6

Why Believe a Computer? Models, Measures, and Meaning in the Natural World

Naomi Oreskes

Doubt is the essence of understanding.—Richard Feynman

For at least three centuries, philosophers have tried to determine what separates science from other forms of human knowledge. In the seventeenth century, Sir Francis Bacon introduced the concept of a "crucial experiment"—a test that would uniquely prove or disprove a scientific idea. By actively challenging our beliefs through crucial experiments and critical observations, Bacon held, we could increase our understanding of the natural world and so improve our lot.[1] History has long since demonstrated that a single test is rarely if ever sufficient to convince anyone of anything, and many experiments appear crucial only in retrospect. Still the idea of crucial experiments has held on, because it speaks to a broader point: Testing is the heart of science. Although there is no foolproof way to define science, testability is the most commonly cited demarcation criterion between scientific theories and other forms of human explanatory effort.[2]

Can computer models be tested? In recent years, there has been an explosive increase in the use of computer simulation models in fields as diverse as economics, aeronautics, cosmology, epidemiology, and forest ecology. Geology is no exception. From geochemistry to hydrology, paleontology to mantle dynamics, computer simulation models are now a standard part of the tool kit of the earth sciences. One of the driving forces behind the increased use of computer models in the earth sciences is their applicability

to systems that are too large, too c
means. The example familiar to m
scientists believe that carbon diox
ties is producing an enhanced g
Earth's average temperature will i
effects will be requires powerful co
complex set of differential equatio
Other examples include models u
nuclear waste repository sites, to e
industrial plants, or to determine
fisheries. These are complex syst
easily recreated in a laboratory exp

In many cases, the predictions
ered as a basis for public policy
agencies may be required by law to
tion has led to the demand for "ve
Claims about model verification a
published scientific literature. Are
model be proved true or false? How

THE PROBLEM

All scientific theory testing involve
several possible configurations of
observed results. Therefore, any er
ory may also be consistent with
many scientists accept the view, fi
that scientific theories can be pro
theories can be *falsified*, but not
with the predictions of a theory,
must be modified; but if the predi
our problems have yet to be encou
corner. This conclusion holds for a
by computers or otherwise.

But the issue is more complica
to cohere with the natural world, t
the fault lies. The problem may lie i
in a faulty piece of equipment, a
background assumption. A famou
the history of astronomy. Parallax
when viewed from different positio
ent backdrops. (Imagine a girl sta
boat, heading north. First you see
back and see her against the sou

to systems that are too large, too complex, or too far away to study by other means. The example familiar to most readers is global climate change. Most scientists believe that carbon dioxide in the atmosphere from human activities is producing an enhanced greenhouse effect and, in the future, the Earth's average temperature will increase as a result. To predict what these effects will be requires powerful computers that can obtain the solutions to a complex set of differential equations involving large quantities of input data. Other examples include models used to predict the behavior of proposed nuclear waste repository sites, to estimate air pollution emission levels from industrial plants, or to determine the circulation of ocean currents affecting fisheries. These are complex systems that cannot be observed simply nor easily recreated in a laboratory experiment.

In many cases, the predictions generated by computer models are considered as a basis for public policy decisions, and government regulators and agencies may be required by law to establish their trustworthiness. This situation has led to the demand for "verification" or "validation" of these models.[3] Claims about model verification and validation are now routinely found in published scientific literature. Are these claims legitimate? Can a computer model be proved true or false? How can we tell when to believe a computer?

THE PROBLEM OF VERIFICATION

All scientific theory testing involves a fundamental ambiguity: There may be several possible configurations of nature that could produce a given set of observed results. Therefore, any empirical data we collect in support of a theory may also be consistent with alternative explanations. For this reason, many scientists accept the view, first developed by Karl Popper in the 1930s, that scientific theories can be proved false but not true. In Popper's terms, theories can be *falsified*, but not *verified*. If empirical data are inconsistent with the predictions of a theory, then something is amiss and the theory must be modified; but if the predictions come true, it may merely mean that our problems have yet to be encountered. Refutation may be just around the corner. This conclusion holds for all scientific knowledge, whether generated by computers or otherwise.

But the issue is more complicated than Popper allowed. If a prediction fails to cohere with the natural world, there is often no simple way to know where the fault lies. The problem may lie in the theory being tested, but it may also lie in a faulty piece of equipment, a bug in a computer program, or a mistaken background assumption. A famous example is the case of "stellar parallax" in the history of astronomy. Parallax is the changing appearance of an object when viewed from different positions, so that the object appears against different backdrops. (Imagine a girl standing on the beach while you sail past in a boat, heading north. First you see her against the northern sky, later you look back and see her against the southern sky.) When Copernicus proposed his heliocentric model of the universe—that the Earth was moving rather than the heavens—astronomers realized that the phenomenon of parallax offered a means to test the theory. If the Earth moved while the sun and stars stayed fixed, then the apparent position of any particular star should change during the course of the year. The star would be seen first against one backdrop, then against another. But when sixteenth-century astronomers searched for stellar parallax, they found none—and they rejected Copernicus's theory. Implicitly, they assumed that the Earth's orbit was large relative to the distance to the stars and therefore the parallax effect would be significant and their telescopes would be able to detect it. These assumptions turned out to be wrong. Because of the enormous distance to the stars, the parallax effect turns out to be very, very small, and it was not until the twentieth century that telescopes became powerful enough to detect it.[4]

Another example comes from the history of geology. One hundred years ago, the great British physicist Lord Kelvin argued that the Earth could not possibly be as old as geologists thought it was. On the basis of the concept of uniformitarianism—the assumption that presently observable geological processes are representative of Earth's history in general—nineteenth-century geologists concluded that the Earth was probably a few billion years old. Given observable rates of erosion and deposition, it would take that long to produce the known rock record. But when Kelvin calculated the time required for a molten body the size of the Earth to cool to its present temperature, he obtained a maximum of 98 million years and promptly declared the entire science of geology invalid. Any conceptual scheme that implied a billion-year-old Earth was fundamentally flawed, he announced. Pursuing the same logic, he dismissed Charles Darwin's theory of natural selection on the grounds of inadequate time for it to operate.[5] For several decades, Kelvin's result held sway and evolutionists were in nearly full retreat, until the discovery of radioactivity proved Kelvin wrong. For his calculation, Kelvin assumed that the Earth had no additional source of heat. In fact, the decay of radioactive elements within the Earth generated heat, and so the Earth had cooled far more slowly than simple physical principles would otherwise suggest.

In hindsight, it is easy to see where others have gone wrong: Astronomers thought their instruments were better than they were; Kelvin thought his knowledge more complete than it was. It is harder to see the flaws in our own reasoning. (If we could see them, presumably we would correct them.) When computer models are involved, it can be more difficult still, because the systems being modeled are very complex and the embedded assumptions can be very hard to see. How *do* we test computer models?

In the earth sciences, models are often tested by "history matching"—by seeing how well the model matches historical data. Hydrologists, for example, may test a groundwater flow model by running it backward in time and comparing the results with published records of water well levels. Climate models can be tested against weather records. If a model accurately reproduces histor-

ical data, modelers commonly clai
can therefore be used to predict
Konikow and his coworkers at the
many "validated" models in hydro
future.[6] Why? The most common re
changes in the modeled system. A r
dry decade, for example, or an unus
change their behavior in ways that
way to predict such occurrences, sc
changes did not occur at all. In effec
eling as though the systems were st
the systems are static—no earth sc
truly static. Nevertheless, scientists
often turns out to be wrong.

These examples give us ground
best work, but they also provide gr
all three cases the mistaken backg
nized. The process of scientific rese
old questions and build better eq
wrongly believing in a false theory c
cases it is. But it is not academic wh
public policy. Then, the veracity o
health and safety and the future of c

THE PARADOX O
REPRESENTATION

The development of fast, inexpensi
sciences because they enable us
remain intractable. Complex earth
increased carbon dioxide, the trans
or the workings of a forest ecosyst
scientific methods. An ecosystem c
Earth's climate cannot be the site
posed adding carbon dioxide to th
experiment would have been rejec
tion models provide an ethical and
natural systems.

A problem arises, however, w
complex models against the natura
of a theory (or model) fail to come
this is not what scientists do. As ph
emphasized, scientists normally m
accommodate observational dis

ical data, modelers commonly claim that the model has been "validated" and can therefore be used to predict the future. However, hydrologist Leonard Konikow and his coworkers at the U.S. Geological Survey have shown that many "validated" models in hydrology fare poorly when extended into the future.[6] Why? The most common reason is that modelers fail to anticipate later changes in the modeled system. A region may experience a particularly wet or dry decade, for example, or an unusual confluence of rare events. Humans may change their behavior in ways that affect the natural system. Lacking a reliable way to predict such occurrences, scientists often build their models as if these changes did not occur at all. In effect, scientists treat the systems they are modeling as though the systems were static. This is not to say that modelers *believe* the systems are static—no earth scientist could imagine any earth system as truly static. Nevertheless, scientists often embed stasis into their models, and it often turns out to be wrong.

These examples give us grounds for humility in our assessment of our own best work, but they also provide grounds for long-range optimism because in all three cases the mistaken background assumptions were eventually recognized. The process of scientific research can be self-correcting as we reexamine old questions and build better equipment and instrumentation. The risk of wrongly believing in a false theory or model may seem academic, and in many cases it is. But it is not academic when scientific knowledge provides a basis for public policy. Then, the veracity of our theories can be a matter of human health and safety and the future of our natural environment.

THE PARADOX OF COMPLEX MODELS: REPRESENTATION VERSUS REFUTABILITY

The development of fast, inexpensive computers has been a boon to the earth sciences because they enable us to study problems that might otherwise remain intractable. Complex earth systems—such as the climate response to increased carbon dioxide, the transport of contaminants through groundwater, or the workings of a forest ecosystem—are difficult to address by traditional scientific methods. An ecosystem cannot be brought into the laboratory; the Earth's climate cannot be the site of controlled experiments. If you had proposed adding carbon dioxide to the Earth's atmosphere to test its effects, the experiment would have been rejected on ethical grounds. Numerical simulation models provide an ethical and pragmatic means to grapple with complex natural systems.

A problem arises, however, when we attempt to test the predictions of complex models against the natural world. In Popper's view, if the predictions of a theory (or model) fail to come true, then the theory must be rejected. But this is not what scientists do. As philosophers Imre Lakatos and Thomas Kuhn emphasized, scientists normally modify their theories in various small ways to accommodate observational discrepancies.[7] Recognizing this, the great chemist and philosopher Pierre Duhem concluded long ago that theories cannot be proved false any more than they can be proved true. In his 1906 essay, "Physical Theory and Experiment," Duhem argued that it is always possible to modify a theory in some manner so as to salvage it in the face of recalcitrant empirical evidence.[8] There is no simple criterion that tells us when we should modify an existing theory and when we should throw it out altogether.

Duhem believed that very simple phenomenological laws, deduced directly from empirical observation, might be refutable. Implicitly, he suggested that the more complex a theory and the instrumentation upon which its empirical support relies, the greater the difficulty of testing it, because the greater the number of parts that can be modified. This suggestion has two consequences. As systems become more complex, it becomes increasingly difficult to determine which part of the system is at fault when something goes wrong. And it becomes increasingly easy to modify a system that is failing in the face of negative evidence.

The implication for computer models is evident. The more sophisticated they become, the more difficult it is to refute them. One can save a complex model from refutation by small adjustments to its components. The constraints provided by laboratory experiments, field study, or theoretical considerations are hardly ever sufficient to define the system completely. Models tested against empirical data can therefore be "fine-tuned" to fit those data. Indeed, this is normal practice: Given a set of observational data, scientists will work on a model until they achieve a fit, a process referred to as calibration. A model that did not match available data would be rejected as false, but a model calibrated to fit available data may also be false.

Given this situation, one can see a justification for the traditional value of simplicity in science: Simpler systems are easier to test. One might therefore strive for simplicity in computer models, and some modelers do. For example, scientists working on artificial intelligence may seek the simplest possible model that can perform a task, such as pilot an aircraft or recognize speech. But these scientists are not trying to create a realistic representation of the human brain in all its complexity. Rather, they are building a system to *do* something. Their goals may be described as *functional* rather than *representational.* Modelers who attempt to represent complex natural systems often eschew simplification as fallacious.

A good example is the development of general circulation models (GCMs) for predicting global climate change. GCMs are computer models that attempt to represent the circulation of the Earth's atmosphere toward the goal of understanding climate patterns and climate change. Early GCMs were conspicuously oversimplified. Computational limitations forced scientists to omit many factors, notably the ability of the world's oceans to take in or give off heat. Simplification was a necessity, but not a virtue. With advances in computation, climate modelers have been able to increase the complexity of their models incrementally. First the oceans were added to GCMs in a static way—as if they

were blankets of water that abso
More powerful computers have l
examine the effects of ocean cur
understanding the greenhouse e
incomplete, nearly all scientists w
in the right direction. Because the
increased complexity in the model
imation to reality.[9]

The history of GCMs illustrat
tems, simplification often means l
known or suspected processes. In
elers by the limits of computation
some computer models took long
were simulating. But the greatly i
puter technology has made it p
complexity than most scientists i
simplicity in model testing is no
which is interpreted as evidence o
presumed realism of the model. T
ambitious it is—the more this is t
harder it is to refute. So we face a
representation of a complex eart
another way, the better the model
harder it is for others to evaluate t
sentation and refutability.

HETEROGENEITY A
OF EM

Complexity is often discussed in
or feedbacks represented in a mo
terms of the *character* of the em
entists often think of empirical
with definite determinable value
salinity of seawater, or the densit
average salt content of seawater
live on land and coastal process
can readily reach. Deltas add fre
the ocean; lagoons are sites of e
the open ocean to measure sali
that modify salinity elsewhere, s
there is a definite value for the s

were blankets of water that absorbed and radiated heat, but did not move. More powerful computers have lately allowed climate modelers to begin to examine the effects of ocean currents, which may turn out to be critical to understanding the greenhouse effect. Although the latest models are still incomplete, nearly all scientists would say that the addition of oceans is a step in the right direction. Because the Earth's climate is a highly complex system, increased complexity in the models is interpreted as evidence of closer approximation to reality.[9]

The history of GCMs illustrates a general point: For modeling earth systems, simplification often means leaving out available information or ignoring known or suspected processes. In the past, this practice was forced upon modelers by the limits of computational power; it used to be a joke in geology that some computer models took longer to run than the geological processes they were simulating. But the greatly increased speed and decreased cost of computer technology has made it possible to construct models of far greater complexity than most scientists imagined only a few years ago. The benefit of simplicity in model testing is now subordinated to the value of complexity, which is interpreted as evidence of the increased sophistication and therefore presumed realism of the model. The greater the scope of the model—the more ambitious it is—the more this is the case. But the more complex a model, the harder it is to refute. So we face a paradox: The closer a model comes to a full representation of a complex earth system, the harder it is to evaluate it. Put another way, the better the model is from the point of view of the modeler, the harder it is for others to evaluate the model. There is a trade-off between representation and refutability.

HETEROGENEITY AND THE (IN)ACCESSIBILITY OF EMPIRICAL INPUT

Complexity is often discussed in terms of the number of different processes or feedbacks represented in a model, but complexity can also be discussed in terms of the *character* of the empirical input parameters required by it. Scientists often think of empirical parameters in terms of physical properties with definite determinable values, like the atomic weight of an element, the salinity of seawater, or the density of the continental crust. Determining the average salt content of seawater is actually a challenging task, because we live on land and coastal processes affect the salinity of the seawater that we can readily reach. Deltas add fresh water that decreases the local salinity of the ocean; lagoons are sites of evaporation that may increase it. If we go to the open ocean to measure salinity, we still have to account for processes that modify salinity elsewhere, such as melting of polar ice. Nevertheless, there is a definite value for the saltiness of the sea, even if we have to work hard to determine it. But in many models of natural systems, the required empirical input may not have definite values and may not be fully determinable. This point requires explanation.

Many models of natural systems are based on continuum theory, in which a material with heterogeneous parts is treated as if it were a single homogeneous entity: the continuum. A familiar example is a household sponge. At some scale the "sponge" does not exist, it is a composite of solid material and the holes within it. The absorbency of sponges is what philosophers call an "emergent" property: It *emerges* from the interrelation of the solid, the holes, and the fluids with which they are in contact. Or consider another example: raisin bread. We all know what raisin bread is and we know what it tastes like. But cut it into small enough pieces and it no longer exists. Instead, we have a pile of raisins and bits of plain bread. The taste, smell, and feel of raisin bread—indeed, the very existence of raisin bread—are all emergent properties. Modeling based on continuum theory relies on the existence of emergent properties in nature; emergent properties permit scientists to describe complex, variable materials in a simple and tractable way.

An important example in earth science is the concept of permeability—the measure of how readily fluids will flow through a rock. Although we think of rocks as solid, really they are not. All rocks contain holes and fractures that can allow fluids to flow through them. The greater the number, size, and interconnectedness of these holes and fractures, the higher the permeability of the rock. Permeability is a crucial measure in siting nuclear waste repositories and toxic disposal sites, because we rely on the surrounding rocks to prevent toxins from escaping the site. On a microscopic level, however, permeability does not exist. On a microscopic level, rocks are complex aggregates of solid mineral grains with no permeability, and the empty spaces between them that have infinite permeability. In principle, the "true" permeability of a rock might be determined by measuring every mineral grain and the size and shape of every space around it. But to describe this complexity in realistic detail would be a hopeless task. A geologist could spend a lifetime describing the patterns of holes in a single rock outcrop.

Continuum theory provides a way to simplify this problem by treating the rock as a continuous material with a single definite property (namely, permeability) that represents the macroscopic behavior of the overall rock mass. But in making this simplification, fine-scale information is lost. We are deliberately ignoring the details. Is the devil in the details? In most cases, probably not, but there is no way to know until we use the model.

In addition to the loss of information at the subcontinuum scale, there is another problem: The emergent properties of the continuum may themselves be heterogeneous. Some parts of a sponge may be more holey than others; some slices of raisin bread may have more raisins; and some parts of a rock layer may be more permeable. Continuum equations traditionally involve the assignment of a single average value (or sometimes simply varying values) for complexly varying emergent properties. When we take an average value, we lose information about highs and lows. And this may matter a lot, because in many geological environments, the flow of fluids is controlled, not by the average properties of the bulk of the rock mass, but by the local presence of zones of very high permeability. In the Earth, heterogeneity matters. We need to know the variability of rocks surrounding a nuclear waste repository, because local zones of above-average permeability may allow contaminants to escape into the hydrosphere and biosphere. We need to know the extremes of weather that global warming may cause, because our houses are built to withstand ordinary weather; it is the extreme events that do damage to life and property. And we need to know the extremes of possible earthquake magnitudes, because one giant quake can do more damage than scores of small ones. From a human perspective, extremes are often more important than averages.

A response to the problem of heterogeneity is to collect spatially or temporally distributed data on heterogeneous properties. For example, we could study the spatial variability of permeability surrounding a proposed nuclear waste repository, and some scientists have done this. But this approach can only take us so far, because exhaustive local sampling presents the risk of modifying the bulk properties we are trying to measure. The insertion of closely spaced drill holes into a rock unit to determine its permeability may change that permeability. A full description of a heterogeneous physical system cannot be obtained without changing or destroying it.

Because of the heterogeneity of the natural world, uncertainty is inherent in the very nature of the systems we are studying. The required empirical input parameters are fundamentally inaccessible, not because our instruments are not sophisticated enough, our budgets not big enough, or we not clever enough, but because *the data do not exist in the form required by the equations at the scale to which we have access.* And when a system is poorly known, there is always uncertainty over whether our samples are representative. One is caught in a circular argument: To judge whether or not our samples are representative, we need to know the characteristics of the system, but if we knew the characteristics of the system, we would not need to sample it!

The heterogeneity of the natural world ensures that any model that strives for realistic representation must contain a large quantity of empirical input. This returns us to the complexity paradox: The more data we have, the more likely our model is to capture nature's diversity and richness. However, the more data we have, the more complex the model becomes and the less likely it is to be unique—another model of equal complexity but with a different mix of empirical input could have produced the same result.

ACCESS TO MODEL PREDICTIONS

Acknowledging these caveats, one may still argue that once a model is built, using the best available information, it can be tested like any other scientific theory by comparing its output with the natural world. Some models perform extremely well when this type of comparison is done. For example, computer models have been developed to predict the positions of the planets, stars, and other celestial bodies, and these predictions can easily be compared with astronomical observations. As the 1996 collision of Comet Shoemaker-Levy with Jupiter shows, these models are successful: Planetary scientists predicted the location and timing of this collision more than a year in advance.

But models in celestial mechanics are the exception that proves the rule: They represent relatively simple physical systems in which the operative forces can be described by a small number of equations involving variables that can be determined with a high degree of precision. Indeed, the equations involved were solved long before the advent of computers. The programs that do this for us today are a matter of convenience, not necessity. Furthermore, the events predicted by celestial mechanics are readily accessible: They take place all the time and in a repetitive way. Any model that did not accurately locate the planets would be subject to modification in very short order. As Karl Popper once explained, astronomical predictions "are possible only because our solar system is a stationary and repetitive system, and this is so because of the accident that it is isolated from the influence of other mechanical systems by immense regions of empty space and is therefore relatively free of interference from outside. Contrary to popular belief the analysis of such repetitive systems is not typical of natural science. These repetitive systems are special cases where scientific prediction becomes particularly impressive—but that is all."[10]

Most of the world is not like the solar system. The systems earth scientists study are not repetitive; they are not isolated from the influence of other systems; and they are not free from outside interference. Although the planets may seem far away—and physically, of course, they are—on a scientific level they are unusually accessible. Ironically, the predictions we make about the Earth we stand on are far less accessible.

For example, global circulation models must predict the Earth's climate over the next several decades. With adequate funding for scientific research, these predictions can be compared with empirical evidence in the years to come. But what exactly is the relevant empirical evidence? A climate model may predict global average temperature, but how does one *measure* the average temperature of the world? The data that are needed to test the model are *themselves modeled*—various point measurements around the globe must be synthesized into a single number, and scientists have argued over how best to achieve that synthesis. Refined GCMs may produce averages for particular regions of the globe—what modelers call model cells—but temperature is not measured in a model cell any more than it is measured in the world at large; temperature is measured at a point. The gap that exists between empirical input and model parameters is mirrored by a gap between model output and the data that could potentially confirm it.

A second problem, which is perhaps more important, is that policy decisions need to be made now. Many earth scientists and environmentalists argue

that we cannot wait to confirm th
be too late to undo the damage
outcome unless action is taken
advance whether the predictions
be too late to use that informatic
be offered as a basis for decision-

Other policy-relevant models
science may no longer exist. F
nuclear waste repository sites m
thousand to ten thousand years.
tion about what we mean by pre
the cultural institution we call sci
and if it does, it will almost certai
us. More important is that our w
even unrecognizable as science,
lectual tradition were to exist tha
ence, it is unlikely that its practit
the predictions we left for them t
up a mechanism to confirm th
future. Even if we could set up s
would be superfluous. Decisions
about a future to which we have
the future. How good a surrogate

An obvious solution to this p
ical scales and shorter time fra
example from hydrologic model
present observations is no guar
future, because natural systems a
ways. As the philosopher Sir Al
Hume's famous skepticism abo
whether the future will resemble
the assumption that it will—bu
philosopher Heraclitus put it, we

No model output ever matc
always whether the match is "go
scientific and one social. On a s
depends on scale. Consider this
two sisters are the same height, s
same to the nearest inch. With a f
in matters of height, such fine
model, however, scale can be vit
cantly affect model fit under th
data are available may generate
larger scales. A model that work

that we cannot wait to confirm the results of GCMs, because, if we wait, it may be too late to undo the damage wrought. When a model predicts a negative outcome unless action is taken, a dilemma develops: We cannot know in advance whether the predictions are correct, but, if we wait to find out, it may be too late to use that information. The result is that testable predictions may be offered as a basis for decision-making in advance of any actual test results.

Other policy-relevant models make predictions on time frames over which science may no longer exist. For example, models of proposed high-level nuclear waste repository sites must predict repository behavior for the next thousand to ten thousand years. Models of this kind raise a fundamental question about what we mean by prediction. If the history of science is any guide, the cultural institution we call science may not exist ten thousand years hence; and if it does, it will almost certainly be in a form that will be unrecognizable to us. More important is that our work will most likely be unintelligible, perhaps even unrecognizable as science, to our successors. Therefore, even if an intellectual tradition were to exist that identified itself as the successor to our science, it is unlikely that its practitioners would be able to make much sense of the predictions we left for them to confirm or deny. There is no way for us to set up a mechanism to confirm the predictions of models 10,000 years in the future. Even if we could set up such tests, by the time we got the results they would be superfluous. Decisions must be made now on the basis of predictions about a future to which we have no access. Models are a surrogate for access to the future. How good a surrogate they are is an open question.

An obvious solution to this problem is to test models on smaller geographical scales and shorter time frames, and earth scientists do this. But, as the example from hydrologic models shows, a match between model results and present observations is no guarantee that the model will do as well in the future, because natural systems are dynamic and may change in unanticipated ways. As the philosopher Sir Alfred J. Ayer argued in a discussion of David Hume's famous skepticism about prediction, the question is not so much *whether* the future will resemble the past—because all science is premised on the assumption that it will—but *how* it will resemble it.[11] Or, as the Greek philosopher Heraclitus put it, we cannot step twice in the same river.

No model output ever matches the natural world exactly. The question is always whether the match is "good enough"—and this raises two issues, one scientific and one social. On a scientific level, evaluation of model fit always depends on scale. Consider this example of scale dependency: If I state that two sisters are the same height, say, five foot three inches, I implicitly mean the same to the nearest inch. With a finer ruler, I could make a finer distinction; but in matters of height, such fine distinctions rarely worry us. In a computer model, however, scale can be vitally important. Small errors that do not significantly affect model fit under the time frame or geographical scale for which data are available may generate large discrepancies when extrapolated over larger scales. A model that works on a small scale may fail on a large scale. A model that seems to be working at present may go wrong in the future. This fact highlights a key difference between natural and engineering systems: Modern engineers start with computer models, but then they build prototypes and pilot plants. If in the building stage a problem arises that was not predicted by the computer model, it can be remedied. Modern aeronautical designs are developed on computers, but no one ever buys a ticket on a commercial jet before a prototype has been flown for many hours.

The social issue at stake when judging model fit is perhaps the most important of all. To ask whether a fit is adequate is to make a judgment call. If we conclude that a proposed nuclear waste repository will not release "significant" radionuclides into the biosphere, we have implicitly invoked a notion of "significance" that rests in turn on prior judgments of socially, politically, and ethically acceptable levels of environmental impact. The same is true for virtually all models that deal with environmental issues. The literature of numerical modeling is filled with words like "adequate," "acceptable," "reasonable," and "significant," yet rarely is there any attempt to define the basis for these judgments. One reason is obvious: Scientists do not like to talk in these terms. Science, at least in our idealized sense of it, is supposed to be about facts, not judgments. Scientists are not trained to talk about questions of moral value. Yet, at root, all environmental issues involve questions of moral value, and judgment is the basis upon which we stake our moral claims. Computer modeling, like all quantification in science, appeals to our sense of objectivity, but even the most mathematically and computationally sophisticated model will not absolve us of the need for judgment, nor of the need to justify our judgments in human terms.

COMPLEXITY IS THE STRENGTH AND WEAKNESS OF NUMERICAL MODELS

The power of contemporary computers allows earth scientists to examine complex processes and systems—the lithosphere, the oceans, the atmosphere, the flow of groundwater—that are not readily studied by other means. These systems sustain life on Earth; and if we fail to understand them, we endanger not merely the quality of our lives, but ultimately life itself. Theirs is no small contribution, for computer models have helped us gain a better understanding of the Earth's complex life-supporting processes. The ability to represent such systems is the obvious strength of models. Their weakness is that these complex models are nonunique, their predictions may be in error, and the scale of their predictions can make them difficult if not impossible to test. Furthermore—and this is perhaps the most important point of all—we cannot separate our judgment of the adequacy of our models from our judgment of the social and moral consequences of the effects we are trying to model.

When scientists construct laboratory experiments to test a theory, they abstract a portion of the physical world in order to ask certain kinds of ques-

tions. Computer modelers do t
abstractions of the natural world t
Not surprising, modelers sometir
The laboratory scientist, in const
complexity of the natural world fo
setup. How serious this sacrifice
not be predetermined. Laborator
the history of science, but they ha
no doubt prove true for numerica
plexity can be preserved but at th
ity. Like the sacrifice of complexit
tation, the significance of the lo
experience tells us what kinds of
tory; future experience will have t

The traditional point of cont
natural world is the experimental
point of contact is the *input*. In p
of large quantities of informatio
practice, the availability of data fr
advances in theory and computat
outsider to judge the quantity and
the impression of being grounded
retical edifice. This reality may b
with computer models. If the ult
grounding in empirical phenome
empirical basis of a model is d
greater rigor in the evaluation of e
illusion that things are better kno

SO WH

All models depend on data; and
direct encounters with the Earth.
cated, the need for data obtained
increase rather than decrease. Fie
are to understand the distributio
fluids in the crust, the patterns
regional scale, or the weather pa
spheric pollutants.

Field evidence also provide
model predictions. No sensible pe
the threat of global warming, bu
spend large sums of money solvi
can help to determine the exten

tions. Computer modelers do the same thing. Numerical simulations are abstractions of the natural world that allow us to ask certain kinds of questions. Not surprising, modelers sometimes call their work "numerical experiments." The laboratory scientist, in constructing a physical experiment, sacrifices the complexity of the natural world for the tractability of a simplified experimental setup. How serious this sacrifice is depends on the problem at hand and cannot be predetermined. Laboratory experiments have provided crucial tools in the history of science, but they have also misled people. The same duality will no doubt prove true for numerical experiments. In numerical modeling, complexity can be preserved but at the cost of physical and intellectual accessibility. Like the sacrifice of complexity inherent in traditional forms of experimentation, the significance of the loss of accessibility is difficult to judge. Past experience tells us what kinds of problems are tackled effectively in the laboratory; future experience will have to do the same for computer modeling.

The traditional point of contact between a laboratory experiment and the natural world is the experimental outcome; but in numerical models, the main point of contact is the *input*. In principle, models allow for input and synthesis of large quantities of information—and this is a very great strength—but in practice, the availability of data from the natural world has not kept pace with advances in theory and computation. And it can be exceedingly difficult for an outsider to judge the quantity and quality of data in a model. A model may give the impression of being grounded in empirical evidence, yet be largely a theoretical edifice. This reality may be what makes some people uncomfortable with computer models. If the ultimate strength of scientific knowledge is its grounding in empirical phenomena, then we should be uncomfortable if the empirical basis of a model is difficult to ascertain. Modeling may lead to greater rigor in the evaluation of earth processes, but it may also propagate the illusion that things are better known than they really are.

SO WHAT DO WE DO?

All models depend on data; and in the earth sciences, data are collected by direct encounters with the Earth. As our models become increasingly sophisticated, the need for data obtained through human encounters with nature will increase rather than decrease. Field-based empirical evidence is essential if we are to understand the distribution of earth materials that control the flow of fluids in the crust, the patterns of ocean circulation that affect climate on a regional scale, or the weather patterns that determine the dispersal of atmospheric pollutants.

Field evidence also provides a complement to reliance on long-term model predictions. No sensible person would wish to court disaster by ignoring the threat of global warming, but neither would any sensible society wish to spend large sums of money solving a problem that does not exist. Geologists can help to determine the extent of global change by studying its effects in presently observable geological processes. Temperature-sensitive natural geological processes like the calving of glacial ice or the evaporation rates in inland lakes can provide direct evidence of climate change. Study of radioactive elements in natural geological environments can teach us about the likely behavior of these and similar elements in nuclear waste repositories without having to wait ten thousand years.

Perhaps the most important perspective geology and geologists can contribute to environmental debates is historical. Uncertainty surrounds environmental questions because there is much about the Earth that we do not understand. In particular, we do not know whether the changes we are currently witnessing are part of the Earth's natural cycle, the result of human activities, or both. The Earth's climate has cooled and warmed many times in geological history: How can we know for sure whether the changes of the twentieth century are caused by human activities or are merely the latest cycle in the Earth's long record of change? Geology can help to provide an answer, because the rock record is the record of Earth's history. Humans have been keeping weather records for a little more than a century and studying the migration of fluids and elements in the crust for a few decades, but rocks have been recording evidence of these processes for nearly four billion years. Geologists can use the rock record to deepen our understanding of the Earth's natural processes and the ways in which human activities may be transforming them.

Computer models of natural processes may be limited by mathematical understanding, computational power, or data. In recent years, there have been great advances in mathematical treatment of complex systems and staggering advances in computational power. But these have not been matched by comparable advances in data from the natural world. Like any chain of reasoning, computer models are only as strong as their weakest link. Without evidence from the physical world, both as input to models and as a check on them, we run the risk of constructing computational houses of cards. And without discussion of the criteria by which we judge our models, we run the risk of obscuring the profound social issues at stake in the shadows of our edifices of computational prowess.

Down to Eartl
at Government

Gc

the top of a low coastal bluff lo
there is a thick, tarnished bra
embedded in a heavy slab of
cumference of this disk are star
below them the word *tidal*, plu
identify uniquely this particular
hundreds of thousands, if not
throughout the United States;
veyed location required for diff
mark the exact corner of a squa
surveyor with the precise geosp
the benchmark was set. These
Coast and Geodetic Survey (ano
U.S. General Land Office Surve
Survey.

Many Geological Survey ber
hilltops or mountaintops that co
tryside. Each constitutes a pern
pose of making a detailed topog
published and sold to the public
topographic quadrangle maps—
contiguous United States, a daur

Preparing and updating th
the U.S. Geological Survey (U
States Congress in 1879. As old