AUTHOR'S PRE-PRINT. See published paper here:
https://doi.org/10.1016/j.envsoft.2020.104928

This is the pre-print (initial submission, before peer-review) of a peer-reviewed article of the same title, which has been published in *Environmental Modelling & Software*. The published version can be found here: https://doi.org/10.1016/j.envsoft.2020.104928

The citation for the published paper is:

Lim, Theodore C. 2021. "Model Emulators and Complexity Management at the Environmental Science-Action Interface." Environmental Modelling & Software 135 (January): 104928. https://doi.org/10.1016/j.envsoft.2020.104928.

Model emulators and complexity management at the environmental science-action interface

Author: Theodore Lim, Dept of Urban Affairs and Planning, Virginia Tech

Highlights

- Emulators can improve management decisions by facilitating non-modeler scenario testing
- Emulator purposes in environmental management: performance, platform, and perception
- Both modelers and managers learned about management during emulation process
- Emulators include response surface emulation and lower fidelity surrogates
- Lower fidelity surrogates more appropriate for environmental management contexts

Abstract

As our understanding of the interactions present in socio-ecological systems models advance, emulation modeling can help reduce the complexity and required computational resources of the models used to represent these systems. While emulation is commonly used in model meta-analyses and parameterization, it has been less explored in the context of environmental management. In this research, I analyze the reflections of a group of experienced watershed modelers on the training and performance of an emulator of a high-resolution hydrological model. I find that decreased simulation run-times are indeed important in enabling stakeholders to interact directly with the model; however, an emulator's ability to act as a platform and to manage stakeholder perceptions of the modeling process are equally important. Further, at the science-action interface, stakeholder perceptions play a significant role in the process of learning about management and the evolving approach to model emulation and balance of model simplicity and complexity.

Keywords

Emulation model; surrogate model; science-action interface; environmental management;

1 Introduction

Model emulators are "models of models," typically implemented in order to reduce the complexity and computational resources of numerical models (Castelletti et al., 2012; Ratto et al., 2012; Razavi et al., 2012a). Within decision-making contexts such as watershed management, the reduction of complexity and computational resources is attractive because it may result in faster scenario testing and lower barriers for stakeholders to directly interact with models. Besides these performance-based benefits however, how emulation models fit into complex decision-making contexts has been less explored (Hong et al., 2017). This is particularly true given recent advances in deep/machine learning which is process-agnostic, but highly predictive of environmental phenomena (e.g.: Nearing et al., 2020). The implications of faster and higher resolution model emulators (further enabled by deep/machine learning advances) on decision-making contexts is less addressed, however (Shen, 2018).

As "boundary objects" between environmental scientists, local and state governments, and many other stakeholders affected by watershed management (the particular area of environmental management that this article focuses on), models play important roles, including facilitating research, communication, and participation (Hamilton et al., 2019). The evaluation of environmental models should therefore extend beyond their ability to predict or explain physical phenomena, to include evaluation of their usefulness in other aspects of the socio-technical processes of environmental management (Elsawah et al., 2015; Hamilton et al., 2019; Voinov and Gaddis, 2008). In this view of the role of environmental models, model emulators are a unique object of study. On the one hand, they could play an important bridge between environmental science and management by increasing accessibility (through reduced complexity) of complex models. On the other hand, the use of emulators could result in less trust in the model or reduced salience for answering key management questions. The balance between increased accessibility and preserving credibility, salience, and legitimacy of the model depends on *how* complexity is reduced and the purpose of the model.

A number of trends continue to strengthen that indicate that important advancements will be made with increasingly computationally or numerically complex models and that model emulation will be utilized to improve the operationalization of such models. First, sustainable management of natural resources requires acknowledgment of the interconnectedness of social-ecological systems. Interconnections involve the (often non-linear) feedbacks that exist between multiple system compartments and models' coupling at different spatial and temporal scales (Liu et al., 2008; Mahmoud et al., 2009; Wagener et al., 2005). Second, process-based models, also called "bottom-up" models are increasingly used to represent emergent phenomena, especially in "no-analog" situations such as large-scale land use change or climate change (Fatichi et al., 2016). Lastly, computational resources are increasingly available for relatively low costs to researchers and scientists. The availability of these computational resources for researchers enables more computational power for increasingly demanding simulations, but has also led to advances in the field of machine learning and artificial intelligence. The combined use of machine learning coupled with physically-based models (Karpatne et al., 2017) exhibits promise for emulation modeling, including the capability to mimic high resolution spatial-temporal output, with vastly reduced simulation times (Lim, in review).

This study explores the role of emulation modeling in reducing complexity at the science-management interface of watershed management. Based on the qualitative analysis of primary gathered data from facilitated focus group responses and program documents from a key regional watershed in the United States, I offer a conceptual model of emulation modeling within the environmental management context. The emulation process conceptual model highlights the relationships between purpose definition and diverse learning activities. The proposed conceptual model captures the iterative nature of managing model complexity through balancing computational and representational complexities. To support the conceptual model, this research aims to make two **contributions**:

AUTHOR'S PRE-PRINT. See published paper here:
https://doi.org/10.1016/j.envsoft.2020.104928

1. Define the primary purposes of emulators used specifically in environmental management and planning.
2. Illustrate emulation as a process of balancing model complexity in a socio-environmental system management context that includes purpose refinement and diverse learning activities.

The following section presents further review of the key concepts relevant to this study.

2 Background

2.1 Emulation models

An emulation model (also known as a "meta-model", or "surrogate model") is a model, often of a reduced form, that mimics the outputs of an original model. Usually emulation modeling is undertaken to save on computational resources or time (Forrester et al., 2008; Ratto et al., 2012; Razavi et al., 2012b). The reduced computational and time expense of the emulation model allows for modeling meta-analyses such as sensitivity analysis, parameter optimization, or systems modeling, which require iterative sampling and many realizations of the model under different parameter or starting conditions (Pianosi et al., 2016). Recently, emulation modeling has been mentioned as an area of opportunity for additional research on the quantification of uncertainty (Fatichi et al., 2016), and for evaluating sensitivity (Pianosi et al., 2016) in the input parameters of deterministic, process-based hydrological models. People have also recognized the potential of model emulators to aid in the context of environmental decision-making to speed scenario testing (Carnevale et al., 2012). In integrated assessment models, which include interdependent submodels of environmental and social systems, emulators can help reduce computational barriers in model coupling and characterize systems' uncertainty (Fletcher et al., 2019; Haasnoot et al., 2014; Little et al., 2019).

There are two major approaches to emulation modeling: response surface surrogates, and lower fidelity surrogates (Razavi et al., 2012a). The former is best understood as a "model of a model," and the primary purpose is simply to reproduce the output of the original model with high fidelity. The "response surface" is the original model's output; the closer the emulator can come to reproducing the surface, given new input parameter values, the better. Less important are the processes that enable the response surface's reproduction. In contrast, lower fidelity surrogates preserve the physical bases of the model's structure that are deemed to be most relevant or important. They achieve computational cost savings through eliminating less relevant detail in the original model. In this article, the distinction between these two emulation approaches is important and I will refer to the former approach as response surface emulation, the latter as lower fidelity surrogates, and both together as types of "emulators." In particular, lower fidelity surrogates are highly related to questions about appropriate levels of complexity in environmental modeling.

2.2 Complexity and processes of model simplification

Academic discussion about appropriate levels of complexity in models has existed for at least as long as advances in computing have enabled large-scale models (Lee, 1973). The reader is referred to Hong et al. (2017), who provide a comprehensive review of model simplification in environmental studies. In brief, simple models are attractive because they are often more generalizable, can better summarize/represent causal relationships, are easier to implement, easier to adapt, and have shorter simulation times and computational requirements (Chwif et al., 2000; Van Nes and Scheffer, 2005). More complex models are attractive, especially in new knowledge creation, because they may be able to represent important processes not available in simpler models such as feedbacks, interactions, and higher spatial/temporal resolutions more specific to heterogeneous applications (Clement, 2011). In hydrological modeling, process-based models (sometimes called "bottom up" models), are often considered more complex, and are particularly beneficial because: they are based on known physical relationships and preserve closure of laws of conservation of mass, energy, and momentum (Fatichi et al., 2016); they are more amenable to coupling to other interdisciplinary or multi-compartment models (Condon and Maxwell, 2014; Lim and Welty, 2017; Williams and Maxwell, 2011); and can ensure decision-making relevance, interpretability and consistency, especially when historical patterns are not applicable for future conditions, such as under conditions of land use change or climate change (Mahmoud et al., 2009; Wagener et al., 2005).

There is agreement that the principle of Occam's Razor, also known as parsimony – or only including in a model what is absolutely needed-- is an appropriate guiding principle to follow in model building. In addition to being costly, data hungry, and prone to "lock in" and issues of equifinality (Beven, 1993), overly complex models are sometimes implemented for non-scientific reasons. Examples of these reasons include: "showoff" factor, "include all" syndrome – prevalent among inexperienced modelers who may not have enough understanding of systems' dynamics to determine what processes are important to include, and "possibility factor" - including processes simply 'because we can' rather than being motivated by concrete hypothesis testing or theory (Chwif et al., 2000). Sometimes models are selected because they happen to be within the expertise area of the experts using them, rather than being intentionally chosen for the question at hand (Voinov et al., 2018). Complex models have also been shown to have a psychological effect of garnering more trust than simple models (Oreskes, 2003). While in hydrological sciences, there has also been the sense that complexity is often a blockage to increased understanding (Kirchner, 2009; McDonnell et al., 2007).

The balance between complexity and simplicity in environmental modeling can be conceived of as a cycle during which science is advanced and models are refined. An adaptation of the "simplicity cycle" (**Figure 1**) depicts how model complexity increases as understanding of a phenomena increases (Ward, 2015). Yet, there is a distinction between "naive simplicity" and "simplicity." The former condition reflects the adage "you don't know what you don't know," while the latter is based on an intentional choice in the processes represented in the model, based on an understanding of what is most important to model (Schwartz et al., 2017). The figure also depicts that scientific understanding can increase while *decreasing* model complexity through "elegant simplicity" and that doing so is actually desirable. Occam's Razor should therefore not be applied to all situations, but rather only in situations where understanding of the system is being advanced. Simplification is not an "event," but rather part of a cyclic process. Van Ness and Scheffer (2005) characterized this cycle as consisting of three parts: scrutinize, synthesize, and simplify. This cycle implies that neither simple nor complex models should be inherently preferred. Instead, the combination of simple and complex and the cyclical process of scrutinizing, synthesizing, and simplifying is a powerful way to gain understanding of complex systems (Van Nes and Scheffer, 2005).

AUTHOR'S PRE-PRINT. See published paper here:
https://doi.org/10.1016/j.envsoft.2020.104928

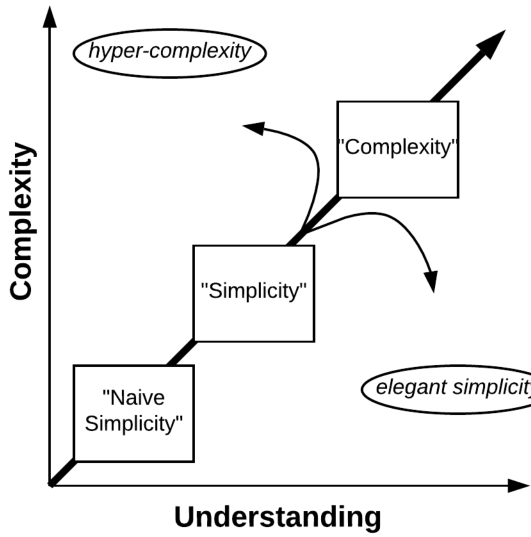


Figure 1. Adaptation of the simplicity cycle by Schwartz et al., 2017

2.3 Models and complexity in environmental planning and management

The choice in level of complexity – and, by extension, the nature and approach of a model emulator, must be matched to the intended application, and this is especially true in the context of environmental planning and management (Haasnoot et al., 2014; Voinov et al., 2018). In environmental decision-making, many "boundaries" must be crossed-- between disciplines, professions, stakeholder interests, geographies, and different forms of knowledge (Cash et al., 2002). Linking the knowledge contained in stakeholders' mental models, is an exercise of finding the types information that are salient, credible, and legitimate across these boundaries (Cash et al., 2002). In this context, models are "boundary objects" that provide a common terminology, conceptual framework, and platform for necessary negotiations and communication to occur (Falconi and Palmer, 2017; White et al., 2010). The process of acquiring knowledge that is salient, credible, and legitimate to the stakeholders in environmental problem-solving is can be though of as negotiating the interface between environmental science and action.

Modeling with stakeholders in at the environmental science-action interface is a key step to defining and framing socio-ecolocial problems and establishing trust and acceptance in the process of environmental management. Voinov et al., (2018) define *participatory modeling* as "a purposeful learning process for action that engages the implicit and explicit knowledge of stakeholders to create formalized and shared representations of reality" (p 233). There are three major rationales for encouraging citizens to participate in model-building in the context of environmental planning and management (Korfmacher, 2001). First, participation enables the co-production of knowledge, since stakeholders are likely to bring specific understanding of relevant social-environmental processes. Second, participation exemplifies democratic ideals of representation in decision-making and exemplifies "post-normal" response to the shortcomings of expert-driven, technocratic problem-solving ideals (Funtowicz and Ravetz, 1993). Third, participation is pragmatic; stakeholders who have been involved in environmental decision-making, are more likely to accept policy and cooperate in policy implementation (Lubell, 2004).

Similar to the cyclical process of model simplification, participatory modeling emphasizes the importance of iterative process of learning in "loops:" single loop learning refers to when stakeholders increase understanding within a single reference frame; in double loop learning, stakeholders may question initial assumptions, biases and beliefs, or reframe questions entirely. Models can aid in both single-loop and double-loop learning processes (Glynn et al., 2017; Pahl-Wostl et al., 2007; Zellner and Campbell, 2015). Stakeholder knowledge incorporation through both qualitative and numerical modeling has been shown to enhance the richness of environmental problem solving (Elsawah et al., 2015).

Environmental models are frequently used to aid scenario testing, an important means of learning in environmental management and planning. Scenario analysis encourages "stretching" the realms of possibility for users (Goodspeed, 2020; Xiang and Clarke, 2003). In environmental planning, there is a long tradition of the use of scenarios for "what if" analyses (Klosterman, 1999). In these situations, models act as decision support systems that allow stakeholders to engage in learning through deductive scenario analysis. Stakeholders' direct interaction with decision support systems have been shown to increase incorporation of diverse viewpoints when levels of detail and complexity are managed appropriately (Arciniegas and Janssen, 2012).

Given the importance of models and group learning in the processes of environmental management discussed above, model emulators may have the following important roles to play:

- Model emulators may facilitate stakeholder engagement through lowering computational barriers to scenario testing and encouraging deductive learning
- Model emulators may increase trust in models as boundary objects by reducing barriers to stakeholders' direct interactions with the model
- The process of model emulation may provide the opportunity to learn about the watershed itself
- The process of model emulation may provide the opportunity to learn about the processes of management
- The process of model emulation may provide the opportunity to learn about the process of model simplification

In this study, I explore these roles and their interrelationships through analyzing the reflections of a group of experienced watershed modelers on the topic of emulation modeling.

3 Methods

3.1 Chesapeake Bay Program Background

The Chesapeake Bay Watershed is known as being one of the most studied, most instrumented, and most-modeled watersheds in the world (Ernst, 2003; Layzer, 2011). Its drainage area spans six states on the East Coast of the United States, and includes decades of human impacts, including urbanization, agricultural intensification and deforestation. Despite decades of efforts to improve water quality in the Chesapeake Bay, both an ecological and natural resource national treasure, progress has been slow. Central to the management activities in the Bay are a suite of CBP environmental models that identify the sources, spatial distribution, transport, and ultimate fate of sediment and nutrients in the watershed. The watershed model is one of the models in this suite of models, and **Figure 2** shows how the watershed model relates to data and model inputs and ultimately to prediction on impacts to water quality in the Bay.

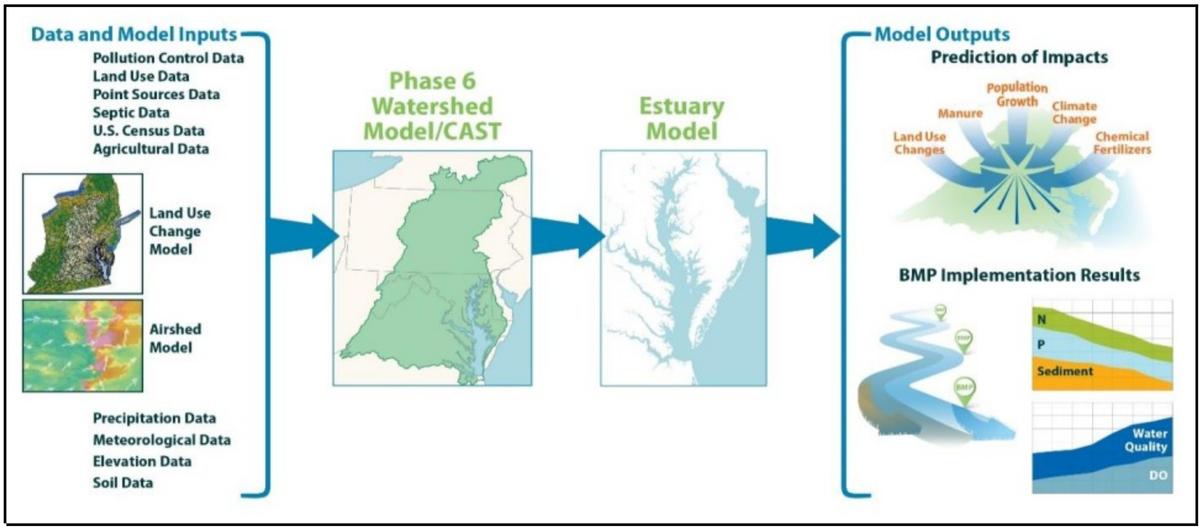


Figure 2. Relationship between the CBP Watershed Model to data and modeled inputs and predicted outputs. Source: Chesapeake Bay Program

There are four purposes of the CBP Watershed Model: (1) to estimate changes in nutrient and sediment load resulting from management actions; (2) to delivery estimated loads to the estuarine model; (3) for calibration and validation against other lines of evidence from collected from monitoring and models of the watershed; (4) for scientific study (*Chesapeake Bay Program Phase 6 Watershed Model: Final Model Documentation for the Midpoint Assessment: Section 1 – Overview*, 2018).

In its latest release, (Phase 6, released in 2017), the CBP watershed model underwent a dramatic change in modeling approach. Reversing a 20-year trend of ever-increasing spatial and temporal resolution and more detailed process representation in the watershed model (Linker et al., 2002; Shenk and Linker, 2013), the Phase 6 model release instead opted for a lower-resolution, simplified approach. The Phase 6 model coincided with the 2017 Midpoint Assessment for the Bay's progress on achieving its Total Maximum Daily Load (TMDL) goals. The Bay's TMDL, passed in 2010, governs how its "pollution diet" is allocated to each of the six contributing states and local jurisdictions. Watershed managers are required to prepare Watershed Implementation Plans projecting plans for land use change and implementation of best management practices to lower point and non-point source contributions of nitrogen, phosphorous, and sediment.

3.2 Data Collection

The primary data used in this study was original qualitative research drawn from focus groups conducted with a core group of modelers involved with the Chesapeake Bay Program (CBP). The transcriptions of the focus groups were supplemented with document analyses of peer-reviewed journal articles authored by members of this core group related to CBP watershed modeling, and key presentation slides and reports outlining the approach taken in Phase 6 of the watershed model. Transcripts and documents were analyzed for themes along with the other collected documents using a two-stage coding process (Charmaz, 2014). In the first stage, open themes were coded to allow for the emergence of new ideas, with the intention to form a conceptual framework for emulation modeling within the context of environmental management, "grounded" in the collected data. In the second stage, axial codes were used to group ideas into larger categories and draw connections among codes. Coding was performed using the software Nvivo 12 Professional for Windows.

Two focus groups were facilitated via online Zoom meetings. Each of a total of seven participants, all of whom are associated with a core group of watershed modelers in the CBP, attended one of the two meetings. Table 1 shows participants' titles and number of years of involvement with the CBP. At the beginning of the meeting, the group was shown a 10-minute pre-recorded video describing the concept of model emulation, and demonstrating the performance of a deep neural network emulator of the process-based hydrological model, ParFlow.CLM. The purpose of the presentation was two-fold. First, I wanted to demonstrate one motivation for model emulation: faster simulation time. Based on previous documentation of the experiences of the CBP, I knew that participants in the focus group would be familiar with the underlying motivations for speeding scenario testing, even if they had not previously referred to that work as "emulation modeling."

Table 1. Focus Group Participants' titles and involvement in the CBP

Position	Number of Years involved in CBP
Contractor to CBP	13
Manager of land use/land cover data inputs	18
Research scientist with CBP	1
Developer of watershed models	25
Modeling coordinator	35
Scenario analysis coordinator	22
Research scientist with CBP	9

Second, the particular kind of emulation shown in the demonstration was an extreme case of response surface emulation. It used deep neural networks to reproduce spatially resolved groundwater variability and hourly time-step surface runoff from a hypothetical site simulated with ParFlow.CLM, a high-resolution, coupled surface-subsurface hydroloogical model (Lim, in review) with high fidelity. The demonstration represented an "extreme case" because of its performance; it reduced ParFlow simulation times from hours/days to a fraction of a second and did not lower the resolution of spatial or temporal outputs. It was also an extreme case because of its basis in deep neural networks – a statistical method that is notoriously abstracted from physical meaning and difficult to interpret. While many emulation models, including the original emulator developed by the CBP team, might represent a mix between lower fidelity surrogate models and response surface emulators, the emulator in the demonstration is closer to a "pure" response surface emulator. Viewing of the demonstration was itself meant to act as a "stretch" scenario to prompt discussion in the focus group, to help them consider the implications of emulation modeling in decision-making processes (Xiang and Clarke, 2003).

Following the 10-minute demonstration, each of the focus groups was asked to relate what they had seen in the demonstration to their own experiences with the CBP. Participants in the group discussed their experiences, filling in gaps in recollection and offering different perspectives. I interjected only to ask clarifying questions of the group, or to solicit response from participants who were less active in the conversation (Given, 2008).

4 Results and Discussion

4.1 Reflections on DNN-emulator demonstration

After viewing the demonstration of the training and performance of the DNN-based emulator of ParFlow.CLM (see Lim in review, for more details), participants immediately drew parallels with their own experiences developing the CBP's watershed model. One participant's first comment in response to the presentation was:

> *Your example is exactly what we did with CAST. CAST was first developed in 2011 and it was an emulation of the time-variable watershed model at that time… I'm not familiar with the term 'deep learning' [from the presentation] but the way we did it is, I ran a bunch of regressions to estimate what the results would be under different input conditions.*

As the focus group conversation continued however, participants began to distinguish the current version of CAST from the emulator shown in the demonstration. The original CAST (then an abbreviation of Chesapeake Approximation Scenario Tool) released in 2011 was developed to allow stakeholders to more quickly create and test scenarios for planning purposes, but for "official" results, model output still needed to come from the original time-variable watershed model. In Phase 6 of the watershed model released in 2017, however, CAST (which now stands for Chesapeake Assessment Scenario Tool) is no longer a approximation of the watershed model, but has itself been adopted as the official watershed model, and its emulation approach is also very different. The Phase 6 CAST model is no longer a response surface emulator, but rather, more similar to a lower fidelity surrogate, with the parameters most relevant to management actions preserved. In particular, the Phase 6 CAST model output is at the annual temporal scale, while the time-variable model, which is still used for scientific research purposes, has an hourly temporal resolution.

Several participants used the terminology "management model" to reflect the current purpose of CAST, distinguishing it from the previous response surface emulator. For example:

> *I don't think I call [CAST] an emulator model. I just call it a management model.*

> *[The CBP] realized that what people needed was a management model, not a scientific model. A management model is really just for accounting purposes.*

Another participant still considers the current CAST an emulator, but distinguished it from the DNN-based emulator shown in the presentation because of the increased, not decreased, interpretability of the lower resolution surrogate:

AUTHOR’S PRE-PRINT. See published paper here:
https://doi.org/10.1016/j.envsoft.2020.104928

> *In my view, CAST is an emulator, but it's not the kind of emulator [that was shown in the presentation]. It's not a deep learning model that comes in and takes out all of the processes. So there's still a lot of interpretability. A lot of all of the [explanations of the] statistical models that go in to building this emulator are incredibly detailed.*
>
> *Stakeholders were tremendously involved in all of the inputs to the Phase 6 model. And so it's absolutely not a black box. The reason for Phase 6 is that it's wide open and not a not of black box at all. [While the DNN-based emulator in the presentation was characterized as a black box.]*

For the purposes of this study, I consider both the response surface emulator and lower resolution surrogate approaches to emulation models, and therefore refer to both the original CAST and the Phase 6 CAST as “emulators.”

4.2 Purpose of emulation in watershed management

Three purpose themes for emulation emerged from the focus groups: improving ***p****erformance*, ***p****erception* and capacity to act as a ***p****latform* (3 Ps). Counting the number of mentions from the focus group transcripts devoted to each of these themes shows that these three purpose themes were represented relatively evenly. **Figure 3** shows the frequency of codes for each purpose theme and combinations of purpose themes in the focus group transcripts.

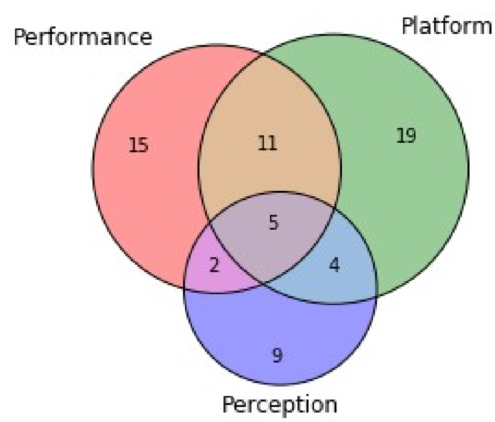


Figure 3: “Three P’s” of model emulator purpose: code frequency found in focus group transcripts. Overlapping areas indicate text coded in multiple themes. Although performance improvement (speed) was the initial motivation, capability to serve as a platform and positive perception by stakeholders became major goals of the emulation model’s development.

4.2.1 Performance

Improving performance, specifically, the speed of the model run, was cited as the first reason for the development of the original CAST model in 2011 (and other emulators that preceded CAST as well):

> *At the time that [CAST] was first developed it was taking the states about three weeks to develop an input deck, submit it to the Bay Program... and get the results back. It was very difficult for the states to do any kind of planning when you have to wait for the results for one scenario for three weeks.*

The 2011 CAST model reduced simulation time from three weeks (including all pre-processing and post-processing of inputs and outputs) to nearly instantaneous (Deveroux and Rigelman, 2014). Participants also discussed *Performance* of the original CAST model in terms of how well it could mimic the original model's output:

For all land uses, the agreement was within 10% of the Watershed Model loads in 95% of the modeling segments. The error was predominantly on agricultural land uses in small number of land–river segments. This high level of agreement with the Watershed Model indicates that CAST is a valuable and useful tool for rapid assessment of load reductions given varying BMP inputs (Devereux and Rigelman, 2014)

4.2.2 Platform

As the conversation continued however, participants began to mention features of CAST that extended beyond reduced simulation times and how close results were to the original model's output. Because the CAST online user interface was designed to meet non-modelers' needs, the emulator as a platform initially focused on how its outputs were summary reports themselves that were maximally relevant to decision-makers and the kinds of scenarios they would be testing. Managers could use the online platform in ways that were salient to their obligations to reduce nutrient and sediment loads in their jurisdiction areas.

> *We had the emulator-- the CAST management model--for a purpose. That was the specific purpose of making a complex model accessible to managers so that they can own the model in a sense, they could operate the model they could do something that would be otherwise complex and unmanageable, put it on their desktop and see the results. It was because of the scale of the management decisions that were being asked...managers [needed] to be able to get their hands on this thing and to run their own scenarios.*

4.2.3 Perception

In addition to improving simulation time (*Performance*) and providing an online hub for scenario testing (*Platform*), *Perception* also emerged as a major theme, both in motivating the emulator, and later, in evaluating its utility. Concerns with the transparency and interpretability (*Perception*) of the original model's output were a major motivator for having a simplified online platform where non-modelers could access and interpret model output:

> *In the mid 2000s and earlier… [modelers] always wanted to be really helpful and they'd share all the data that anybody ever asked for, send[ing] text files to people who asked for the data.*

> *[But] there was a different text file for every land river segment in the Chesapeake Bay watershed. It was direct output using Fortran [a programming language], that didn't have headers and nobody knew what it was. People would get angry. There's a way to use Fortran code to marry all that together and I think the assumption was: we'll send it all to you and that [you would know] how to put all those files together. I don't think the Bay Program meant to obfuscate any of the information, they just weren't thinking about it from a user perspective and it sent the impression that they were trying to hide something, which wasn't their intention at all.*

Having an online interface where managers were in complete control of the process themselves helped them access the results of decades of modeling advances. However, even after the online interface simplified scenario testing and reporting, there continued to be issues with the perception of the emulator, which, at that time was a response surface emulator. There were times when the approximation differed significantly from the "real" watershed model's output:

> *We had an emulator model of Phase 5, [where] we had a situation where in [certain areas] the emulator model was pretty close, but then in other areas we had a lot of trouble, so the emulator kind of broke down.*
>
> *Some people were suspicious and they would run scenarios in CAST and then they submit the same input data to the Bay Program, [to] do a comparison.*

This discrepancy between the emulator and the original model contributed to efforts to revise the goals of the watershed model altogether.

4.2.4 Evolving purpose: from response surface emulator to lower fidelity surrogate

In 2012, following generally positive feedback from users about the usefulness of the response surface emulator and how it was aiding in scenario testing of management decisions, the CBP began to advance plans to simplify the watershed model itself. The first two slides extracted from a 2012 presentation given by one of the participants to the CBP's Science and Technical Advisory Committee (STAC), "Partnership Priorities for 2017 and a Suggested Watershed Modeling Approach" (**Figure 4)** illustrate all three purpose themes for the emulation model. In the first slide is a quotation from a state government representative that "we [as non-modelers] want to be able to explain the models to our stakeholders and have them be relevant at the local scale." The same slide also includes elements of the 3 Ps of emulator purpose. *Performance* is reflected through "simplicity" and "scalability"; *Platform* is reflected in "scalability" and "ease of use"; and *Perception* is reflected in "understandability" and "ease of use." The second slide illustrates what is meant by "understandability" and how this is related to both procedural fairness and transparency of the model output. The first explanation, an answer to the question "What's my reduction from Nutrient Management?" is at least eight lines long, and fades into an ambiguous blur at the end, representing the answer's lack of salience from the user's perspective, despite the question's importance to stakeholders and managers. The second explanation is reduced to three easily understandable factors: crop type, timing, and geographic region.

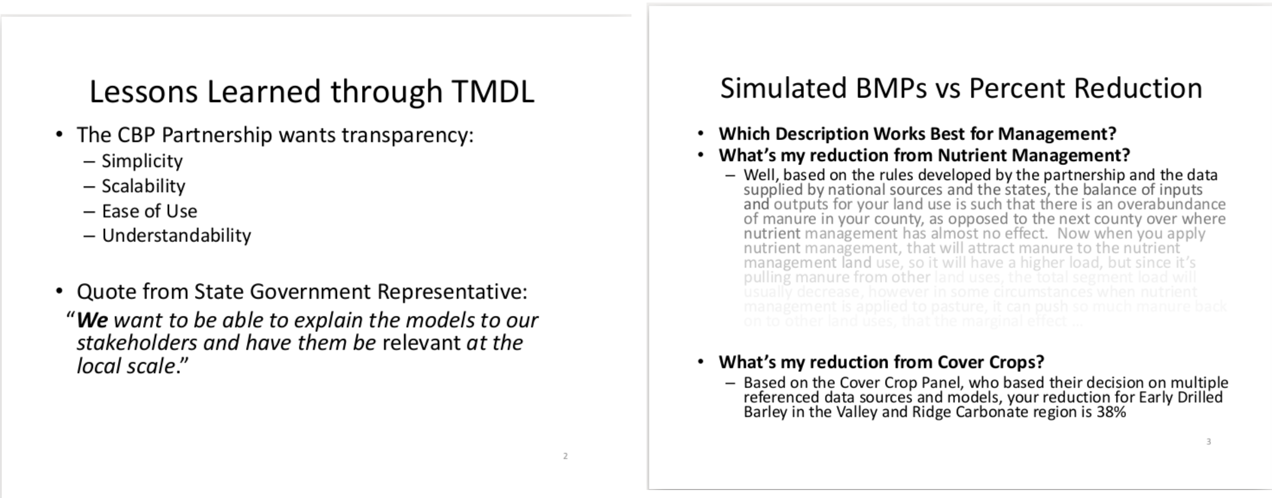


Figure 4. First two slides excerpted from a 2012 presentation given from the CBP watershed modeling team to the CBP STAC.

The purpose of emulation therefore evolved beyond an initial focus on reducing simulation run-times and improving the ability for non-modelers to run the model; reflecting on Phase 6 of the watershed model, participants emphasized perception and expanded platform functionalities.

The capability to act as a platform did not merely mean a platform for manager-users, but also for integrating diverse knowledge from all the extensive studies carried out over the decades in the Chesapeake Bay and involving the stakeholders themselves in the co-production of knowledge:

> *[Phase 6 CAST is] not emulating anything it's incorporating things. It's rolling out the knowledge that we think we have about the watershed. Modelers and academics know something about transport and so we get the inputs from them. The stakeholders, they know about their own land use and they know about their animals and the manure that they produce and the crops that they have and the yields that those crops have.*

One participant pointed out that acceptance of the simplified Phase 6 watershed model was contingent on the trust (*Perception*) cultivated over decades of engagement between stakeholders, committees, and modelers in the CBP.

> *The striking difference that I've seen, between [my previous] experience [in a different watershed] and my Chesapeake Program experience is the level of trust of the stakeholders… [CBP has] been working with stakeholders and partners for a long time... [the previous watershed I worked in] wouldn't have been able to go through the [emulation] process … because [stakeholders] needed to know everything that went behind that result. They were like: what kind of coefficient have you used? What kind of soil permeability have you used? … I think the Chesapeake Bay Program has the advantage of having built, over years, a level of trust that even if the stakeholders might not agree on the results there's much less of a concern*

> *about whether the model is solid or whether you need a detailed explanation of what's going on behind the scenes.*

Performance and platform themes were co-mentioned more frequently than other combinations of purpose themes (**Figure 3**). There were two major reasons for this. First, increased speed enabled the creation of an online tool that acts as a platform for users to develop and test their own scenarios, giving them the autonomy to interact with the model independently from the modeling team using their own desktop computers. Second, especially in the Phase 6 CAST, decreased complexity in process representation allowed for an opening up of the model, increasing its capacity to be linked to other models and integrate multiple lines of evidence and data sources. In this sense, the emulator itself becomes a platform to be coupled to other models, what one participant referred to as a "model of models."

> *In the process of emulating a complex model, it opens it up for it to be used in conjunction with other lines of evidence.*

> *[The transition to the Phase 6 model] opened it up for using other models. Not just process-based models, but also statistical models, and creating a new set of models.*

4.3 Learning themes

The evolving purposes that emerged from the focus group were accompanied by reflections that carried themes of "learning." Three major learning themes emerged: learning about the watershed, learning about the model, and learning about environmental management. Of these, learning about management was the theme that appeared most frequently from the transcripts (**Figure 5**).

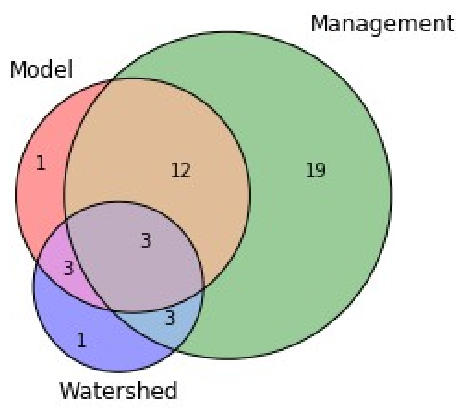


Figure 5 Learning themes: code frequency found in focus group transcripts. Overlapping areas indicate text coded in multiple learning themes. The majority of learning around the emulator had to do with management processes.

4.3.1 Managers learning about management options

Learning about management options by managers emerged as an important theme, given that initial purpose of the original CAST model was to enable watershed managers to more quickly test options to achieve water quality targets:

> *The management model [CAST] is really to do these what-if scenarios and a lot of that is related to the TMDL [Total Maximum Daily Load] in the Chesapeake Bay. It's: if I do this set of BMPs [Best Management Practices], in this region, what do I get? What is the best bang for the buck?*

An example of the deductive learning that comes from managers personally being able to interact with the model to develop their own scenarios included:

> *[Managers] are actually ending up with much better products because they can do many more what-if scenarios than what we could do at the office ourselves running it.*

4.3.2 Modelers learning about modeling

There were also examples of how modelers' changed their perspectives of the role of the models in the environmental management context. For example, when the initial CAST response surface emulator was developed, the original time-variable model was viewed as the "truth" that the emulator was attempting to approximate. Later, modelers learned that the higher resolution, more complex model was not necessarily "better" than a simpler model.

> *When CAST was an emulator model, we conceived of [the] Phase 5 [dynamic watershed model] as something like a research model. And, not that we consider that to be the truth but we consider that to be the best estimate. And then the emulator model was trying to get close to that best estimate but when we really needed a management decision, we had to go back to that best estimate which was coming from the research model. We had some problems with that conceptualization of modeling. Because we were modeling 200 different counties in the watershed and 30 different land uses in each one, we had 6,000 simulations that were going on. And in some locations, since this was a complex model, they came up with very unreasonable calculations and so we spent many months on a problem with phosphorus in pasture that was essentially caused by instability in the dynamic model.*

Participants recognized that "truth" instead should be compared to monitoring data:

> *We got rid of this way of looking at it where we had a research type model and then an emulator model, but instead went to producing a simple model on which a lot of other models went into that we found that we have actually had much better spatial Nash Sutcliffe Efficiencies [a*

> *measure or model performance, compared to monitored data] than when we used the previous model.*

This insight was then related back to the demonstration of the DNN-based emulator of ParFlow.CLM, which did not present any connection to a “real” watershed’s monitoring data. The ParFlow.CLM simulations were based on a hypothetical watershed, and the goal of the DNN emulator was merely to reproduce the modeled output from the original ParFlow.CLM model as best as possible:

> *Relating this back to your presentation, I was a little uncomfortable with treating the research model as the ideal that the emulator model is trying to get to.*

4.3.3 Modelers learning about management

Participants brought up aspects of both the model and its input data that were significant because of the environmental management context. One problem that resulted from the instability of the original dynamic model described above was the implication that model errors had on the fairness of allocating necessary management actions:

> *If you're looking over lots and lots of scenarios, lots and lots of places, you're like, well [the errors introduced by the dynamic model are within the] acceptable tails of a good simulation. But for a management model, somebody lives in that county, someone owns a pasture in that county, and they are getting very unfairly tagged with really high [pollutant] loads because of instability of a complex model.*

Another participant brought up realizing the importance of stakeholder perceptions of the data input into the model. Specifically, they reflected on learning the importance of having high resolution land cover data that matches with how stakeholders perceive the landscape, even though from a traditional model evaluation perspective, increasing data resolution did not significantly improve model outputs:

> *With the availability of Google Earth... people know what the landscape looks like from above and then when they see a coarse version of that in a gridded format, 30 meter resolution LANDSAT interpreted imagery... it doesn't look like things they're familiar with...[high resolution data do not] necessarily improve the calibration but they improve the credibility of the model...That was one of the reasons why the [Chesapeake Bay] Partnership invested over a million dollars in getting us [1 meter resolution land cover] data.*

4.3.4 Learning about the watershed through the model simplicity cycle

By simplifying the model, CAST could act as a platform for more diverse types of co-produced knowledge (from both experts and stakeholders), as well as act as a platform for the incorporation of other models. Opportunities to incorporate more diverse sources of data and models contributed to better understanding of the watershed itself.

> *CAST did rely on a complex model, which was the Phase 5.3.2 model but that was not the only aspect of it. [Emulation] opened it up for using other models. Not just process-based models, but also statistical models, and creating a new set of models.*

Now that the simplified, lower resolution surrogate, Phase 6 CAST has been officially adopted however, discussions of what is included in the model is continuing to evolve. Already, participants recounted how the “simplicity cycle” is pushing toward including more detail. Now however, the increased complexity is driven not by modelers and researchers, but by the stakeholders themselves, through their interactions with CAST:

> *Creating a mass balance for manure application started with counting the animals and ended up with applications to streams to crops and storage and handling losses and things like that. The stakeholders really determined how all those counts happened. And they made that model so much more complex and a lot of the difficulties that we're having now has to do with the complexity of the model of farmer behavior that the stakeholders built*
>
> *Over time, it's interesting how the sophisticated users want to be able to change things that they can't do easily any more. Like change the rate of applications of manure fertilizers, something like that. So it becomes a matter of if we have enough users who want to add [a new] capability, but you have to always weigh that against making the tool more complicated for users as a whole.*

4.4 Connections between purpose themes and learning themes: complexity management at the science action interface

Lastly, I tabulated frequencies of purpose and learning themes within the focus group transcripts to better understand how the evolving purpose of the emulator related to the different kinds of learning that occurred from the release of the initial CAST emulator in 2011 to today. **Figure 6** shows a conceptual diagram with the strength of the connections between paired purpose-learning themes represented by the thicknesses between red double headed arrows. Thicknesses of the double headed arrows are based on the total number of words coded in the purpose-learning theme pairs. I used word counts over codes counts because I noticed that the long segments in the transcripts `represented “importance” of the connection but may have only been marked as a single code. Using code counts would have artificially deflated the importance of these segments.

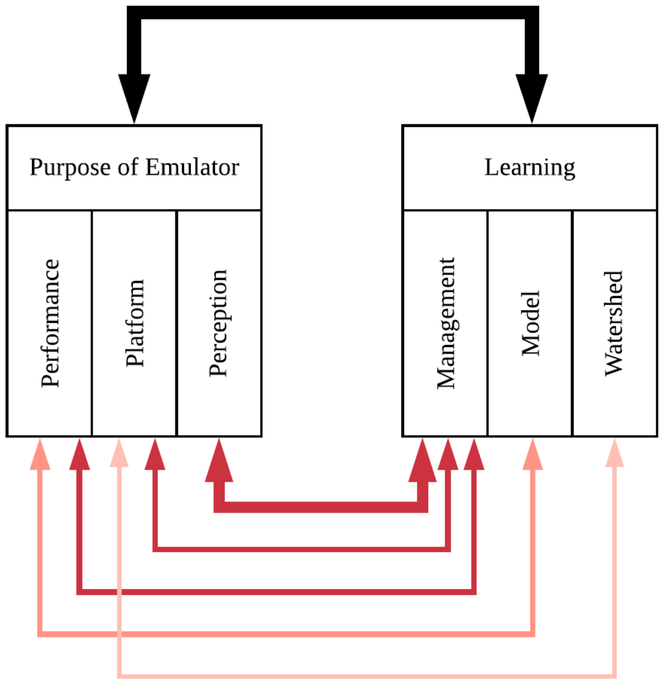


Figure 6. Connections between purpose themes and learning themes. Within the context of watershed management, emulation is an iterative process involving the linked development of purposes for the emulator and learning about management, the model, and the watershed. Darker red, and thicker connections indicate linkages in the themes discussed by the focus group participants.

I found that the most important connections occurred between "Perception" and "Learning about Management" (777 words), followed by "Performance" and "Learning about Management" (706 words), "Platform" and "Learning about Management" (705 words), "Performance" and "Learning about Model" (625 words), and "Platform" and "Learning about Watershed" (475 words). Other pairs had less than 400 words.

In complex socio-ecological management problems, cultivation of stakeholder trust governance of information is extremely important. In this study, the *Perception* purpose theme included preliminary codes such as: "trust," "credibility,""interpretability," "explainability," "optics," "involvement," and "inclusion." The example given above of the rationale behind the CBP's investment in high resolution land cover data because of stakeholder perceptions of their own landscapes, and despite evidence that this data would not significantly improve the calibration of the model is a particularly good example of the relationship between perception and learning about socio-ecological management. Another participant referred to some of parts of the integration of other models into the simplified Phase 6 CAST model as important to increasing the confidence of managers and justifying potential decisions:

> *Within the CAST management model we were able to specifically point to things and say: "See? There, this is where that study came in", or, "This is where that other modeling effort came in."*

> *And so it was very much in that sense maybe more optics in terms of being able to specifically point to a thing and say "Yeah you see this is this is not one model but this is a model of models that's getting to this end point that you're using to make your own decisions", which are often, fantastically expensive to these localities and states.*

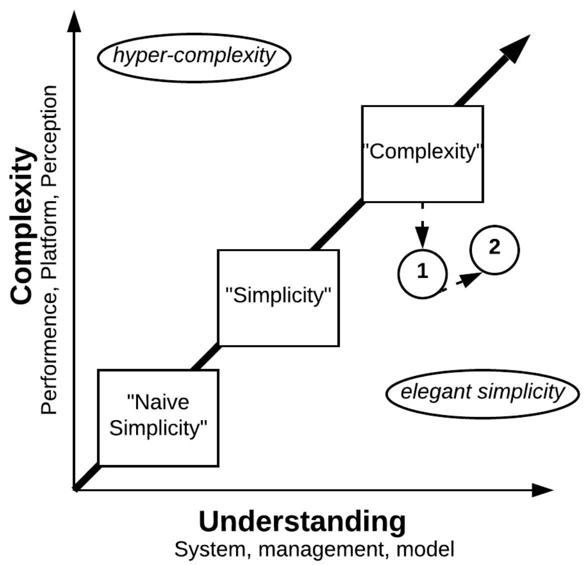
Perhaps in contrast to the traditional cycle of simplification, in environmental management contexts, stakeholder perception plays an important role in determining what "stays in" elegantly simplified models.

It is also important to note that the learning themes are most prominently connected to purpose themes through learning about management. Shorter model run-times (performance) enabled learning about management. Viewing the emulator as a platform also connected to the realization that management requires the incorporation of diverse sources of knowledge, data, models, and research findings. This is a reflection of where the CBP is in its history. While much of the past thirty years was focused on refining scientific understanding of the causes of the degradation of the Chesapeake Bay, the passage of the Bay TMDL and the midpoint assessment in 2017 shifted focus from science to action. The motivations for model emulation were a direct recognition to facilitate better management decisions and learning about management options. Initially, this focused on improving model run times, but also on viewing the model as a platform, and on the roles of preserving positive perceptions of the model and its input data to encourage action.

One motivation for using DNN-based emulators for this research was to show how they might be attractive for their ability preserve high resolution outputs as response surface emulators. However, through the focus groups, participants reflected that high resolution outputs were not expected to be important to the majority of the users in the case of the Chesapeake Bay watershed. Participants attributed the reasons for this to the requirements of water quality planning. Temporally, resolutions higher than the annual scale are not required. Spatially, some managers might be interested in modeling scenarios for smaller scales (for example, field-scale), but currently, the barrier to such applications is a lack of data at smaller scales to support the modeling.

Revisiting the "simplicity cycle" diagram discussed in the Background section, the role of emulation modeling in the CBP is illustrated in **Figure 7**. This figure includes adaptations specific to the science-action interface. The "complexity" axis explicitly acknowledges that model complexity may incorporate performance-related, platform-related or perception-related complexity. The "understanding" axis includes understanding not only of the system itself, but of management of the system and of the models used in environmental science and management. The original 2011 CAST surface response emulator's primary goal was to reduce the computational (performance) complexity of the model (position 1 in the figure). The 2017 Phase 6 CAST has seen some increases in platform functionality and in data inputs. But, increased understanding of watershed management and roles of the models in decision-making have also increased.

AUTHOR'S PRE-PRINT. See published paper here:
https://doi.org/10.1016/j.envsoft.2020.104928



5 Conclusions

In this research, I explored the concept of model emulation at the environmental science-action interface through analyzing watershed modelers' reflections on a presentation of an extreme case of one kind of emulator, a deep learning-based response surface emulator. Participants were able to relate their experiences in model development in one of the most well-known watershed management cases in the world, the Chesapeake Bay Watershed. Their responses revealed that emulation played a key role in complexity management at the environmental science-action interface. After a twenty-year trend of increasing model complexity in the Chesapeake Bay, the formalization of regulatory requirements through the Bay's TMDL document shifted emphasis on facilitating planning and management actions from diverse state, local, and community actors. Non-modelers' ability to use the model on their own to test "what if scenarios" motivated the original emulation models, starting in the mid-2000s, including the first version of CAST, in 2011.

I showed that the response surface emulator approach to the first CAST was an improvement in accessibility to non-modeler users, however, continued learning about the nature of environmental management eventually motivated a shift in emulator approach to something more like a lower resolution surrogate. Three purpose themes emerged from the participants' reflections, the "3Ps": (1) performance, (2) platform, and (3) perception. Three learning themes also emerged: learning about (1) management, (2) modeling, and (3) the watershed itself. The most frequent connections between "purpose" and "learning" themes were between perception and learning about management. These reflections underscored the importance of stakeholder trust in the model, how the model and its input data were being perceived, and the "optics" of model complexity management. This implies that although advances in deep/machine learning could theoretically result in advances in performance in emulation models, perceptions by stakeholders and lay users could limit the utility of such emulators at the science-action interface.

Acknowledgements
Funding: This work was supported by the Virginia Tech Institute for Society, Culture and Environment. Special thanks to Gary Shenk and Lewis Linker for their assistance on this project.

AUTHOR'S PRE-PRINT. See published paper here:
https://doi.org/10.1016/j.envsoft.2020.104928

References

Arciniegas, G., Janssen, R., 2012. Spatial decision support for collaborative land use planning workshops. Landscape and Urban Planning 107, 332–342. https://doi.org/10.1016/j.landurbplan.2012.06.004

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Advances in Water Resources, Research Perspectives in Hydrology 16, 41–51. https://doi.org/10.1016/0309-1708(93)90028-E

Carnevale, C., Finzi, G., Guariso, G., Pisoni, E., Volta, M., 2012. Surrogate models to compute optimal air quality planning policies at a regional scale. Environmental Modelling & Software, Emulation techniques for the reduction and sensitivity analysis of complex environmental models 34, 44–50. https://doi.org/10.1016/j.envsoft.2011.04.007

Cash, D., Clark, W.C., Alcock, F., Dickson, N.M., Eckley, N., Jäger, J., 2002. Salience, Credibility, Legitimacy and Boundaries: Linking Research, Assessment and Decision Making (SSRN Scholarly Paper No. ID 372280). Social Science Research Network, Rochester, NY.

Castelletti, A., Galelli, S., Ratto, M., Soncini-Sessa, R., Young, P.C., 2012. A general framework for Dynamic Emulation Modelling in environmental problems. Environmental Modelling & Software, Emulation techniques for the reduction and sensitivity analysis of complex environmental models 34, 5–18. https://doi.org/10.1016/j.envsoft.2012.01.002

Charmaz, K., 2014. Constructing Grounded Theory, Second edition. ed. SAGE Publications Ltd, London ; Thousand Oaks, Calif.

Chesapeake Bay Program Phase 6 Watershed Model: Final Model Documentation for the Midpoint Assessment: Section 1 – Overview, 2018.

Chwif, L., Barretto, M.R.P., Paul, R.J., 2000. On simulation model complexity, in: 2000 Winter Simulation Conference Proceedings (Cat. No.00CH37165). Presented at the 2000 Winter Simulation Conference Proceedings (Cat. No.00CH37165), pp. 449–455 vol.1. https://doi.org/10.1109/WSC.2000.899751

Clement, T.P., 2011. Complexities in Hindcasting Models—When Should We Say Enough Is Enough? Groundwater 49, 620–629. https://doi.org/10.1111/j.1745-6584.2010.00765.x

Condon, L.E., Maxwell, R.M., 2014. Groundwater-fed irrigation impacts spatially distributed temporal scaling behavior of the natural system: a spatio-temporal framework for understanding water management impacts. Environ. Res. Lett. 9, 034009. https://doi.org/10.1088/1748-9326/9/3/034009

Devereux, O., Rigelman, J.R., 2014. CAST: An Online Tool for Facilitating Local Involvement in Watershed Implementation Plans for the Chesapeake Bay Total Maximum Daily Load. JWMM. https://doi.org/10.14796/JWMM.C364

Elsawah, S., Guillaume, J.H.A., Filatova, T., Rook, J., Jakeman, A.J., 2015. A methodology for eliciting, representing, and analysing stakeholder knowledge for decision making on complex socio-ecological systems: From cognitive maps to agent-based models. Journal of Environmental Management 151, 500–516. https://doi.org/10.1016/j.jenvman.2014.11.028

Ernst, H.R., 2003. Chesapeake Bay Blues: Science, Politics, and the Struggle to Save the Bay, 1St Edition edition. ed. Rowman & Littlefield Publishers, Inc., Lanham.

Falconi, S.M., Palmer, R.N., 2017. An interdisciplinary framework for participatory modeling design and evaluation—What makes models effective participatory decision tools? Water Resources Research 53, 1625–1645. https://doi.org/10.1002/2016WR019373

Fatichi, S., Vivoni, E.R., Ogden, F.L., Ivanov, V.Y., Mirus, B., Gochis, D., Downer, C.W., Camporese, M., Davison, J.H., Ebel, B., Jones, N., Kim, J., Mascaro, G., Niswonger, R., Restrepo, P., Rigon, R., Shen, C., Sulis, M., Tarboton, D., 2016. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. Journal of Hydrology 537, 45–60. https://doi.org/10.1016/j.jhydrol.2016.03.026

Fletcher, S., Strzepek, K., Alsaati, A., Weck, O. de, 2019. Learning and flexibility for water supply infrastructure planning under groundwater resource uncertainty. Environ. Res. Lett. 14, 114022. https://doi.org/10.1088/1748-9326/ab4664

Forrester, A., Sobester, A., Keane, A., 2008. Engineering Design via Surrogate Modelling: A Practical Guide. John Wiley & Sons.

Funtowicz, S.O., Ravetz, J.R., 1993. The Emergence of Post-Normal Science, in: Von Schomberg, R. (Ed.), Science, Politics and Morality. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-015-8143-1

Given, L., 2008. Interactive Focus Groups, in: The SAGE Encyclopedia of Qualitative Research Methods. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States. https://doi.org/10.4135/9781412963909.n221

Glynn, P.D., Voinov, A.A., Shapiro, C.D., White, P.A., 2017. From data to decisions: Processing information, biases, and beliefs for improved management of natural resources and environments. Earth's Future 5, 356–378. https://doi.org/10.1002/2016EF000487

Goodspeed, R., 2020. Scenario Planning for Cities and Regions: Managing and Envisioning Uncertain Futures. Lincoln Institute of Land Policy.

Haasnoot, M., van Deursen, W.P.A., Guillaume, J.H.A., Kwakkel, J.H., van Beek, E., Middelkoop, H., 2014. Fit for purpose? Building and evaluating a fast, integrated model for exploring water policy pathways. Environmental Modelling & Software 60, 99–120. https://doi.org/10.1016/j.envsoft.2014.05.020

Hamilton, S.H., Fu, B., Guillaume, J.H.A., Badham, J., Elsawah, S., Gober, P., Hunt, R.J., Iwanaga, T., Jakeman, A.J., Ames, D.P., Curtis, A., Hill, M.C., Pierce, S.A., Zare, F., 2019. A framework for characterising and evaluating the effectiveness of environmental modelling. Environmental Modelling & Software 118, 83–98. https://doi.org/10.1016/j.envsoft.2019.04.008

Hong, E.-M., Pachepsky, Y.A., Whelan, G., Nicholson, T., 2017. Simpler models in environmental studies and predictions. Critical Reviews in Environmental Science and Technology 47, 1669–1712. https://doi.org/10.1080/10643389.2017.1393264

Karpatne, A., Atluri, G., Faghmous, J.H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., Kumar, V., 2017. Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. IEEE Transactions on Knowledge and Data Engineering 29, 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168

Kirchner, J.W., 2009. Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. Water Resources Research 45, n/a–n/a. https://doi.org/10.1029/2008WR006912

Klosterman, R.E., 1999. The What If? Collaborative Planning Support System. Environ Plann B Plann Des 26, 393–408. https://doi.org/10.1068/b260393

Korfmacher, K.S., 2001. The Politics of Participation in Watershed Modeling. Environmental Management 27, 161–176. https://doi.org/10.1007/s002670010141

Layzer, J.A., 2011. The Environmental Case: Translating Values Into Policy, 3rd ed. CQ Press, Washington, D.C.

Lee, D.B., 1973. Requiem for large-scale models. Journal of the American Institute of Planners 39.

Lim, T.C., Welty, C., 2017. Effects of spatial configuration of imperviousness and green infrastructure networks on hydrologic response in a residential sewershed. Water Resources Research 53, 8084–8104. https://doi.org/10.1002/2017WR020631

Linker, L.C., Shenk, G.W., Wang, P., Hopkins, K.J., Pokharel, S., 2002. A SHORT HISTORY OF CHESAPEAKE BAY MODELING AND THE NEXT GENERATION OF WATERSHED AND ESTUARINE MODELS. proc water environ fed 2002, 569–582. https://doi.org/10.2175/193864702785665021

Little, J.C., Hester, E.T., Elsawah, S., Filz, G.M., Sandu, A., Carey, C.C., Iwanaga, T., Jakeman, A.J., 2019. A tiered, system-of-systems modeling framework for resolving complex socio-environmental policy issues. Environmental Modelling & Software 112, 82–94. https://doi.org/10.1016/j.envsoft.2018.11.011

Liu, Y., Gupta, H., Springer, E., Wagener, T., 2008. Linking science with environmental decision making: Experiences from an integrated modeling approach to supporting sustainable water resources management. Environmental Modelling & Software 23, 846–858. https://doi.org/10.1016/j.envsoft.2007.10.007

Lubell, M., 2004. Resolving Conflict and Building Cooperation in the National Estuary Program. Environmental Management 33, 677–691. https://doi.org/10.1007/s00267-003-0066-6

Mahmoud, M., Liu, Y., Hartmann, H., Stewart, S., Wagener, T., Semmens, D., Stewart, R., Gupta, H., Dominguez, D., Dominguez, F., Hulse, D., Letcher, R., Rashleigh, B., Smith, C., Street, R., Ticehurst, J., Twery, M., van Delden, H., Waldick, R., White, D., Winter, L., 2009. A formal framework for scenario development in support of environmental decision-making. Environmental Modelling & Software 24, 798–808. https://doi.org/10.1016/j.envsoft.2008.11.010

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

Nearing, G., Kratzert, F., Sampson, A.K., Pelissier, C.S., Klotz, D., Frame, J., Prieto, C., Gupta, H., 2020. What Role Does Hydrological Science Play in the Age of Machine Learning? (preprint). EarthArXiv. https://doi.org/10.31223/osf.io/3sx6g

Oreskes, N., 2003. The role of quantitative models in science, in: Canham, C.D., Cole, J.J., Laurenroth, W.K. (Eds.), Models in Ecosystem Science. Princeton University Press, NJ, USA, pp. 13–31.

Pahl-Wostl, C., Craps, M., Dewulf, A., Mostert, E., Tabara, D., Taillieu, T., 2007. Social learning and water resources management. Ecology and Society 12.

Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., Wagener, T., 2016. Sensitivity analysis of environmental models: A systematic review with practical workflow. Environmental Modelling & Software 79, 214–232. https://doi.org/10.1016/j.envsoft.2016.02.008

Ratto, M., Castelletti, A., Pagano, A., 2012. Emulation techniques for the reduction and sensitivity analysis of complex environmental models. Environmental Modelling & Software, Emulation techniques for the reduction and sensitivity analysis of complex environmental models 34, 1–4. https://doi.org/10.1016/j.envsoft.2011.11.003

Razavi, S., Tolson, B.A., Burn, D.H., 2012a. Review of surrogate modeling in water resources: REVIEW. Water Resour. Res. 48. https://doi.org/10.1029/2011WR011527

Razavi, S., Tolson, B.A., Burn, D.H., 2012b. Numerical assessment of metamodelling strategies in computationally intensive optimization. Environmental Modelling & Software, Emulation techniques for the reduction and sensitivity analysis of complex environmental models 34, 67–86. https://doi.org/10.1016/j.envsoft.2011.09.010

Schwartz, F.W., Liu, G., Aggarwal, P., Schwartz, C.M., 2017. Naïve Simplicity: The Overlooked Piece of the Complexity-Simplicity Paradigm. Groundwater 55, 703–711. https://doi.org/10.1111/gwat.12570

Shen, C., 2018. A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. Water Resources Research 54, 8558–8593. https://doi.org/10.1029/2018WR022643

Shenk, G.W., Linker, L.C., 2013. Development and Application of the 2010 Chesapeake Bay Watershed Total Maximum Daily Load Model. J Am Water Resour Assoc 49, 1042–1056. https://doi.org/10.1111/jawr.12109

Van Nes, E.H., Scheffer, M., 2005. A strategy to improve the contribution of complex simulation models to ecological theory. Ecological Modelling 185, 153–164. https://doi.org/10.1016/j.ecolmodel.2004.12.001

Voinov, A., Gaddis, E.J.B., 2008. Lessons for successful participatory watershed modeling: A perspective from modeling practitioners. Ecological Modelling, Special Issue dedicated to the memory of Yuri Svirezhev 216, 197–207. https://doi.org/10.1016/j.ecolmodel.2008.03.010

Voinov, A., Jenni, K., Gray, S., Kolagani, N., Glynn, P.D., Bommel, P., Prell, C., Zellner, M., Paolisso, M., Jordan, R., Sterling, E., Schmitt Olabisi, L., Giabbanelli, P.J., Sun, Z., Le Page, C., Elsawah, S., BenDor, T.K., Hubacek, K., Laursen, B.K., Jetter, A., Basco-Carrera, L., Singer, A., Young, L., Brunacini, J., Smajgl, A., 2018. Tools and methods in participatory modeling: Selecting the right tool for the job. Environmental Modelling & Software 109, 232–255. https://doi.org/10.1016/j.envsoft.2018.08.028

Wagener, T., Liu, Y., Gupta, H.V., Springer, E., Brookshire, D., 2005. Multi-resolution integrated assessment and modelling of climate change impacts on water resources in arid and semiarid regions, in: Seventh IAHS Scientific Assembly. Presented at the Regional Hydrological Impacts of Climatic Variability and Change, Foz do Iguaçu, Brazil, p. 8.

Ward, D., 2015. The Simplicity Cycle: A Field Guide to Making Things Better Without Making Them Worse. Harper Business.

White, D.D., Wutich, A., Larson, K.L., Gober, P., Lant, T., Senneville, C., 2010. Credibility, salience, and legitimacy of boundary objects: water managers' assessment of a simulation model in an immersive decision theater. Sci. and Pub. Pol. 37, 219–232. https://doi.org/10.3152/030234210X497726

Williams, J.L., Maxwell, R.M., 2011. Propagating Subsurface Uncertainty to the Atmosphere Using Fully Coupled Stochastic Simulations. Journal of Hydrometeorology 12, 690–701. https://doi.org/10.1175/2011JHM1363.1

Xiang, W.-N., Clarke, K.C., 2003. The Use of Scenarios in Land-Use Planning. Environ Plann B Plann Des 30, 885–909. https://doi.org/10.1068/b2945

Zellner, M., Campbell, S.D., 2015. Planning for deep-rooted problems: What can we learn from aligning complex systems and wicked problems? Planning Theory & Practice 16, 457–478. https://doi.org/10.1080/14649357.2015.1084360