Water Resources Research

EDITORIAL
10.1029/2019WR026398

Key Points:
- "Information" can be considered both a "physical" quantity and a "statistical" basis for representing incompleteness in knowledge
- Information-based approaches have the potential to explain complexity, that is, emergent patterns of form and function that we see in Nature
- Information Theory helps characterize cause and effect, diagnose model inadequacy, study coevolution, and assess explanatory power

Citation:
Kumar, P., & Gupta, H. V. (2020). Debates—Does Information Theory provide a new paradigm for Earth science?. *Water Resources Research*, *56*, e2019WR026398. https://doi.org/10.1029/2019WR026398

Received 23 SEP 2019
Accepted 4 JAN 2020

Debates—Does Information Theory Provide a New Paradigm for Earth Science?

Praveen Kumar[1] and Hoshin V. Gupta[2]

[1]Department of Civil and Environmental Engineering, University of Illinois, Urbana, IL, USA, [2]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA

Abstract The basis for all knowledge is "information" that we compile about the world, expressed through models that support understanding, prediction, and decision making. This overview paper provides a contextual basis for the four papers that make up the "debate series" compiled under the above title. We briefly introduce Information Theory, discuss how "information" can be considered to be both a "physical" quantity and a "probabilistic" basis for representing incompleteness in knowledge, discuss the core motivation for this debate series, and briefly summarize the major arguments advanced by each of the debate papers. Our purpose is to facilitate an understanding of how these papers are related and how they approach the debate series question from different perspectives, while pointing to future directions for research. Finally, we invite further discourse and debate to advance the understanding and prediction of natural system dynamics using Information Theory, including the assessment of its limitations and complementarity to existing physics and machine learning approaches. Ultimately, our goal is to press for the development of philosophical and methodological advances that will enable the Earth science community to address some of the compelling unsolved problems in our field.

1. Introduction

The basis for all knowledge is "information" that we compile about the world. We use this information to construct models that support understanding, prediction, and decision making and to test hypotheses (expressed through models) regarding the nature of reality. While we often find it useful to use empirical relationships extracted from available data to support decision making, the purpose of science is ultimately to go beyond empiricism and establish representations that embody the dynamical causal relationships that underlie the structures, patterns, and behaviors that we observe.

Significant advances in Information Theory have been made during the past three decades, which go significantly beyond the original developments that paved the way for breakthroughs in statistical physics and communications technologies and that led to numerous applications in myriad other fields. "Information" is now accepted as both the metaphysical foundation for understanding nature and as the ontological basis for its characterization through relationships encapsulated in models. While these foundations are compelling, much remains to be resolved. The richness of the exciting advances now being made offers multiple perspectives by which to address the "nature" of Nature. These viewpoints are akin to the metaphorical "blind persons and the elephant" story, in which each blind person, having experienced contact with only a specific part of an elephant, argued that the elephant was fundamentally "like" a rope, snake, tree trunk, fan, spear, or a wall, depending on which part they were able to investigate (i.e., the tail, trunk, legs, ear, tusk, or the body).

So, while aspects of Information Theory have been around now for several decades, recent advances suggest that we are still at the initial stages of a scientific revolution with potentially significant ramifications to Earth science. Here, we argue that rapid advance can be achieved by reconciling the multiple different perspectives by which Information Theory enables a deeper understanding of nature, through detection and characterization of cause-effect relationships, through improved representation via models, and through more powerful (i.e., informative) tools for their evaluation. This debate series brings forward at least some of the different perspectives by which Information Theory can help to advance our understanding of nature.

In this overview paper, we seek to provide a contextual basis for the reader seeking to delve more deeply into the four papers that make up the "debate series" compiled under the common title "Does Information Theory provide a new paradigm for Earth science?" Section 2 provides a brief historical introduction to

10.1029/2019WR026398

Information Theory and the complementary perspectives by which "information" can be considered to be both a "physical" quantity and a "probabilistic" basis for representing incompleteness in knowledge, thereby providing a useful lens for the predictive understanding of physical processes. The debate that we wish to foster is regarding how these different perspectives can be both advanced and reconciled, thereby leading us toward an accurate characterization of nature. Consistent with the "blind-persons" metaphor, the debate offered by these four different, but unavoidably connected, perspectives seeks to enable a more coherent description of the "elephant"—in our case natural systems. There are two aspects to this—the debate offered *within* the boundaries of each perspective and the debate enabled by our collective attempt to articulate these complementary viewpoints, which when combined and resolved would be transformative. In other words, taken together, these perspectives create the opportunity for significant scientific advance.

Section 3.1 discusses the core motivation for this debate series and poses several questions that are important to consider, because doing so can affect how we approach the study of the Earth sciences. Sections 3.2–3.5 provide brief summaries of the major arguments advanced by each of the four papers in this compilation, thereby helping to facilitate both an understanding of how they are related and how they approach the debate series question from different perspectives, while pointing to future directions for research.

Finally, in section 4, we summarize the main points raised herein, argue for the need to more deeply investigate these issues, and point to the need for further debate. Ultimately, our goal is to press for the development of philosophical and methodological advances that will enable the Earth science community to address some of the compelling unsolved problems in our field.

2. Background

2.1. The Basis for Treating Information as a Physical Quantity

Using the idea of a virtual intelligent being, that later came to be known as "Maxwell's demon," James Clerk Maxwell (Maxwell, 1872) made the following argument in an 1867 letter to Peter Guthrie Tait:

> Consider a vessel full of air at a uniform temperature, in which the molecules are moving with a range of velocities. Assume that an intelligent being, carefully situated, can know the velocities of the molecules. Suppose that the vessel is divided into two equal halves A and B, with a trap door between them such that the being can open and close it, thereby allowing faster molecules to enter side A of the vessel and the slower ones to enter side B. By knowing the velocities, this intelligent being could create an energy gradient, which could be used to do work.

This thought experiment vividly demonstrates the fact that the knowledge or "information" about the velocities is a fundamentally different attribute than the velocities themselves, and the idea became the very early basis for quantifying information, with "Maxwell's demon" receiving considerable attention by the field of thermodynamics (Leff & Rex, 2002).

Following this argument, Szilard in 1929 (Szilard, 1929) developed a similar thought experiment but now using a single molecule of gas in a vessel with volume V at temperature T, to show that if we knew which half of the vessel the molecule was in, we could derive an amount of work equal to $W_{ext} = k_B T \ln(2)$ where k_B is the Boltzmann constant (i.e., $W_{ext} = \int_{V/2}^{V} p \mathrm{d}V = \int_{V/2}^{V} \frac{k_B T}{V} \mathrm{d}V$ as $pV = k_B T$ for a single molecule of gas). Since the probability of finding the molecule in any particular half of the vessel is 0.5, the amount of information gained from an error-free measurement associated with identifying its location is equivalent to the answer of a yes-no question (i.e., is the molecule present in this half?).

Tukey later referred to the amount of information associated with a binary yes-no question as a "bit" (Shannon, 1948*; pg. 380*), and this terminology now serves as a fundamental unit for measuring information. Importantly, the work, W_{ext}, associated with a binary decision process represents a clear link between energy (or work) and information, and this intimate link between the quantitative characterization of information and thermodynamics led Landauer (1991) to argue that information is a physical quantity. Although both Maxwell's and Szilard's efforts were "thought experiments," they have since been experimentally realized with the help of modern technologies (Toyabe et al., 2010; Koski et al., 2014, 2015; Vidrighin et al., 2016) that have paved the way for potential conversion of information to net energy, among other possible applications.

10.1029/2019WR026398

2.2. Information as a Measure of Surprise

The formal foundations for a quantitative characterization of information were established by the remarkable contributions of Shannon (1948) in the context of communication theory, who argued that information is associated with the amount of "surprise" associated with the occurrence of an event (i.e., the less likely the occurrence of an event, the larger is the information gained when the event occurs). Simply put, a more surprising outcome is associated with more information. Anchored by this idea, Shannon showed that the information, also call the surprisal, $I(X)$ associated with an event X is mathematically expressed as $I(X) = \log(1/p(x))$ where $p(x)$ is the probability of event X. Therefore, when integrated over all possible events in a probability distribution, the average information is

$$H(X) = \sum p(x) \log\left(\frac{1}{p(x)}\right)$$

This quantity, $H(X)$, is referred to as the "entropy" associated with our knowledge regarding the potential occurrence of event X, due to the similarity of this expression with that of thermodynamic entropy and its link therewith (see section 3). When a logarithm to the base 2 is used ($\log_2$) this measure of information is expressed in units of bits, and when a logarithm to the base e is used ($\log_e$) this measure of information is expressed in units of nats.

As such, entropy is a statistical property (i.e., the expected value of $\log\left(\frac{1}{p(x)}\right)$) of any probability distribution. However, due to its being related to the surprise associated with an event, it may be also regarded as a measure of ignorance, thus providing dual approaches to inference associated with a probability distribution function. In communication theory, this "information entropy" quantifies the upper bound for communication through a noisy channel and has thereby served as the foundation for development of myriad systems of communication.

If we apply this concept to a simple binary Bernoulli process with probability p of occurrence of an event X, it is easy to see that the entropy $H(X) = -p \log p - (1-p) \log(1-p)$ and has a maximum when the probability of the event $p = 1/2$. Conversely, when $p = 0$ or 1 (i.e., when we have a certain outcome), the entropy $H(X) = 0$. So, in general, the higher the a priori uncertainty associated with an event is, the higher is the entropy (ignorance) and hence the surprise. In this sense, therefore, entropy $H(X)$ provides a quantitative measure of the uncertainty associated with an event, as expressed by its probability distribution.

2.3. Maximum Entropy as a Basis for Representing What We Know

This idea of using entropy to help quantify uncertainty was brought to the fore through the work of Jaynes (1957), who posed the following question: *Suppose that you know certain attributes of a distribution (e.g. mean, variance, etc.), but not the whole distribution, how does one select a candidate for the distribution based on such partial information?* Jaynes showed that the best choice is the distribution (from the set of candidate distributions that satisfy the constraints of what is known) that has the *maximum entropy*. In other words, we should plead maximum ignorance or uncertainty about what we do not know, and this idea is known as the *Principle of Maximum Entropy* (POME). By applying this idea in the context of statistical mechanics, Jaynes showed that the result is a thermodynamic entropy that is equivalent to information entropy up to a multiplicative Boltzman constant (k_B), thereby further reinforcing the physical nature of information.

Early applications of these concepts to the Earth sciences include, for example, studies of landscape evolution to predict river profiles as the most probable states given the constraints of geologic structure and lithology (Leopold & Langbein, 1962), as mechanisms for dealing with limited sets of observations, for design of systems (including observational networks) under uncertainty, and selection among competing hypotheses, among several others. Please see Singh (1997) for a detailed review and refer to various accessible books on the subject (Singh, 2015).

2.4. Information as a Lens for Predictive Understanding of Physical Processes

During the past three decades, a number of developments have profoundly reshaped our physical understanding of the universe in terms of information and have therefore led to the development of novel paradigms for its characterization. Through his provocative notion "It from bit," Wheeler (Wheeler & Ford,

1998) pointed out that all of our physical understanding is rooted in information. Specifically, he argued that "laws of physics" are *informational* statements, in that they provide the best description of what we know. This contrasts with the more traditional notion that laws of physics represent reality and serve as the basis or genesis of information.

In this regard, Lloyd (2002) argued elegantly that the universe itself is a computer that is therefore continually producing information, and, in fact, the *Quantum Computing* paradigm (which has been rapidly gaining traction), initially proposed by Benioff (1980), relies on tapping this computational behavior of nature. The point is that, while the laws of physics help us to understand elementary behavior, they have not been as successful in explaining the foundations of complexity, that is, the emergent patterns of form and function that we see in nature, which perhaps arise due to its computational nature.

Information, therefore, enters the discourse about Nature from two perspectives. First, through the contention that information is physical, that is, it is ontological. Second, by the fact that it is the primary attribute by which we understand the world around us, that is, it is epistemological. It is therefore natural to explore how this duality can be exploited to develop a better understanding of Earth system processes.

3. Benefits of an Information Theoretic Perspective to the Earth Sciences

3.1. New Perspectives Through Information Theory

This brings us to one of the core motivations for the series of papers that responds to this debate topic: *Does Information Theory provide a new paradigm for Earth science?* Or framed another way, *Can information-based approaches help us develop a paradigm shift in the way we understand and predict the complex patterns, forms and behavior of Earth systems?* These are, perhaps, loaded questions with several interesting ramifications. For example,

1. Should we rethink our conceptualizations of natural processes in terms of information and information flows, and if so, how do we do this?
2. Can the lens of information make visible the interactions, dependencies, and/or signaling between processes that would otherwise remain hidden when viewed from the viewpoint of "laws of nature"?

To clarify, the laws of nature are often stated in terms of conservation of mass, momentum, and energy. The question is whether this is sufficient to properly characterize the diversity of spontaneous emergent forms, or whether information-based approaches could add potency in this regard. This leads us to explore the following possibilities:

1. Can information-based approaches better inform the development of Earth system models that are intended to be isomorphic representations of natural processes?
2. Can we characterize the information content of models themselves, and can this characterization be used for model improvement?
3. Can natural system dynamics be represented and predicted using information-theoretic frameworks, in forms that are independent and/or complementary to physics-based modeling approaches?
4. Does information have a role in the deterministic characterization of natural processes?

The goal of this debate series is to argue, based on emerging research, that answers to all of these questions are in the affirmative, with the goal to explore the pros and cons associated with information-based thinking. We expect that such exploration will enable us to identify and overcome bottlenecks and help to drive innovative approaches (by the broader community of Earth system scientists) that explore novel avenues to address some of the compelling unsolved problems in our field (Blöschl et al., 2019). Potential benefits of the Information Theory perspective range from stimulating novel paradigms of thinking about natural processes (as it has in the past) to developing tools for addressing practical issues.

From our perspective, the rapid pace of advance associated with Information Theory during the past several decades, even simply in the context of physics and physical processes, has been truly astounding, and we believe that information-based approaches offer the potential to address some of the most compelling challenges in the Earth sciences. The four papers that make up the initial contributions to this debate provide further perspective on this conversation, through a careful articulation of pros and cons. We provide an overview of each of the papers below.

10.1029/2019WR026398

3.2. Information Theory and the Study of Cause and Effect

In their contribution to this debate series, Goodwell et al. (2020) focus on the use of Information Theory to understand interactions among observed variables and to thereby detect and characterize cause-effect dependencies. Their article explores the problem of extracting such attributes from data measured in the natural environment, particularly where an interventional approach for ascertaining cause-effect relationships is not possible.

Specifically, their approach is based on the notion that detecting propagating fluctuations between variables can reveal interdependencies between them. These fluctuations may be either externally imposed or internally generated in a system but can be used to understand cause and effect interactions, forcings, and feedback. The authors argue that recent advances in Information Theory, such as the concepts of transfer entropy (Schreiber, 2000), partial information decomposition (Williams & Beer, 2010), momentary information transfer (Pompe & Runge, 2011), and the causal history approach (Jiang & Kumar, 2019) have paved the way for novel thinking and methods to characterize interdependencies among several variables through their time series signatures. *Transfer entropy* measures the reduction in uncertainty of a predicted variable due to the knowledge of the history of another variable. Temporally averaged, this reduction in uncertainty represents a measure of *causal dependency* in the Granger sense (Granger, 1969). *Momentary information transfer* takes this a step further by characterizing the strength of information transfer from a variable at some specific time in the past to a different variable at the current time. Causal history then considers the sum total of the influences of all interacting variables in the past on the current state of any variable.

What makes such approaches different from traditional approaches, such as correlation-based methods, is that they exploit temporal asymmetries (i.e., the future cannot cause the past) to enable a characterization of cause-effect dependencies. As explained by the authors, different attributes of dependencies can be discerned from multivariate time series data, thereby enabling deeper understanding of a variety of complex systems. In particular, this knowledge enables the construction of process networks that consist of variables as nodes and asymmetrically directed information flows between them as links. A causal history analysis of such networks enables a "whole system" characterization to be achieved, thereby providing a broader context for the consideration of causal dependencies.

However, progress in this area requires that the "curse of dimensionality" challenge be addressed. As the number of variables to be considered increases, the dimensions of the associated (conditional) probability density functions increase rapidly, which geometrically drives up the amounts of data required to robustly estimate the relevant information measures. While important methods have been developed in this regard, much remains to be done. A further challenge resides in the need for careful physical interpretation of the information measures; that is, do they simply provide a statistical interpretation of interdependence, or do they truly characterize physical relationships, thereby providing the new dimension to understanding physical systems envisioned earlier.

3.3. Information Theory and the Testing of Scientific Hypotheses

In their contribution to this debate series, Nearing et al. (2020) take an information theoretic perspective to the problem of developing and testing Earth system models. They point out that (i) despite the perspective advanced by most popular philosophical theories about hypothesis-driven science, Earth system models are fundamentally not "truth-apt," and (ii) a major and unavoidable problem with the idea of basing inference in the characterization of "uncertainty" is that strictly reliable uncertainty accounting is impossible; that is, we cannot know what we do not know. Further, because explaining a phenomenon requires modeling it using a *collection* of hypotheses, only the entire model and not the individual hypotheses can be tested directly. The fatal consequences of this, when employing classical (probabilistic or statistical) inference, are model misspecification and the potential for unbounded errors in inference.

The authors argue that instead of approaching the modeling/inference problem from the perspectives of "true/false" and "uncertainty," we might instead focus on a more tractable problem of characterizing "information." This is based in the recognition that what we really want is to measure the ability of models to provide accurate information about some phenomena of interest and, as such, the motivating question for inference becomes: "How much information do we have and how well do we use it?"

Accordingly, the authors reframe the problem of hypothesis testing as being one where the proposition to be examined is "whether there exists any information in the available experimental data that has not been captured by the model" and argue that a conservatively bounded estimator can be constructed by which to test this proposition by resorting to a benchmark model based in a machine-learning approach. They further argue that it is possible to use information measures to bound the effects of model error as segregated from data error.

The issues raised by this paper have several interesting implications for the Earth sciences. A number of studies have previously demonstrated that machine-learning-based models can robustly outperform traditional Earth system models at the task of estimating key states and fluxes that those models were designed for, where information loss typically arises from a combination of both parameter error and model structural inadequacy (Nearing et al., 2018). Further, multimodel ensembles for characterizing model structural uncertainty are unable to avoid the problem of perpetuating substantial bias.

As such, it becomes important to recognize at least two complementary types of model performance, namely, *predictive accuracy* and *functional accuracy* (Ruddell et al., 2019). *The former refers to* the ability of a model to reproduce observed data, while the latter refers to the ability of a complex model to reliably simulate the various dynamical interactions in a complex system across different scales. This distinction is particularly critical when we cannot rely on the future looking like the past and must instead require accurate simulations of system dynamics. In other words, a model that adequately describes the system dynamics should not have to compromise functional accuracy to achieve better predictive accuracy.

As with section 3.2 above, the important challenges lie in quantifying the information contained in high-dimensional data sets (the curse of dimensionality issue mentioned previously) and the need for increasing amounts of data to overcome the loss of the regularizing effect associated with strongly posed (and therefore potentially incorrect) modeling assumptions, when instead using *Jaynes*' POME for the posing of model structural hypotheses. The potential benefit is the ability to test scientific hypotheses in a meaningful way, thereby providing an alternative perspective on how we think about the diagnosis of relationships between our models and the real world.

3.4. Information Theory and Physics

In their contribution to this debate series, Perdigão et al. (2019) address the question of whether or not Information Theory has a role in the deterministic characterization of natural processes. They point out that, historically, the fields of *statistical thermodynamics* and *Information Theory* have characterized entropy as an aggregate system level descriptor. In other words, in *statistical thermodynamics* entropy quantifies the diversity of distinguishable microstates within the aggregate over which statistics are drawn, while in *Information Theory* it quantifies the number of independent probing actions required to elicit the nature of the system states (e.g., the number of questions that must be asked; see section 2). Further, *thermodynamics* provides a physical context for the production of entropy through irreversible processes that have a dissipative component.

The authors also review the debate about whether entropy is simply a measure of uncertainty or an actual physical quantity and point out that it can be either/both depending on the context (i.e., depending on what we are looking at). Importantly, they point out that beneath the statistically motivated concepts of entropy lie fundamental physics that characterize the system in terms of its underlying dynamic mechanisms, via the phase space, whose dimensionality quantifies the number of state variables plus their rates of change.

Viewed from this perspective, *thermodynamic* and *information-theoretic* considerations can be connected to the *kinematic-geometric* properties of the system footprint in its phase space. This provides a way to not only capture the diversity of microstates and the overall topology of the system but to also characterize the dynamics of the system in terms of its state variables and momenta, thereby helping to elicit the underlying mechanisms that shape its dynamics, that is, the geometric footprint associated with the underlying dynamical physical laws.

In this context, the authors argue that dynamic couplings are often wrongly perceived as coevolution, when in reality they entail kinematic-geometric covariation of systems in ergodic balance, that is, the variables change in time but the overall system does not. One consequence is that coupled models, such as those in classical climate dynamics and sociohydrology, prescribe time variation of local codependencies (cross derivatives) *but do not allow for the system as a whole to evolve*. This is because the representations are rigid and grounded on ergodic invariants, and so any simulated variability comes out of artificial stochastic noise, itself devoid of evolutionary meaning. In that regard, while such models enable deterministic bifurcations and intermittence between regimes, they *do not allow for any true system innovation*, *thereby attempting to simulate long-term dynamics with laws/representations that are only valid in the short term*.

Accordingly, the authors argue that the dynamic laws of physics do not need to be *deterministic* in the classical Newtonian sense. While classical mechanics assumes rigidity of the "invariants of motion" through stringent deterministic models, which then require stochastic ad hoc patches to reflect dynamic diversity, *such invariance is no longer required in the emerging nonergodic coevolutionary approaches*, and we can therefore begin to actually capture coevolutionary features seen across Earth system dynamics (the authors provide examples). In this regard, dynamical system evolution can be seen as a learning process akin to a natural adaptive machine-learning engine.

Of course, in practical applications where data are limited, the operational challenge resides in deriving kinematic-geometric and information-theoretic diagnostics from sparse data records. In this regard, the authors point to the growing literature on use of Information Theory to infer dynamical relationships among system variables and on information estimators for sparse data.

Importantly, however, the primary scientific challenge is to ask the right questions about the system, and the authors point out that (a) *classical correlation* measures only capture the tip of the iceberg; (b) *information-theoretic* measures can capture more features; but (c) the study of *kinematic-geometric* couplings can reveal underlying dependency structures. Accordingly, the fundamental question remains one of how best to read nature, and the field of *information physics* can provide important perspectives in this regard.

3.5. Model Complexity and Algorithmic Information Theory

Finally, the contribution by Weijs and Ruddell (2020) to this debate series approaches the topic from the perspective of *Algorithmic Information Theory*, wherein modeling is viewed as a form of data compression. In other words, a "perfect" model is an optimally compressed lossless representation of the information contained in available data about a system. They point out that both "science" and "data compression" have the same objective, which is the discovery of patterns in (observed) data with a view to describing them in as compact a form as possible. Accordingly, the success of any modeling endeavor can be measured in terms of the "compression efficiency" that has been achieved, and Information Theory provides powerful tools that enable us to measure and optimize model complexity versus performance in a common unit—the "bit" (see section 2).

To motivate this perspective, the authors refer to the philosophical principle known as "Occam's razor," which states that we should prefer the simplest hypothesis/model that can achieve a particular level of model predictive performance. In this regard, they point out that while model performance, that is, the measure of missing information needed to describe the original observations given the model, and model complexity, that is, the measure of model description length in bits, can be viewed from a multiple-criteria (Pareto-optimal) tradeoff perspective (see section 3.2), because both can be meaningfully quantified using the common unit of "bits," the success of any modeling endeavor can actually be assessed by collapsing the two into a single measure of lossless model size that, when minimized, leads to the optimal ability to both generalize and make predictions.

Importantly, given the current resurgence of machine learning in the Earth sciences (e.g., see recent WRR special issue on "Big data and machine learning in water sciences"), such an approach enables both "data-driven" and "physics-based" models to be placed on a meaningful continuum, which has the potential to enable insights regarding how to combine the strengths of both while avoiding the weaknesses of each. This is important because, while scientists typically add to (rather than subtract from) model complexity as more is learned about the physics of the associated system, several studies have reported the seemingly counterintuitive result that simple statistical models can often provide superior predictions when compared with detailed process-based models (see discussion in Nearing et al., 2020).

One way to understand this seeming paradox is that all theory stems ultimately from observations and that both "physics-based" models and "data-driven" models are attempts to simplify and generalize from narrow observations to the broader reality. In other words, all models are attempts to compress information and achieve parsimony in the explanation of the phenomenon. The authors delve into the different approaches for doing this, noting that if a compression/model does not perfectly describe the original observations, then it should be considered to be a "lossy" model. In this regard, while one might hope that the lossless size of a "physics-based" model is smaller than the size of the original observations from which it is constructed, this is often not the case because of mistaken physics, errors in observations, flawed assumptions, overgeneralization from dissimilar contexts, and/or overly complex model representations.

The authors further make the connection to *Bayesian inference* and *Maximum-Likelihood Theory*, by demonstrating that the description length associated with model loss is minimized when the likelihood of the data given the model is maximized. Both the philosophy of *Maximum Likelihood* and *Maximum A Posteriori Probability* estimation can therefore be justified from a "maximum compression" perspective.

Finally, the authors discuss the power offered by *Algorithmic Information Theory*, in which rigorous quantification of the information content in models and data is pursued in terms of the lengths of descriptions/programs. This perspective results in a notion of "strong parsimony," in which a theoretical limit to compression is defined as the (bit length) size of the smallest program that can reproduce the data. Given that Earth system models are run on digital computers, they can (in principle, as with any other computation) be assigned description lengths in bits that include both performance and complexity, and, consequently, their relative explanatory power (ability to infer patterns from observations) can be assessed. Such an approach would allow the study of whether adding "physics-based" insights into such a model, in the search for improved understanding, will actually add information that helps to improve predictions, or may instead hurt prediction performance.

Accordingly, quantifying the information content contained in "physics-based" and "data-based" knowledge will be essential to the task of merging the pattern detection and functional-characterization power of emerging machine learning approaches with the causal explanatory power expressed by accumulated scientific knowledge.

4. Summary and Conclusions

In summary,

1. While the "laws of physics" successfully help us to understand elementary behavior, they have not been as successful in explaining the expressions of "complexity," that is, the emergent patterns of form and function that we see in nature.
2. "Information" enters the discourse about *nature* from two perspectives—through the contention that information is "physical" (ontological) and by virtue of the fact that all of our physical understanding is rooted in information (epistemological); that is, the "laws of physics" are informational statements. It is therefore natural to explore how this duality can be exploited to develop a better understanding of Earth system processes.
3. "Information" can be rigorously quantified in terms of "bits," which characterize the number of binary ("yes-no") questions that must be answered in order to unambiguously characterize any state of affairs.
4. The works of Shannon, Jaynes, and many others provide the firm philosophical and mathematical foundations needed for the use of information-theoretic metrics in the characterization of scientific knowledge. In this regard, the POME provides a meaningful basis for incorporating "ignorance" (lack of knowledge) into our scientific hypotheses and/or explanations about reality.

Given these facts, this series of papers argues that Information Theory does, in fact, provide a new paradigm for Earth science by enabling better understanding and prediction of the complex patterns, forms, and behaviors that characterize Earth systems (see questions posed in section 3.1). While the laws of physics provide important constraints on the evolution of system behavior, information-based approaches have the potential to enable better characterization of complexity and emergent behavior and to address some of the compelling unsolved problems in our field.

As such, the papers compiled in this collection represent an attempt to stimulate further debate and to encourage deeper exploration of the questions raised herein. Goodwell et al. (2020) point to the need for framing causal interactions and feedback as information propagation and using it for the detection and characterization of multivariate dependencies from the information contained in available time series data about system behavior, thereby enabling deeper understanding of complex systems. Nearing et al. (2020) point to the need for hypothesis testing approaches that characterize whether a given hypothesis/model has captured all of the information contained in the available experimental data, thereby providing an alternative perspective on how we diagnose the relationships between our models and the real world. Perdigão et al. (2019) point to the need to go beyond the *statistical thermodynamic* and *information-theoretic* perspectives, to explore the *kinematic-geometric* properties that characterize the underlying mechanisms that shape the dynamics of a system, thereby enabling the study of nonergodic system evolution and coevolution. Weijs and Ruddell (2019) point to the need to recognize that models of Nature are, ultimately, attempts at data compression in which the patterns expressed by Nature are described in as compact a form as possible, thereby enabling a meaningful assessment of their relative explanatory power. Each of these papers points clearly to future directions for investigation.

Ultimately, of course, the primary scientific challenge is one of how best to read Nature, by both asking the right questions about the system and by developing tools that effectively enable exploration and answers to those questions. Through this debate series, we propose that Information Theory can provide an important basis, indeed an emerging paradigm, for the study of dynamical Earth systems. It seems particularly timely to adopt this perspective given the current resurgence of machine learning in the Earth sciences and the need to bridge across, and ultimately merge, the physics-based and machine-learning approaches to characterize knowledge. In fact, all of the issues raised through this collection of papers would seem to be highly pertinent to doing so.

In conclusion, we have pointed to four overlapping directions for future development, but no doubt there are others that have not been represented herein. We also acknowledge that not everyone may agree with our characterizations of either Earth science or the related benefits of Information Theory. We therefore invite further debate on the question of whether Information Theory does, in fact, provide a new paradigm for Earth science, with the dual purpose of exposing limitations in our current approaches and bringing forward philosophical perspectives and methodologies that will enable rapid future progress

Acknowledgments

We are grateful to the *Karlsruhe Institute of Technology* and the *Environmental Hydraulic Institute* (IHCantabria) of the *Universidad de Cantabria* for hosting several workshops and a summer school where the authors contributing to this debate series met to discuss their differing but complementary perspectives on this topic. It was at the *2018 Summer School in Information Theory in the Earth Sciences* (SITES; http://sites-2019.ihcantabria.com; Prieto et al., 2018) held at IHCantabria that the idea for this WRR debate series was born. P. K. acknowledges support from NSF Grant EAR-1331906, and H. V. G. acknowledges partial support by the Australian Centre of Excellence for Climate System Science (CE110001028).

References

Benioff, P. (1980). The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by turing machines. *Journal of Statistical Physics*, *22*, 563.

Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., et al. (2019). Twenty-three unsolved problems in hydrology (UPH)—A community perspective. *Hydrological Sciences Journal*, *64*(10), 1141–1158. https://doi.org/10.1080/02626667.2019.1620507

Goodwell, A. E., Jiang, P., Ruddell, B. L., & Kumar, P. (2020). Debates—Does Information Theory provide a new paradigm for Earth science? Causality, interaction, and feedback, submitted to. *Water Resources Research*.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424–438.

Jaynes, E. T. (1957). Information Theory and statistical mechanics. *Physics Review*, *106*(4), 620–630. https://doi.org/10.1103/PhysRev.106.620

Jiang, P., & Kumar, P. (2019). Information transfer from causal history in complex system dynamics. *Physical Review E*, *99*, 012306. https://doi.org/10.1103/PhysRevE.99.012306

Koski, J. V., Kutvonen, A., Khaymovich, I. M., Ala-Nissila, T., & Pekola, J. P. (2015). On-chip Maxwell's demon as an information-powered refrigerator. *Physical Review Letters*, *115*(26), 260602. https://doi.org/10.1103/PhysRevLett.115.260602

Koski, J. V., Maisi, V. F., Pekola, J. P., & Averin, D. V. (2014). Experimental realization of a Szilard engine with a single electron. *Proceedings of the National Academy of Sciences*, *111*(38), 13,786–13,789. https://doi.org/10.1073/pnas.1406966111

Landauer, R. (1991). Information is physical. *Physics Today*, *44*(5), 23. https://doi.org/10.1063/1.881299

Leff, H., & Rex, A. F. (Eds) (2002). *Maxwell's demon 2: Entropy, classical and quantum information, computing*. Boca Raton, FL: CRC Press. ISBN: 978-0-7503-0759-8.

Leopold L. B., and W. B. Langbein (1962), The concept of entropy in landscape evolution, Geological Survey Professional Papers 500-A, United States Department of the Interior. https://doi.org/10.3133/pp500A

Lloyd, S. (2002). Computational capacity of the universe. *Physical Review Letters*, *88*(23), 237901. https://doi.org/10.1103/PhysRevLett.88.237901

Maxwell, J. C. (1872). *Theory of heat*, (Second ed.p. 324). London: Longmans, Green, and Co. ISBN: 978-1-297-65733-7.

Nearing, G. S., Ruddell, B. L., Clark, M. P., Nijssen, B., & Peters-Lidard, C. (2018). Benchmarking and process diagnostics of land models. *Journal of Hydrometeorology*, *19*(11), 1835–1852.

Nearing, G. S., Ruddell, B. L., Bennett, A. R., Prieto, C., & Gupta, H. V. (2020). Debates—Does Information Theory provide a new paradigm for Earth science? Hypothesis testing, submitted to. *Water Resources Research*.

Perdigão, R. A. P., Ehret, U., Knuth, K. H., & Wang, J. (2019). Debates—Does Information Theory provide a new paradigm for Earth science? Emerging pathways of information physics, submitted to. *Water Resources Research*.

Pompe, B., & Runge, J. (2011). Momentary information transfer as a coupling measure of time series. *Physical Review E*, *83*, 051122. https://doi.org/10.1103/PhysRevE.83.051122

Prieto, C., Ehret, U., & Nearing, G. (2018). Using Information Theory in Earth sciences. *Eos*, *99*. https://doi.org/10.1029/2018EO107389

Ruddell, B., Drewry, D. T., & Nearing, G. S. (2019). Information Theory for model diagnostics: Tradeoffs between functional and predictive performance in ecohydrology models. *Water Resources Research*, *55*(8), 6534–6554. https://doi.org/10.1029/2018WR023692

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, *85*(2), 461–464. https://doi.org/10.1103/PhysRevLett.85.461
Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
Singh, V. P. (1997). The use of entropy in hydrology and water resources. *Hydrological Processes*, *11*, 587–626. https://doi.org/10.1002/(SICI)1099-1085(199705)11:6<587::AID-HYP479>3.0.CO;2-P
Singh, V. P. (2015). *Entropy theory in hydrologic science and engineering*, (p. 805). New York, N. Y: McGraw Hill.
Szilard, L. (1929). über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen (On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings). *Zeitschrift für Physik*, *53*(11–12), 840–856. https://doi.org/10.1007/BF01341281
Toyabe, S., Sagawa, T., Ueda, M., Muneyuki, E., & Sano, M. (2010). Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. *Nature Physics*, *6*, 988–992. https://doi.org/10.1038/nphys1821
Vidrighin, M. D., Dahlsten, O., Marco, B., Kim, M. S., Vedral, V., & Walmsley, I. A. (2016). Photonic Maxwell's demon. *Physical Review Letters*, *116*(5), 050401. https://doi.org/10.1103/PhysRevLett.116.050401
Weijs, S. V., & Ruddell, B. R. (2020). Debates — Does Information Theory provide a new paradigm for Earth Science? Sharper Predictions using Occam's Digital Razor, to appear in Water Resources Research.
Weijs, S. V., Ruddell, B. R., Ehret, U., Knuth, K. H., & Wang, J. (2020). Debates—Does Information Theory provide a new paradigm for Earth science? Sharper predictions using Occam's digital razor, submitted to. *Water Resources Research*.
Wheeler, J. A., & Ford, K. (1998). *Geons, black holes, and quantum foam: A life in physics*. New York: W.W. Norton & Co. ISBN: 0-393-04642-7.
Williams, P. L., & Beer, R. D. (2010). Nonnegative decomposition of multivariate information, CoRR. Available at http://arxiv.org/abs/1004.2515