the long-time history over which Earth's landscapes evolve, the time scale of human observation is extremely short. There can be short-term patterns in geoscience phenomena that are very different from the long-term pattern one is trying to predict or explain; hence, extrapolations from short-term observations may not be reliable (e.g., the short-term wind direction you observe may not be indicative of the prevailing long-term wind direction that predominantly shapes the landscape). Also, different features of a landscape (and the corresponding different processes) can become salient as different time scales are considered. The processes that are most relevant on a short-time scale (such as storm events) may be insignificant on a long-time scale, as well as the reverse (e.g., uplift phenomena are negligible over the short term, but are such stuff as the Himalayas are made of over the long term). Hence, inadequate attention to these issues of time, both in the construction and application of the model, can be a significant source of uncertainty. The second source of uncertainty, *space*, is analogous to these problems of time. For example, to understand how water moves through the ground on a small spatial scale, the type of soil or rock (e.g., its porousness) might be most relevant to model, while on a large scale, the overall topology of the landscape (e.g., whether it is on a steep slope) and whether it has large-scale rills (cracks or channels) might be more relevant. The third source of uncertainty Schumm calls *location*, which relates to the uniqueness of geomorphic systems (e.g., there is a sense in which no two rivers are exactly the same, and hence models developed for one location, might not be applicable to other locations).

In the next cluster, Schumm identifies *convergence* as a fourth source of uncertainty. Convergence is the idea that different processes or causes can produce similar effects. For example, sinuous rills on the Moon look like dried river beds formed by flowing water, but were later concluded to be the result of collapsed lava tubes [41.65, p. 59]. Hence, one needs to be careful in inferring cause from effect, and in drawing an analogy from the causes of an effect at one location to the causes of a very similar effect at another location. The fifth source of uncertainty, *divergence*, is the opposite of convergence: the same cause can produce different effects. Schumm gives the example of glacio-eustasy, or the change of sea levels due to the melting of glaciers and ice sheets. He explains [41.65, p. 64]:

> "With the melting of the Pleistocene continental ice sheets the assumption is that a global sea-level rise will submerge all coastlines. However, the results are quite variable [...] [a]s a result of isostatic uplift following melting of the continental ice sheets."

Isostatic uplift refers to the rebounding or rise of land masses that were depressed under the massive weight of the ice sheets (this rebound is still ongoing and averages at the rate of a centimeter per year: see, e.g., *Sella* et al. [41.72]). In other words, the melting of glaciers and icesheets can cause sea levels both to rise and to fall (depending on the location): one cause, two different (and opposite) effects.

The sixth source of uncertainty Schumm identifies is what he calls *efficiency*, which he identifies with the assumption that the more energy expended, the greater the response or work done. He notes that this will not generally be the case [41.65, p. 66]:

> "When more than one variable is acting or when a change of the independent variable, such as precipitation, has two different effects, for example, increased runoff and increased vegetation density, there may be a peak of efficiency at an intermediate condition."

He gives as an example the rate of abrasion of a rock by blown sand, which has a maximum abrasion efficiency at relatively low rates of sand feed (presumably due to an interference of rebounding particles with incoming particles).

The seventh source of uncertainty he identifies is *multiplicity*, which is the idea that there are often multiple causes operating in coordination to produce a phenomenon, and hence one should adopt a *multiple explanation approach*. This concept originated in the work of the American geologist Thomas C. Chamberlin (1843–1928), specifically in his method of multiple working hypotheses, a method which he urged was beneficial not only to scientific investigation, but also to education and citizenship. In his 1890 article introducing this method he considers the example of explaining the origin of the Great Lake Basins. *Chamberlin* writes [41.73, p. 94]:

> "It is practically demonstrable that these basins were river-valleys antecedent to the glacial incursion, and that they owe their origin in part to the pre-existence of those valleys and to the blocking-up of their outlets [...]. So, again, it is demonstrable that they were occupied by great lobes of ice, which excavated them to a marked degree, and therefore the theory of glacial excavation finds support [...]. I think it is furthermore demonstrable that the earth's crust beneath these basins was flexed downward, and that they owe [...] their origin to crustal deformation."

What might initially appear to be a scientific controversy involving rival hypotheses or competing explanations, in fact turns out to be a case where each hypothesis correctly has part of the story. Chamberlin concludes that one benefit of considering diverse explanations for observed phenomena is that it forces the geologist to move beyond hasty or simplistic explanations, and instead to consider the possibility that more than one relevant process has been involved. (For a philosophical discussion of the method of multiple hypotheses in the case of plate tectonics, see *Rachel Laudan* [41.74].)

An example of this from paleontology is the long-standing debate about the cause of the Cretaceous (K–T) mass extinction (in which 70% of all species, including all the (nonavian) dinosaurs, went extinct). The favored explanation of this extinction event is the impact hypothesis: that the extinction was caused by a large comet or asteroid that hit Earth near present-day Chicxulub, Mexico. While the fact that this impact occurred is not in doubt, some scientists question whether the impact hypothesis can explain the gradual and step-wise extinction pattern that is observed in the fossil record. They favor instead an explanation that appeals to massive volcanism and climate change, which was already underway. While often viewed as rivals, these two explanations might be complementary [41.75]. *Schumm* concludes, "if there is more than one cause of a phenomenon, unless all are comprehended, extrapolation will be weak and composite explanations are needed" [41.65, pp. 74–75]. (For a more general philosophical discussion of explanation in the Earth sciences, including a discussion of the explanation of the K–T extinction, see *Cleland* [41.76].)

The final three sources of uncertainty Schumm identifies are *singularity*, the idea that landforms, though also having many commonalities, have features that make them unique, and hence respond to changes in slightly different ways or at different rates; *sensitivity*, the idea that small perturbations to a system can have significant effects, especially when a system involves either internal or external thresholds; and the *complexity* of geomorphic systems, which means they have numerous interconnected parts interacting in typically nonlinear ways. An example of an important threshold in the geosciences is the velocity at which a sediment particle of a given size is set in motion by a particular fluid (e.g., water or wind). This is an example of an extrinsic threshold involving changes in an external variable. There can, however, also be intrinsic thresholds in which there is an abrupt change in a system without there being a corresponding change in an external variable. For example, under constant weathering conditions the strength of materials can be weakened until there is an abrupt adjustment of the system (such as a landslide). Another example of an intrinsic threshold is when a bend or loop in a meandering river will suddenly be cut off by the formation of a new channel. More generally, geomorphic systems often exhibit what are called *autogenic behaviors*, in which there can be a sudden and pronounced change in the system's behavior or characteristics, not due to an external cause, but rather due to internal feedbacks in the system, in which gradual changes can result in sudden, threshold-like responses (for a discussion see *Murray* et al. [41.77]; for an example of an autogenic behavior discovered in the St. Anthony's Falls physical model discussed earlier, see *Paola* et al. [41.78]). *Schumm* concludes [41.65, p. 84]:

> "The recognition of sensitive threshold conditions appears to be essential in order that reasonable explanations and extrapolations can be made in geomorphology, soil science, sedimentology and stratigraphy, and many environmental and ecosystem areas."

So far we have reviewed five sources of uncertainty arising during stages of the modeling process and 10 sources of uncertainty arising from the complexity of geoscience systems. A further complication arises from the fact that even models with these sorts of errors can generate predictions that agree reasonably well with observations – a case of getting the right answer for the wrong reason. Hence, on pain of committing the fallacy of affirming the consequent, one cannot deductively conclude that one's model is right, just because it produces predictions that match observations. More generally, this is related to the fact that more than one model or theory can account for a given set of observations: the data underdetermine the model or theory choice. In the philosophical literature this is known as the problem of underdetermination (e.g., see *Duhem* [41.79], or for contemporary discussion, see *Stanford* [41.80]; for a philosophical discussion of underdetermination in the Earth sciences see *Kleinhans* et al. [41.2]). In the geoscience literature the problem of underdetermination is sometimes referred to as the problem of nonuniqueness, or *equifinality* [41.81]. *Beven* and *Freer* write [41.82, p. 11]:

> "It may be endemic to mechanistic modeling of complex environmental systems that there are many different model structures and many different parameter sets within a chosen model structure that may be [...] acceptable in reproducing the observed behavior of that system. This has been called the equifinality concept."

In other words, the data are not sufficient to uniquely pick out a model structure or parameter set. (A similar sort of equifinality was seen in the nonuniqueness of inverse problems discussed earlier.) Moreover, the acceptable parameter sets may be scattered throughout parameter space (i. e., not localized around some optimum parameter set). This problem of equifinality is not just hypothesized, but has been demonstrated in computer simulations, which are now cheap and efficient enough to allow explorations of the parameter space of models of a variety of geoscience systems.

The problem of equifinality has led Beven et al. to develop a method to deal with uncertainty that they call the generalized likelihood uncertainty estimation (GLUE) methodology [41.83]. GLUE involves a kind of Monte Carlo method with a random sampling of the space of possible model–parameter combinations, in which each possible set of parameters is assigned a likelihood function (assessing the fit between model predictions and observations). The idea is not to pick one *best* model–parameter set, but rather to take into account the predictions of all acceptable models (models not ruled out by current data or knowledge), weighted by their relative likelihood or acceptability, in something like a Bayesian averaging of models and predictions. (For a recent review and discussion of objections to the GLUE methodology see *Beven* and *Binley* [41.84].)

41.9 Multimodel Approaches in Geosciences

The GLUE methodology is just one of several different approaches that try to use multiple models in concert to reduce uncertainty. The GLUE methodology requires a large number of runs to adequately explore the parameter space. However, this is not typically feasible in computationally intensive models. An alternative approach that can be used with more complex models is the *metamodel* approach (for a review see *Kleijnen* [41.85]). A metamodel is a simplified surrogate model that is abstracted from the primary model and used to aid in the exploration of the primary model and its parameter space. While metamodels have long been used in engineering research, they have only recently started to be applied to models in the geosciences.

Odoni [41.86], for example, has applied the metamodel approach to the study of a landscape evolution model (LEM) developed by *Slingerland* and *Tucker* [41.87] known as GOLEM (where GO stands for geomorphic-orogenic). GOLEM has been used, for example, to model the evolution of a catchment landscape of the Oregon Coast Range around the headwaters of the Smith River over a period of 100 000 years. In order to understand how equifinality manifests itself in GOLEM, Odoni selected 10 parameters (related to mass movement, channel formation, fluvial erosion, and weathering processes) to vary over a range of values that was determined to be consistent with the location based on published data and calibration. The model outputs used to describe the landscape at 100 000 years include sediment yield, drainage density, sediment delivery ratio, and a topographic metric. Rather than trying to solve the full GOLEM model for the immense number of possible parameter value combinations, Odoni derived a metamodel, or set of regression equations, that described each model output as a function of the GOLEM parameters. As he explains, "The parameter space is then sampled rapidly and densely ($>> 1 \times 10^6$ times), using each metamodel to predict GOLEM's output at each sample point" [41.86, p. i]. In this way metamodels yield a clearer picture of what drives model output (leading to a possible further simplification of the model) and an understanding of where equifinality may be lurking. It is important to note that this equifinality is not just an abstract cooked-up possibility, but a genuine, wide-spread practical problem, making it yet another example of what Belot termed down-to-earth *underdetermination*.

More common than both the GLUE and metamodel approaches are classic *intermodel comparison* projects. The most well known here are the large-scale, multiphase intercomparison projects used by the IPCC in their assessments. The most recent coupled model intercomparison project (CMIP5), for example, compares the predictions of dozens of climate models running the same set of scenarios. The aim of such multimodel ensembles is to "sample uncertainties in emission scenarios, model uncertainty and initial condition uncertainty, and provide a basis to estimate projection uncertainties" [41.88, p. 369]. *Lloyd* has emphasized the strength of such multimodel approaches, arguing that it is "a version of reasoning from variety of evidence, enabling this robustness to be a confirmatory virtue" [41.89, p. 971].

The proper assessment of such intermodel comparisons for robustness and uncertainty reduction involves some subtleties, however (see, e.g., *Parker* [41.90, 91]; *Lenhard* and *Winsberg* [41.92]). Models can, for example, agree because they share some common model structure, rather than indicating model accuracy. As *Masson* and *Knutti* explain [41.93, p. 1]:

> "All models of course contain common elements (e.g., the equations of motion) because they describe the same system, and they produce similar results. But if they make the same simplifications in parameterizing unresolved process, use numerical schemes with similar problems, or even share components or parts thereof (e.g., a land surface model), then their deviations from the true system or other models will be similar."

In such cases an agreement among climate models does not indicate that modelers are on the right track. It remains unclear how best to conceptualize and assess model independence [41.23, p. 485]. More generally, the spread of an ensemble of models is often taken to approximate the uncertainty in our predictions; however, as *Knutti* et al. [41.94] have argued, these are *ensembles of opportunity*, not systematic explorations of model or parameter space. They suggest a number of ways forward, including having a larger diversity of models to help find constraints valid across structurally different models, and developing new statistical methods for incorporating structural model uncertainty [41.94, p. 2755]. There are many other multimodel approaches used in the geosciences, including *coupled models* and *hierarchical modeling*.

41.10 Conclusions

The geosciences provide a rich and fruitful context in which to explore methodological issues in scientific modeling. The problem of understanding and articulating scientific uncertainty has particularly come to the fore in these fields. The complex and multiscale nature of geological and geophysical phenomena require that a wide variety of kinds of models be deployed and a broad spectrum of sources of uncertainty be confronted. Most modelers do not expect their models to give specific, quantitative predictions of the detailed behavior of the systems under investigation. Rather, they are understood as providing a tool by which scientists can test hypotheses (including causal ones), evaluate the relative importance of different elements of the system, develop model-based explanations [41.95, 96], and generate qualitatively accurate projections of future conditions. Indeed, it is precisely by grappling with these many sources of uncertainty that geoscientists gain insight and understanding into the various processes that shape the Earth, their relative importance and patterns of dependence, and the emergent structures that they produce.

The geosciences, as we have seen, constitute a significant portion of scientific research today. Our philosophies of science and our understanding of the nature of model-based inquiry are inadequate if we do not take this research into account. As we hope this review has made clear [41.44, p. 100]:

> "the earth sciences are profoundly important, not only because they challenge conventional philosophical portraits of how scientific knowledge is produced, tested, and stabilized, but also because they matter for the future of the *world*."

References

41.1 N. Oreskes: How earth science has become a social science. In: *Special Issue: Climate and Beyond: The Production of Knowledge about the Earth as a Signpost of Social Change*, ed. by A. Westermann, C. Rohr, Historical Soc. Res. **40** (2015) 246–270

41.2 M. Kleinhans, C. Buskes, H. de Regt: Terra incognita: Explanation and reduction in earth science, Int. Stud. Phil. Sci. **19**(3), 289–317 (2005)

41.3 G.K. Gilbert: *Report on the Geology of the Henry Mountains* (Government Printing Office, Washington 1877)

41.4 G.E. Grant, J.E. O'Connor, M.G. Wolman: A river runs through it: Conceptual models in fluvial geomorphology. In: *Treatise on Geomorphology*, Vol. 9, ed. by J.F. Shroder (Academic, San Diego 2013) pp. 6–21

41.5 W.M. Davis: The systematic description of land forms, Geogr. J. **34**, 300–318 (1909)

41.6 W.M. Davis: The geographical cycle, Geogr. J. **14**, 481–504 (1899)

41.7 I. Kant: Universal natural history and theory of the heavens or essay on the constitution and the mechanical origin of the whole universe according to Newtonian principles. In: *Kant: Natural Science*, ed. by E. Watkins (Cambridge Univ. Press, Cambridge 2012), transl. by O. Reinhardt, originally published in 1755
41.8 P.-S. Laplace: *Exposition du Système du Monde* (Cambridge Univ. Press, Cambridge 2009), originally published in 1796
41.9 N. Oreskes: From scaling to simulation: Changing meanings and ambitions of models in the earth sciences. In: *Science without Laws: Model Systems, Cases, and Exemplary Narratives*, ed. by A. Creager, E. Lunbeck, M.N. Wise (Duke Univ. Press, Durham 2007) pp. 93–124
41.10 A. Daubrée: *Études Synthétiques de Géologie Expérimentale* (Dunod, Paris 1879), in French
41.11 A. Bokulich: How the tiger bush got its stripes: How possibly versus how actually model explanations, Monist **97**(3), 321–338 (2014)
41.12 M.K. Hubbert: Strength of the earth, Bull. Am. Assoc. Petroleum Geol. **29**(11), 1630–1653 (1945)
41.13 R. Bagnold: *The Physics of Blown Sand and Desert Dunes* (Dover, Mineola 2005), originally published in 1941
41.14 D. Green: Modelling geomorphic systems: Scaled physical models. In: *Geomorphological Techniques (Online Edition)*, ed. by S.J. Cook, L.E. Clarke, J.M. Nield (British Society for Geomorphology, London 2014), Chap. 5, Sect. 3
41.15 M. Weisberg: *Simulation and Similarity* (Oxford Univ. Press, Oxford 2013)
41.16 E. Winsberg: Computer simulations in science. In: *The Stanford Encyclopedia of Philosophy*, ed. by E. Zalta http://plato.stanford.edu/archives/sum2015/entries/simulations-science (Summer 2015 Edition)
41.17 M. Kirkby, P. Naden, T. Burt, D. Butcher: *Computer Simulation in Physical Geography* (Wiley, New York 1987)
41.18 G. Tucker: Models. In: *Encyclopedia of Geomorphology*, Vol. 2, ed. by A. Goudie (Routledge, London 2004) pp. 687–691
41.19 G. Tucker, S. Lancaster, N. Gasparini, R. Bras: The channel-hillslope integrated landscape development model (CHILD). In: *Landscape Erosion and Evolution Modeling*, ed. by H. Doe (Kluwer Academic/Plenum, New York 2001)
41.20 T. Coulthard, M. Macklin, M. Kirkby: A cellular model of holocene upland river basin and alluvial fan evolution, Earth Surf. Process. Landf. **27**(3), 268–288 (2002)
41.21 A. Rowan: Modeling geomorphic systems: Glacial. In: *Geomorphological Techniques*, ed. by L.E. Clark, J.M. Nield (British Society for Geomorphology, London 2011), Sect. 5, Chap. 5.6.5 (Online Version)
41.22 CMIP5: World Climate Research Programme's Coupled Model Intercomparison Project, Phase 5 Multi-Model Dataset, http://cmip-pcmdi.llnl.gov/cmip5/ (2011)
41.23 J. Katzav, W. Parker: The future of climate modeling, Clim. Change **132**, 475–487 (2015)
41.24 N. Oreskes: The role of quantitative models in science. In: *Models in Ecosystem Science*, ed. by C. Canham, J. Cole, W. Lauenroth (Princeton UP, Princeton 2003)
41.25 A. Nicholas, T. Quine: Crossing the divide: Representation of channels and processes in reduced-complexity river models at reach and landscape scales, Geomorphology **90**, 318–339 (2007)
41.26 A.B. Murray, C. Paola: A cellular model of braided rivers, Nature **371**, 54–57 (1994)
41.27 T. Coulthard, D. Hicks, M. Van De Wiel: Cellular modeling of river catchments and reaches: Advantages, limitations, and prospects, Geomorphology **90**, 192–207 (2007)
41.28 A.B. Murray: Contrasting the goals, strategies, and predictions associated with simplified numerical models and detailed simulations. In: *Prediction in Geomorphology*, ed. by P. Wilcock, R. Iverson (American Geophysical Union, Washington 2003) pp. 151–165
41.29 B.T. Werner: Complexity in natural landform patterns, Science **284**, 102–104 (1999)
41.30 A. Bokulich: Explanatory models versus predictive models: Reduced complexity modeling in geomorphology, Proc. Eur. Philos. Sci. Assoc.: EPSA11 Perspect. Found. Probl. Philos. Sci., ed. by V. Karakostas, D. Dieks (Springer, Cham 2013)
41.31 A.B. Murray: Reducing model complexity for explanation and prediction, Geomorphology **90**, 178–191 (2007)
41.32 S. Hall: At fault?, Nature **477**, 264–269 (2011)
41.33 J. Wainwright, M. Mulligan: Mind, the gap in landscape evolution modelling, Earth Surf. Process. Landf. **35**, 842–855 (2010)
41.34 T. Kuhn: *The Structure of Scientific Revolutions* (Univ. Chicago Press, Chicago 2012), [1962]
41.35 P. Suppes: Models of data, Proc. Int. Congr. Logic, Methodol. Philos. Sci., ed. by E. Nagel, P. Suppes, A. Tarski (Stanford Univ. Press, Stanford 1962) pp. 251–261
41.36 I. Lakatos: Falsification and the methodology of scientific research programmes, Proc. Int. Colloquium Phil. Sci.: Crit. Growth Knowl., Vol. 4, ed. by I. Lakatos, A. Musgrave (Cambridge Univ. Press, Cambridge 1970), London, 1965
41.37 E. Rykiel: Testing ecological models: The meaning of validation, Ecol. Model. **90**, 229–244 (1996)
41.38 P. Edwards: *Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming* (MIT Press, Cambrindge 2010)
41.39 E. Lloyd: The role of complex empiricism in the debates about satellite data and climate models, Stud. Hist. Philos. Sci. **43**, 390–401 (2012)
41.40 R. Benson, P. Mannion: Multi-variate models are essential for understanding vertebrate diversification in deep time, Biol. Lett. **8**(1), 127–130 (2012)
41.41 A. Mc Gowan, A. Smith (Eds.): *Comparing the Geological and Fossil Records: Implications for Biodiversity Studies* (Geological Society, London 2011), No. 358. The Geological Society Special Publication

41.42 R. Giere: Using models to represent reality. In: *Model-Based Reasoning in Scientific Discovery*, ed. by L. Magnani, N. Nersessian, P. Hagard (Springer, New York 1999)
41.43 S. Norton, F. Suppe: Why atmospheric modeling is good science. In: *Changing the Atmosphere: Expert Knowledge and Environmental Governance*, ed. by C. Miller, P. Edwards (MIT Press, Cambridge 2001) pp. 67–106
41.44 N. Oreskes: Models all the way down (review of Edwards *A Vast Machine*), Metascience **21**, 99–104 (2012)
41.45 K. Beven: *Environmental Modelling: An Uncertain Future? An Introduction to Techniques for Uncertainty Estimation in Environmental Prediction* (Routledge, New York 2009)
41.46 H. Chang: *Inventing Temperature: Measurement and Scientific Progress* (Oxford Univ. Press, Oxford 2004)
41.47 N.K.S. Oreskes: Frechette, K. Belitz: Verification, validation, and confirmation of numerical models in the earth sciences, Science **263**, 641–646 (1994)
41.48 N. Oreskes, K. Belitz: Philosophical issues in model assessment. In: *Model Validation: Perspectives in Hydrological Science*, ed. by M. Anderson, P. Bates (Wiley, West Sussex 2001) pp. 23–42
41.49 N. Oreskes: Evaluation (not validation) of quantitative models, Environ. Health Perspect. **106**(supp. 6), 1453–1460 (1998)
41.50 G. Lauder: On the inference of function from structure. In: *Functional Morphology in Vertebrate Paleontology*, ed. by J. Thomason (Cambridge Univ. Press, Cambridge 1995) pp. 1–18
41.51 J. Hutchinson, M. Garcia: Tyrannosaurus was not a fast runner, Nature **415**, 1018–1021 (2002)
41.52 M. Weisberg: Robustness analysis, Phil. Sci. **73**, 730–742 (2006)
41.53 B. Calcott: Wimsatt and the robustness family: Review of Wimsatt's re-engineering philosophy for limited beings, Biol. Phil. **26**, 281–293 (2011)
41.54 A. Saltelli, K. Chan, M. Scott: *Sensitivity Analysis* (Wiley, West Sussex 2009)
41.55 J. Hutchinson: On the inference of structure using biomechanical modelling and simulation of extinct organisms, Biol. Lett. **8**(1), 115–118 (2012)
41.56 D. Hamby: A review of techniques for parameter sensitivity analysis of environmental models, Environ. Monit. Assess. **32**, 135–154 (1994)
41.57 A. Saltelli, M. Ratto, T. Andres, F. Campologno, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola: *Global Sensitivity Analysis: The Primer* (Wiley, West Sussex 2008)
41.58 R. Snieder, J. Trampert: Inverse problems in geophysics. In: *Wavefield Inversion*, ed. by A. Wirgin (Springer, New York 1999) pp. 119–190
41.59 G. Backus, J. Gilbert: Numerical applications of a formalism for geophysical inverse problems, Geophys. J. R. Astron. Soc. **13**, 247–276 (1967)
41.60 M. Sen, P. Stoffa: Inverse theory, global optimization. In: *Encyclopedia of Solid Earth Geophysics*, Vol. 1, ed. by H. Gupta (Springer, Dordrecht 2011)
41.61 W. Sandham, D. Hamilton: Inverse theory, artificial neural networks. In: *Encyclopedia of Solid Earth Geophysics*, ed. by H. Gupta (Springer, Dordrecht 2011) pp. 618–625
41.62 G. Belot: Down to earth underdetermination, Phil. Phenomenol. Res. XCI **2**, 456–464 (2015)
41.63 T. Miyake: Uncertainty and modeling in seismology. In: *Reasoning in Measurement*, ed. by N. Mössner, A. Nordmann (Taylor Francis, London 2017)
41.64 E. Tal: The Epistemology of Measurement: A Model-Based Account, Ph.D. Thesis (Univ. Toronto, London 2012)
41.65 S. Schumm: *To Interpret the Earth: Ten Ways to be Wrong* (Cambridge UP, Cambridge 1998)
41.66 S. Lane: Numerical modelling: Understanding explanation and prediction in physical geography. In: *Key Methods in Geography*, 2nd edn., ed. by N. Clifford, S. French, G. Valentine (Sage, Los Angeles 2010) pp. 274–298, 2003
41.67 J. O'Reilly, K. Brysse, M. Oppenheimer, N. Oreskes: Characterizing uncertainty in expert assessments: Ozone depletion and the west antarctic ice sheet, WIREs Clim. Change **2**(5), 728–743 (2011)
41.68 W. Parker: Predicting weather and climate: Uncertainty, ensembles, and climate, Stud. Hist. Phil. Mod. Phys. **41**, 263–272 (2010)
41.69 R. Frigg, S. Bradley, H. Du, L. Smith: Laplace's demon and the adventures of his apprentices, Phil. Sci. **81**, 31–59 (2014)
41.70 E.L. Thompson: Modelling North Atlantic Storms in a Changing Climate, Ph.D. Thesis (Imperial College, London 2013)
41.71 N. Odoni, S. Lane: The significance of models in geomorphology: From concepts to experiments. In: *The SAGE Handbook of Geomorphology*, ed. by K. Gregory, A. Goudie (SAGE, London 2011)
41.72 G. Sella, S. Stein, T. Dixon, M. Craymer, T. James, S. Mazzotti, R. Dokka: Observation of glacial isostatic adjustment in stable North America with GPS, Geophys. Res. Lett. **34**(2), 1–6 (2007), L02306
41.73 T. Chamberlin: The method of multiple working hypotheses, Science **15**(366), 92–96 (1890)
41.74 R. Laudan: The method of multiple working hypotheses and the development of plate tectonic theory. In: *Scientific Discovery: Case Studies*, Boston Studies in the Philosophy of Science, Vol. 60, ed. by T. Nickles (Springer, Dordrecht 1980) pp. 331–343
41.75 M. Richards: The cretaceous-tertiary mass extinction: What really killed the dinosaurs?, http://hmnh.harvard.edu/file/366291 (2015) Lecture given on February 3rd, 2015 at the Harvard Museum of Natural History
41.76 C. Cleland: Prediction and explanation in historical natural science, Br. J. Phil. Sci. **62**, 551–582 (2011)
41.77 A.B. Murray: Cause and effect in geomorphic systems: Complex systems perspectives, Geomorphology **214**, 1–9 (2014)
41.78 C. Paola, K. Straub, D. Mohrig, L. Reinhardt: The unreasonable effectiveness of stratigraphic and geomorphic experiments, Earth Sci. Rev. **97**(1–4), 1–43 (2009)

41.79 P. Duhem: *The Aim and Structure of Physical Theory* (Princeton Univ. Press, Princeton 1954), trans. P. Weiner, 1906
41.80 K. Stanford: Underdetermination of scientific theory. In: *Stanford Encyclopedia of Philosophy*, ed. by N. Edward, E. Zalta http://plato.stanford.edu/archives/win2013/entries/scientific-underdetermination (Winter 2013 Edition)
41.81 K. Beven: Prophecy, reality and uncertainty in distributed hydrological modelling, Adv. Water Resour. **16**(1), 41–51 (1993)
41.82 K. Beven, J. Freer: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, J. Hydrol. **249**(1–4), 11–29 (2001)
41.83 K. Beven: Equifinality and uncertainty in geomorphological modeling, Proc. 27th Binghampton symp. geomorphol.: Sci. Nat. Geomorphol., ed. by B. Rhoads, C. Thorn (Wiley, Hoboken 1996) pp. 289–313
41.84 K. Beven, A. Binley: GLUE: 20 years on, Hydrol. Process. **28**(24), 5897–5918 (2014)
41.85 J.P.C. Kleijnen: Experimental design for sensitivity analysis, optimization, and validation of simulation models. In: *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*, ed. by J. Banks (Wiley, New York 1998) pp. 173–223
41.86 N. Odoni: Exploring Equifinality in a Landscape Evolution Model, Ph.D. Thesis (Univ. Southhampton, School of Geography, Southhampton 2007)
41.87 R.L. Slingerland, G. Tucker: Erosional dynamics, flexural isostasy, and long-lived escarpments, J. Geophys. Res. **99**, 229–243 (1994)
41.88 R. Knutti, J. Sedlácek: Robustness and uncertainties in the new CMIP5 climate model projections, Nat. Clim. Change **3**, 369–373 (2013)
41.89 E. Lloyd: Confirmation and robustness of climate models, Phil. Sci. **77**, 971–984 (2010)
41.90 W. Parker: When climate models agree: The significance of robust model predictions, Phil. Sci. **78**, 579–600 (2011)
41.91 W. Parker: Ensemble modeling, uncertainty, and robust predictions, WIREs Clim. Change **4**, 213–223 (2013)
41.92 J. Lenhard, E. Winsberg: Holism, entrenchment, and the future of climate model pluralism, Stud. Hist. Phil. Mod. Phys. **41**, 253–262 (2010)
41.93 D. Masson, R. Knutti: Climate model genealogy, Geophys. Res. Lett. **38**, L08703 (2011)
41.94 R. Knutti, R. Furrer, C. Tebaldi, J. Cermak, G. Meehl: Challenges in combining projections from multiple climate models, J. Clim. **23**(10), 2739–2758 (2010)
41.95 A. Bokulich: How scientific models can explain, Synthese **180**(1), 33–45 (2011)
41.96 A. Bokulich: Models and explanation. In: *Handbook of Model-Based Science*, ed. by L. Magnani, T. Bertolotti (Springer, Dordrecht 2016)