EOS
EOS, TRANSACTIONS, AMERICAN GEOPHYSICAL UNION

VOLUME 95 NUMBER 32
12 August 2014
PAGES 285–292

Exploratory Modeling: Extracting Causality From Complexity

PAGES 285–286

On 22 May 2011 a massive tornado tore through Joplin, Mo., killing 158 people. With winds blowing faster than 200 miles per hour, the tornado was the most deadly in the United States since modern record keeping began in the 1950s.

As with any similar disaster, the event posed a number of questions: How and why did this happen? What forces caused the tornado to form? What were the properties of the air and the land that spurred it to such great strengths? In many cases, researchers work to answer these questions not just for the disaster in question but in a general sense.

Ever-advancing computing resources tempt researchers to simulate environmental systems in ever-increasing detail. An atmospheric model able to represent fine-scale turbulent eddies coupled to a land surface model, for instance, could be used to reproduce the Joplin tornado to try to figure out exactly why that storm grew so large. However, although such detailed simulations may be suitable for scenario-based predictions such as a specific historical event, they often are too contextually dependent to investigate fundamental cause-effect relationships. In short, advanced models may tell us how specifically but not why in general.

A major contemporary scientific challenge is to develop ways to resolve causality, not just correlation, in large-scale, nonlinear Earth systems. The goal is to answer the question of what makes a tornado stronger rather than what made the 2011 Joplin tornado so strong.

One way to attempt to resolve causality in a complex system is to adopt an exploratory modeling approach (see Figure 1). In such an approach, simple models are run repeatedly with different combinations of parameter values, governing mechanisms, or levels of mechanistic detail. At each stage of this iterative process, hypotheses are generated, tested, and refined in conjunction with field observations.

This approach has been refined in the past 20 years in the geomorphological sciences. Here we explore the new opportunities afforded by increasing computing power and expanding sensor networks used in conjunction with exploratory modeling.

Appropriately Minimalist Modeling

Scientists who study nonlinear, dynamic systems have long been aware that complex phenomena may arise from simple processes. Simple sets of physical interactions within certain ranges of parameters can produce chaotic behavior, spatially periodic patterning, catastrophic shifts, phase transitions, or (multi) fractal scaling [*Schertzer and Lovejoy,* 2011].

Exploratory modeling is, in essence, a philosophical approach to identifying these underlying processes. The reduced detail and complexity in exploratory models enable hypothesis testing over large spatial and time scales and enable rapid testing of many combinations of parameters and processes without the need for an expensive supercomputer.

Exploratory modeling works by intentionally leaving out or simplifying physical details like turbulence or the role of nutrient transport on detrital sediment production, not necessarily because those details are poorly understood or computationally demanding but because doing so helps elucidate the essential causative processes.

To ensure that essential processes are represented in a way sufficient to reproduce observed emergent processes, exploratory models may go far beyond so-called toy—or bare bones—models with the most minimal level of detail. However, they remain differentiated from more figurative models in that secondary dynamics are absent or represented with very low levels of detail. Key to developing a successful exploratory model is selection of the appropriate level of detail, which should be just sufficient to reproduce the emergent phenomena of interest and any secondary details useful for selection between multiple models. Exploratory models generally work at the coarsest scale of process description commensurate with the problem being investigated.

Examples of Successful Exploratory Models

In geomorphology, the use of exploratory models to identify fundamental cause-effect relationships dates back 2 decades to the work of *Murray and Paola* [1994]. In their research the pair distilled braided stream dynamics down to an essential set of rules—in this case, how riverbed slope guides discrete parcels of water and sediment.

Using their simplified model, the authors were able to resolve a long-standing question in geomorphology. Namely, by using multiple model runs where different rules for sediment transport were switched on or off, Murray and Paola found that readily erodible banks unencumbered by topography are a key requirement for the formation of braided streams.

More recently, *Rozier and Narteau* [2014] used a few simple rules representing the motion of sand grains coupled to a model of airflow to elucidate the fundamental processes guiding barchan dune formation. The model's ability to reproduce dune merging and calving behavior suggests that simple rules of sediment motion are sufficient for the development of these behaviors in a broad sense.

As computational power increases, the tendency is to want to add complex physical processes into model simulations. In both of these examples, however, details such as spatially explicit turbulence would have been an unnecessary complication if the goal were to define first-order drivers and sensitivities.

Simple Models Facilitate Understanding Coupled Dynamics

Exploratory models are particularly powerful for understanding coupled physical-human-biological dynamics when there are many potential drivers or many spatiotemporal scales that are potentially important.

For example, to understand the critical drivers of landscape pattern formation in the Florida Everglades and its sensitivities to human stressors, *Larsen and Harvey* [2010] developed an exploratory model that coupled flow, vegetation, and sediment dynamics. So that the model could represent millennia of time, simplifications were made in how flow was simulated by obtaining approximate solutions to governing equations and decoupling the solution of vertical velocity profiles from that of horizontal flow. Phosphorus, a sensitive driver of Everglades vegetation but one hypothesized not to play a major role in

---

By L. Larsen, C. Thomas, M. Eppinga, and T. Coulthard

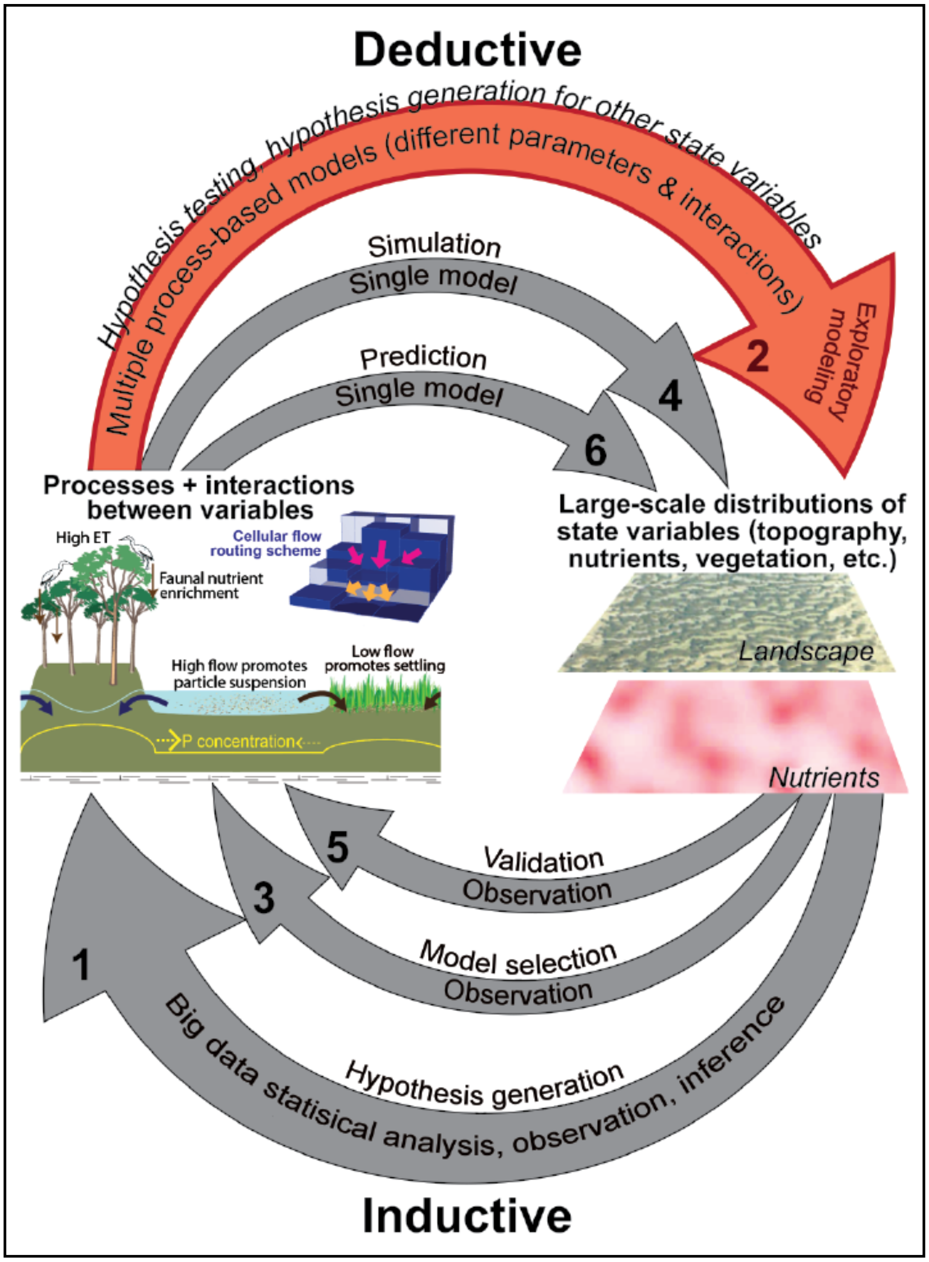

Fig. 1. Exploratory modeling can play one part within a comprehensive approach to studying complex environmental systems. In this conception, models and field observations become progressively more refined toward the center. The iterative refinement is complete when the model is sufficient to meet specified objectives. ET means evapotranspiration.

landscape patterning, was represented only indirectly in a term for a peat accretion rate.

Despite these simplifications, the authors used a detailed representation of how water flows through different vegetation communities that was developed in close association with field experiments and supported by data [*Larsen and Harvey,* 2010]. Consequently, restoration managers deemed the model trustworthy enough to use in simulating how flow affects sediment transport and landscape development, information that continues to influence subsequent restoration decisions.

In another recent example that combines paleoclimatology, hydrology, ecology, and archaeology, *Coulthard et al.* [2013] used an exploratory model to test the hypothesis that early humans were able to migrate out of Africa through the Sahara Desert by following green corridors along rivers. Critically, a simplified exploratory flow model allowed full hydrodynamic simulations of river flow and flooding at continental scales. The research showed that in previous wetter climates there could have been enough water to permit migration along these corridors.

In these and other examples, the strategies used to simplify complex processes and interactions to the appropriate level of detail are wide ranging. Strategies include rule-based cellular automata approaches, reduced spatial dimensionality of governing equations, decoupling equations for dynamics that are only weakly linked, and hierarchical techniques for representing cross-scale dynamics [*Hewitson and Crane,* 1996; *Royle and Dorazio,* 2008].

Exploratory Modeling as a Research Guide

Although predicting emergent behavior from a set of known interactions can be straightforward, ascertaining the critical interactions that explain observed emergent behavior is more challenging. The problem is that several different processes or parameter values can be combined to create the same emergent outcome, an issue known as equifinality.

Researchers using exploratory modeling address equifinality by constructing a variety of models that follow many potential pathways to produce an emergent phenomenon. Each of these different model constructions will behave slightly differently, and observations can be used to discriminate between alternate explanatory mechanisms. The benefit of exploratory modeling is that these various model constructions can guide specific observational research.

As an example, multiple mechanisms have been proposed to explain the maze-like ridge-hollow patterns ubiquitous in boreal peatlands. By turning simple formulations of these mechanisms on and off in a factorial design modeling experiment, *Eppinga et al.* [2009] showed that one of two mechanisms was responsible for the patterning: either a water stress feedback by itself or a water stress feedback coupled to a nutrient accumulation feedback. Further modeling research showed that nutrient concentrations would be lowest under ridges in the former scenario and highest under ridges in the latter.

Targeted sampling, motivated by the exploratory modeling, revealed that nutrient patterns in a continental peatland were consistent with the nutrient accumulation feedback, whereas those in a maritime peatland were consistent with the water stress feedback [*Eppinga et al.,* 2010]. Critically, the exploratory modeling provided recognition that nutrient patterns were mechanistically discriminatory and highlighted the most efficient strategy for field sampling.

Identifying New Causal Mechanisms

The models highlighted here were constructed on the basis of conceptual models and hypotheses developed from expert knowledge or from the mechanisms known to drive similar systems. By necessity, the causal understanding that developed through the application of these models was born out of trial and error through many iterative refinements. There is no guarantee that the exploratory modeling vetted all plausible hypotheses.

Fortunately, emerging statistical approaches in the geosciences may improve hypothesis generation and the efficiency with which models converge on plausible causal mechanisms. Granger causality [*Detto et al.,* 2012] and transfer entropy statistics [*Ruddell and Kumar,* 2009] resolve true causal relationships between variables as well as their time scale of interaction. These analyses enable inductive delineation of "process networks" from time series of multiple variables. Process networks constitute hypotheses of potential causal mechanisms, which would then be tested through, for example, an exploratory modeling process.

As the iterative process of model refinement and field observations proceeds, questions and hypotheses become more specific. Thus, the level of detail in the later models may extend beyond what one would typically consider to be an "exploratory" model.

In their braided stream research, the abstract exploratory models developed by *Murray and Paola* [1994] could replicate general channel dynamics yet exhibited scaling problems and failed to replicate observed braiding intensity or high-sinuosity meanders [*Doeschl-Wilson and Ashmore,* 2005; *Ziliani et al.,* 2013]. Later research that employed solutions to the full shallow-water equations with secondary circulation corrections (rather than Murray and Paola's simplified version) fixed these problems but necessitated the use of a supercomputer [*Nicholas,* 2013].

In this way, the exploratory model may be perceived as a deductive counterpart to exploratory statistics. Like a principal component analysis, it is useful as a way to identify the mechanisms responsible for the emergence of a large percentage of the coarse-scale stream behavior, but more detailed models may be needed to resolve the finer detail of that pattern.

A Foundation of Simplicity

Simplified models have flourished in the geosciences since the advent of computer programming. Yet progress on using simplified models to resolve causality in complex Earth and environmental systems has been uneven.

There seems to be an emerging recognition that exploratory models occupy a niche distinct from that of detailed simulation models, analogous to the way exploratory statistics occupy a niche distinct from predictive statistics. Exploratory modeling and exploratory statistics both yield a coarse, fundamental understanding of primary drivers and sources of variability and may serve as an important precursor to more detailed subsequent modeling.

Combined with continually advancing computing power and the introduction of new statistical techniques, exploratory modeling will provide geoscientists with ever more opportunities to enhance their understanding of complex systems. In fact, they are rapidly emerging as a key foundation for any endeavor that seeks to test hypotheses to better understand system drivers.

Acknowledgments

This feature reflects the contributions of a larger group of participants in the National Science Foundation–funded workshop "The Art and Science of Reduced Complexity Modeling in the Environmental Sciences" (NSF EAR-1263851). Participants' names and workshop resources, including toy models for educational use, can be accessed online at https://sites.google.com/site/rcmworkshop/home. We thank Norm Miller and two anonymous reviewers for helpful suggestions.

References

Coulthard, T. J., J. A. Ramirez, N. Barton, M. Rogerson, and T. Brücher (2013), Were rivers flowing across the Sahara during the last interglacial? Implications for human migration through Africa, *PLoS ONE, 8*(9), e74834, doi:10.1371/journal.pone.0074834.

Detto, M., A. Molini, G. Katul, P. Stoy, S. Palmroth, and D. Baldocchi (2012), Causality and persistence in ecological systems: A nonparametric spectral Granger causality approach, *Am. Nat., 179*(4), 524–535, doi:10.1086/664628.

Doeschl-Wilson, A. B., and P. E. Ashmore (2005), Assessing a numerical cellular braided-stream model with a physical model, *Earth Surf. Processes Landforms, 30*(5), 519–540.

Eppinga, M. B., P. C. de Ruiter, M. J. Wassen, and M. Rietkerk (2009), Nutrients and hydrology indicate the driving mechanisms of peatland surface patterning, *Am. Nat., 173*(6), 803–818, doi:10.1086/593132.

Eppinga, M. B., M. Rietkerk, L. R. Belyea, M. B. Nilsson, P. C. D. Ruiter, and M. J. Wassen (2010), Resource contrast in patterned peatlands increases along a climatic gradient, *Ecology, 91*(8), 2344–2355.

Hewitson, B. C., and R. G. Crane (1996), Climate downscaling: Techniques and application, *Clim. Res., 7*(2), 85–95.

Larsen, L. G., and J. W. Harvey (2010), How vegetation and sediment transport feedbacks drive landscape change in the Everglades and wetlands worldwide, *Am. Nat., 176*(3), E66–E79.

Murray, A. B., and C. Paola (1994), A cellular model of braided rivers, *Nature, 371*(6492), 54–57.

Nicholas, A. P. (2013), Modelling the continuum of river channel patterns, *Earth Surf. Processes Landforms, 38*(10), 1187–1196, doi:10.1002/esp.3431.

Royle, J. A., and R. M. Dorazio (2008), *Hierarchical Modeling and Inference in Ecology: The Analysis of Data From Populations, Metapopulations, and Communities,* 408 pp., Academic, Amsterdam.

Rozier, O., and C. Narteau (2014), A real-space cellular automaton laboratory, *Earth Surf. Processes Landforms, 39*(1), 98–109, doi:10.1002/esp.3479.

Ruddell, B. L., and P. Kumar (2009), Ecohydrologic process networks: 1. Identification, *Water Resour. Res., 45,* W03419, doi:10.1029/2008WR007279.

Schertzer, D., and S. Lovejoy (2011), Multifractals, generalized scale invariance, and complexity in geophysics, *Int. J. Bifurcation Chaos, 21*(12), 3417–3456, doi:10.1142/S0218127411030647.

Ziliani, L., N. Surian, T. J. Coulthard, and S. Tarantola (2013), Reduced-complexity modeling of braided rivers: Assessing model performance by sensitivity analysis, calibration, and validation, *J. Geophys. Res. Earth Surf., 118,* 2243–2262, doi:10.1002/jgrf.20154.

Author Information

LAUREL LARSEN, University of California, Berkeley; email: laurel@berkeley.edu; CHRIS THOMAS, British Geological Survey, Edinburgh, UK; MAARTEN EPPINGA, Utrecht University, Utrecht, Netherlands; and TOM COULTHARD, University of Hull, Hull, UK