BioScience, 2026, **0**, 1–12
https://doi.org/10.1093/biosci/biag010
Advance access publication date: 0 2026
Communicating Science

The PAVE pathway: how to talk about models with people who built them, and people who didn't

Andrew Reid Bell [1], Christian Troost [2], Tatiana Filatova [3], Christian Vincenot [4], Cheng-Zhi Qin [5,6], Sebastian Fiedler [7], Carsten Lemmen [8], Iseult Lynch [9], Maria Pierce [10,11], Bruce Edmonds [12,13], Eliot McIntire [14], Hsiao-Hsuan Wang [15], Tomasz E. Koralewski [16], William E. Grant [16], Ming Wang [17], Ben Black [18,19] and Volker Grimm [17,20]

[1]Andrew Reid Bell (andrew.bell@cornell.edu) is the Schleifer Family Associate Professor of Sustainability at the Ashley School of Global Development and the Environment, Cornell University, Ithaca, NY, USA. [2]Christian Troost is a Senior Researcher and Lecturer at the Department of Land Use Economics and the Computational Science Hub, University of Hohenheim, Stuttgart, Germany. [3]Tatiana Filatova is a Full Professor at the Policy Analysis Section, Department of Multi-Actor Systems, Delft University of Technology, Delft, The Netherlands. [4]Christian Vincenot is an Associate Professor at the Complex Systems Group (NEXUS:CSR), Department of Engineering, Faculty of Science, Technology and Medicine (FSTM), The University of Luxembourg, Esch-sur-Alzette, Luxembourg. [5]Cheng-Zhi Qin is a Professor at the State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China. [6]Cheng-Zhi Qin is a Professor at the College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China. [7]Sebastian Fiedler is a Post-doctoral Researcher at the Department of Plant Ecology, Technische Universität Berlin, Berlin, Germany. [8] Carsten Lemmen is a Senior Researcher at the Institute of Coastal Systems—Modeling and Analysis, Helmholtz-Zentrum Hereon, Geesthacht, Germany. [9]Iseult Lynch is a Professor of Environmental Nanosciences at the School of Geography, Earth and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, UK. [10]Maria Pierce is the Head of the Work Area of Individual-Based Modelling at the Thünen Institute of Baltic Sea Fisheries, Rostock, Germany. [11]Maria Pierce is a Research Associate at the Modeling and Simulation Group at the Institute for Visual and Analytic Computing, University of Rostock, Germany. [12]Bruce Edmonds is an Adjunct Professor at Umeå University, Sweden. [13]Bruce Edmonds is an Adjunct Professor at The Arctic University of Norway, Tromsø, Norway and Director of the Centre for Policy Modelling.. [14]Eliot McIntire is a Research Scientist at the Pacific Forestry Centre, Victoria, British Columbia, Canada. [15]Hsiao-Hsuan Wang is an Associate Professor in the Department of Entomology, Texas A&M University, College Station, Texas, USA. [16]Tomasz E. Koralewski is an Associate Research Scientist, and William E. Grant is a Professor at the Department of Ecology and Conservation Biology, Texas A&M University, College Station, Texas, USA. [17]Ming Wang is a Post-doctoral Researcher and Volker Grimm is an Honorary Professor at the Department of Ecological Modelling, Helmholtz Center for the Environment—UFZ, Leipzig, Germany. [18]Ben Black is a Post-doctoral Researcher at the Department of Civil, Environmental, and Geomatic Engineering, ETH Zurich, Switzerland. [19]Ben Black is the Head of the Working Group for Data and Modelling Infrastructure for Living Labs at the Leibniz Centre for Agricultural Landscape Research, Müncheberg, Germany. [20]Volker Grimm is an Honorary Professor at the University of Potsdam, Plant Ecology and Nature Conservation, Potsdam, Germany.

Abstract

Do you want your model findings to be used and have a place in academic or policy discourse? Then PAVE that place. For modelers, PAVE (Purpose, Assumptions, Validity, and Exploration) is a semi-structured method to communicate the suitability of models to inform a particular purpose. For users of model results, PAVE is a method for identifying the strength with which model findings inform a question of interest. For all audiences, PAVE distills more comprehensive model documentation protocols into a simple set of concepts for modelers and users alike to discuss. In doing so, PAVE solves a dilemma: there is an audience for whom formal protocols for documenting model development, testing, and analysis are everything. And there is an audience for whom such technical tools are a terrifying abyss, neither to be approached nor looked upon. PAVE is meant to help these audiences talk to each other about model findings.

Keywords: modeling, communication, policy/ethics

Among those who will read past the title of this work, we expect at least two distinct groups, asking questions such as the following:

(a) I'm not a modeler—what should I look for when I read a modeling study, use modeling results, or join a participatory modeling process?
(b) I'm a modeler—how should I communicate my contribution to nonmodelers?

For modelers, PAVE (Purpose, Assumptions, Validity, and Exploration; Box 1) is a semi-structured method to demonstrate how to communicate clearly to others the suitability of our models to inform a particular (research, management, or other) purpose. For users of model results (for research, policy, or management decision-making, or other), PAVE is a method for identifying the strength with which model findings inform a question of interest. For all audiences, PAVE distills the elements of more comprehensive model documentation protocols to a simple set of concepts for modelers and users alike to discuss.

In doing so, PAVE solves a dilemma: there is an audience for whom formal protocols for documenting model development, testing, and analysis are everything. And there is an audience for whom such technical tools are a terrifying abyss, neither to be approached nor looked upon. PAVE is meant to help these audiences talk to each other about model findings and get them prepared to look up technical details when needed. As such, it acts as a boundary object among modelers and nonmodelers alike to enable nimble discussion and use of model-generated knowledge in practical and policy decision-making. Technical protocols that describe a model support the communication of the modeling process. In the same way, we hope that PAVE enables communication of the

Received: June 5, 2025. **Revised:** January 12, 2026. **Accepted:** January 22, 2026

usefulness of modeling outputs among modelers and nonmodelers alike.

Introduction

The utility of any model rests on how well its essence, assumptions, and outputs can be communicated to those who encounter it, and whether its purpose resonates with others. On the one hand, modeling tools can help to solve, as well as transform understanding of, complex problems; to engage community members; and to build consensus around contentious issues (Loblay et al. 2023). On the other hand, bringing together modelers and nonmodeling stakeholders—including community members, nontechnical professionals, or decision-makers, among others—is fraught with challenges (Voinov et al. 2016). Decision-makers often understand models as "black boxes" (Brugnach et al. 2007, Lorscheid et al. 2012, Boschetti et al. 2018) and struggle with the totality of assumptions, uncertainty, and limitations that comprise a model. A "gap" separating those who build models and those who might use them is well documented (Borowski and Hare 2007, Janssen et al. 2009, Jones et al. 2020), with several key suggestions in the literature on how to resolve it.

Most broadly, "more communication" between modelers and nonmodelers is encouraged (Jones et al. 2020), at regular intervals and beyond the end of projects (Will et al. 2021), but "more communication" alone will not always reduce the gap (Arnejo et al. 2025). Good communication is audience- and problem-specific, and should be in service of resolving confusions around model concepts, or reconciling different or incompatible understandings of the overall system (Hall et al. 2014, Arnejo et al. 2025).

Participatory approaches to modeling that engage nonmodeling audiences are another key part of reconciling the gap (Borowski and Hare 2007, Brugnach et al. 2007, Mannix et al. 2022, Elsawah et al. 2023, Loblay et al. 2023, Zellner 2024), building shared understanding and "demystifying" models and the modeling process (Herman et al. 2019). Still, even participatory processes have interfaces across knowledge domains that must be spanned, and every participatory process has people on the outside looking in, needing to understand what has transpired. Boundaries between modelers and model users are pervasive: differing understandings, different goals and objectives, and different timelines (Borowski and Hare 2007, Doyle et al. 2019, Will et al. 2021). Where there are boundaries, boundary objects (artifacts that ground the translation between different perspectives; Star 2010) are necessary (Vincenot 2018). We propose PAVE (Purpose-Assumptions-Validity-Exploration) as an ideal boundary object for the purpose of enabling discourse among those who build and implement models, and those who wish to use models or make decisions informed by their outputs. Participation on its own does not resolve the need to communicate the model to those who did not participate in building it.

Underlying the pushes for both participation and communication is the need for a common language (Brugnach et al. 2007, Gray et al. 2017, Arnejo et al. 2025) to better enable a dialog among modelers and model users. One part of this would be unambiguous language and concepts to describe models, model findings, and the modeling process. Across different fields, modelers have solved this part of the problem many times independently—materials modelers have the Modelling Data (MODA) standard, toxicology modelers have the QSAR model report format (QMRF) (Kolokathis et al. 2024), whereas discrete event simulators have STRESS (Monks et al. 2019). In the domain closest to the authors of the present manuscript, agent-based modelers have the Overview, Design concepts, and Details (ODD) protocol (Grimm et al. 2006) and its many progeny (the ODD+D for models including human decision-making, Müller et al. 2013; the ODMAP for modeling species distributions, Zurell et al. 2020). Perhaps the most careful efforts at standardized descriptions of models come from those who need to connect different models together, such as through the Basic Model Interface (BMI) in the geosciences (Hutton et al. 2020), or the AgentBlocks approach now growing in use among agent-based modelers (Filatova et al. 2025). However, these efforts are largely prepared by modelers, for modelers, and tackle the challenge of clear model communication across an audience that largely speaks the same (modeling) language (Rosello et al. 2025). We put some of these standards into focus in this manuscript, to walk through what they do and do not do in terms of communicating our models to others. (As a group of modelers engaged most closely in the modeling of human-environment and socio-ecological systems modeling, we focus on standards relevant to our fields. We expect, however, that similar frontiers exist in other fields and that the reader might also consider the analogs that shape model communication in their areas of interest.)

In our parts of the world of modeling and modelers, the ODD protocol (Grimm et al. 2006, 2010, 2020b; Polhill et al. 2008) is a notable highlight for how rapidly it has been adopted as a standard (Vincenot 2018) for describing the purpose, design, and all details of agent-based models. ODD stands for Overview (the big picture), Design concepts (including theory and other assumptions), and Details (everything else that is needed to replicate the model, its initialization, and its input of external data). The ODD and ODD+D protocols provide a means for standard description of modeling assumptions and details among modelers, but were not tailored to the needs that nontechnical users of a model might have.

The TRACE document ("TRAnsparent and Comprehensive Ecological modelling documentation," Schmolke et al. 2010) is targeted to more thoroughly communicate a model and its history. It was refined and streamlined in 2014 (Grimm et al. 2014), with further guidance published in 2021 (Ayllón et al. 2021), and is defined as "a tool for planning, performing, and documenting good modelling practice" (Grimm et al. 2014). TRACE documents are developed by a modeling team along their iterative modeling process, to "provide convincing evidence that a model was thoughtfully designed, correctly implemented, thoroughly tested, well understood, and appropriately used for its intended purpose."

TRACE provides a clear and exhaustive reference guide to describe what is happening in a model, how it came to be that way, and what the motivations are (Grimm et al. 2014), but (like the ODD) its use at the policy interface or as a boundary object among modelers and nonmodelers has not been documented. What makes TRACE a useful tool for communication *among* modelers may well be what limits use outside of modeler communities—it is rich, complete, and intimidating (or "excessive," Daly et al. 2022, p. 1), often comprising 20–30 pages or more. Some will dig in, but many others—those who are senior, those who are in and out of the project, those who have limited time to gather input for a decision, or those who are just reading a published product—will need something simpler, self-contained, and accessible. That is, they will need an "executive summary" describing not just what the model shows, but what it was asking and how.

One step in this direction is POE (Grimm et al. 2020a), which points decision-makers to ask three screening questions before

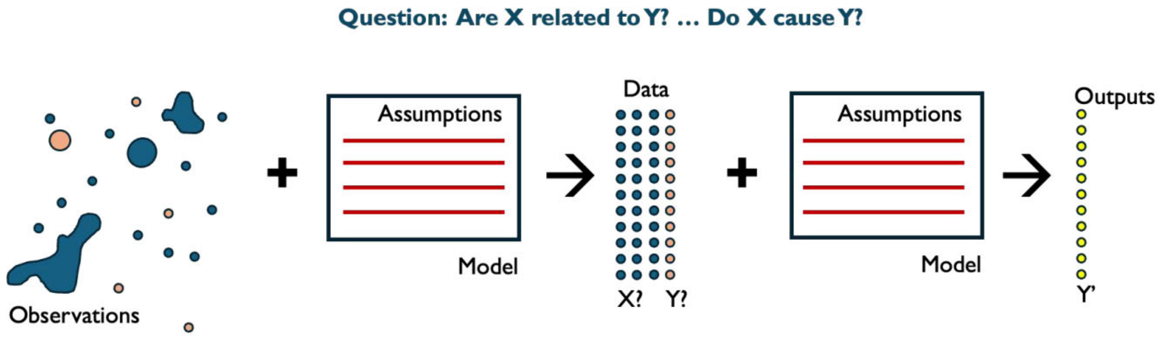

Figure 1. Observations and assumptions are fundamental ingredients in the building of models (as sets of assumptions for treating observations and data). Model outputs are the product of assumptions for translating observations (themselves derived from many different ways of knowing, and taking many different forms) into datasets, as well as of assumptions for representing real-world processes in the system of interest.

they take any model output into account: What is the model's purpose (P)? How is it organized (O)? Is there evidence that the model is realistic enough for its purpose (E)? POE was triggered by the use and abuse of models during the COVID-19 pandemic where, in the surge of model development and publication, models developed to demonstrate a certain concept (like "flattening the curve") were used for a different purpose (like making predictions). POE is, however, still limited, in that to answer these questions, someone with modeling expertise has to help.

Although this expertise will always be needed when it comes to details, here we are trying to suggest a concise approach for talking about models that is more comprehensive than POE. We propose a streamlined method for communicating key model assumptions and findings, drawing on some strengths of the ODD protocol on the one hand (simple, self-contained overview of model components) and on some strengths of TRACE on the other (description of modeling process). We suggest that our PAVE (Purpose, Assumptions, Validity, and Exploration) pathway, presented below, may better create space for modelers to communicate their work to nonmodelers, and for nonmodelers to interpret model findings and participate in the iterative process of model critique and development.

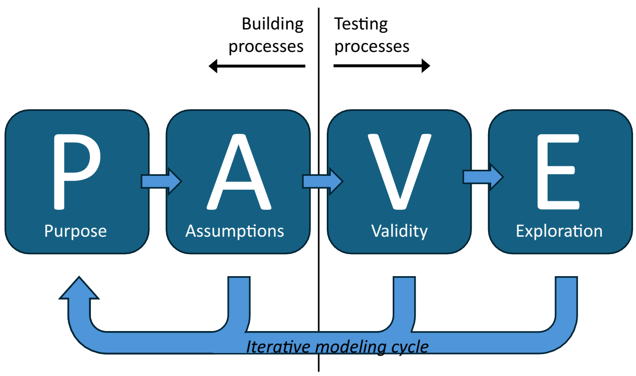

Figure 2. The PAVE pathway for describing findings from a modeling exercise. A PAVE description will always be a static snapshot of the iterative modeling cycle, which pairs theory building processes (stating problems and assembling assumptions into theory and models) and theory testing processes (testing validity of assumptions to inform a particular problem, and exploring the consequences of sets of assumptions in experiments).

How do I PAVE?

To begin PAVEing, we first present a working definition of what we mean by "model."

We define models in the current work as the sets of assumptions that describe what we believe about the real-world process or system we are interested in, and that help translate our observations of that system into data and predictions (figure 1). This definition of model—as a representation of a set of real-world processes—places it close to but distinct from theory—as an explanation of real-world processes (Martens 2020).

These sets of assumptions typically do not come together from nothing, instead emerging from iterative cycles of model development that commonly build upon or adapt existing theory or models to respond to novel purposes and questions. Within this cycle, Purpose (P) and Assumptions (A) are model "building" processes—drawing together novel assumptions along with (possibly) prior theory or models to inform a purpose—whereas V (Validity) and E (Exploration) are model "testing" processes—evaluating the fit of this novel set of assumptions to the purpose (figure 2). Outputs from both validity-testing processes (V; such as sensitivity analyses, calibration, and validation exercises, or other—see appendix 1) and novel experiments and exploration (E) aimed at informing the model's purpose (P) can lead to changes in the question or problem of interest (P), or the set of assumptions present in the model (A)—setting up new demands for validity testing and new opportunities for experiment.

The iterative cycle implies that models are not ever finished; however, sometimes they are ready enough to be used and shared. Wherever model findings are shared or published, PAVE acts as a snapshot of the current state of that iterative cycle, forming an unbroken chain that always begins with the clearly stated purpose (P).

P—Purpose: What objective or question is your model informing?

Whether you frame the purpose of your model as a problem, an objective, or a research question to be answered, all good models have (at least) one, stated explicitly. There are many different reasons to build and test models that may be predictive or explanatory in nature (Hall et al. 2014). Perhaps the purpose is to predict a change in forest cover resulting from a new policy, or to identify how different strategies for managing risk pay off, or test whether one theory is more appropriate to explain an ecosystem outcome than another. Edmonds and colleagues (2019) distinguish seven different model purposes (prediction, explanation, description, theoretical exploration, illustration, analogy, and social learning), whereas with POE, Grimm and colleagues (2020a) distill and disentangle only three (demonstration, understanding, and prediction). Taking a different approach, Troost and colleagues (2023) describe model purpose through six dimensions (focus of interest, target situations, and requirements for model interoperability, resolution and transparency, computational resource constraints, and precision and accuracy). These different constructions of purpose all make the same critical point—a topic, objective, or research question alone is rarely sufficient to unambiguously show the "why" of your model; research questions commonly require qualifiers like "for whom" or "for what use." In this way, a model's purpose also defines its scope, or domain of applicability. Using a model beyond this range is inappropriate, but unfortunately quite common. For example, a team of researchers, planners, or managers developing a model to inform a question like "how will coastal flooding under future environmental change affect flood-prone homeowners?" might have the goals of (i) ranking possible options or investments based on some set of criteria; (ii) estimating uncertainty around outcomes like flood depth or duration; or (iii) identifying surprises or extreme events a planner may wish to consider (adapted from Fischhoff and Davis' 2014 discussion of the different modes of observing and communicating uncertainty). These are all related but different purposes, tied to the same question—they differ in how explicitly they embed goals and preferences, versus leaving users to formulate their own.

Moreover, these are only an illustrative few examples of the many different purposes one may have in building and applying a model. Whatever it is, the model's stated purpose should bound what set of assumptions is necessary for the model to include, in order to inform it. Without knowing (unambiguously) the purpose, it will not be possible to assess whether your model assumptions (A) make the model appropriate for its purpose (P in PAVE).

A—Assumptions: What set of assumptions (theory or model) are you drawing together to inform your purpose (P)?

Importantly, a model is chosen or assembled to respond to a specific purpose, so a description of a model's assumptions should allow a reader to see how the different parts of a model's purpose or question are informed by it. If a purpose is given as a question like, "How does X impact Y under conditions of D and E?" then the set of assumptions should capture existing theory or understanding of how conditions D and E influence X, Y, and the impact that X has on Y.

How does the model developer describe their assumptions appropriately for others outside their core discipline? This can be challenging to do well, as the language that communicates unambiguously within one's field may be impenetrable outside it.

At a very general level, modeling and simulation build on one or both of two fundamentally different approaches. The first is "extrapolation" from or "training" on observed system behavior without explicitly mirroring internal system processes; this is often called "data-driven" or "empirical" modeling. The second draws on knowledge about how the system works internally to build model components that mimic selected system processes. This is often called "mechanistic," "process-based," or "theory-driven" modeling. In practice, many models combine these two approaches (Troost et al. 2023).

It is key to communicate the assumptions regarding both approaches (figure 1): Within data-driven modeling, assumptions about how to translate available observations into data, how to interpret them, the expected continuity of observed behavioral patterns and driving factors from observed to simulated cases, and assumptions on identifiability, sampling, error and sampling bias involved in estimation strategies must be made clear. To describe assumptions about constructing and interpreting data, we suggest answering the question "what *needs* to be true of these observations for them to be appropriate for your purpose?" as follows:

(a) Observations must contain sufficient variation in variables X, D, and E.
(b) Observations must be about different points of time.
(c) Observations must be independent of each other.
(d) Observations must be normally distributed.

Within mechanistic or process-based modeling, when communicating assumptions about processes in our system of interest, we suggest recognizing that quantitative analysis (such as modeling and simulation) is commonly about comparing interacting (countervailing, amplifying, etc.) processes to understand how they compound to produce an observed outcome (Frisch 1933). Economic analysis compares costs and benefits to anticipate choices. Physics compares gravity and lift to deduce whether an object will fly. Ecologists compare mortality and fertility to deduce survival of a population, and the balance of competitive and synergistic interactions across populations to predict the composition of a community.

We suggest first to explain, in the modeler's area of interest, which counteracting processes influence the behavior of Y or the effect of X on Y, according to established knowledge. Which processes accelerate or increase the effect on variable Y? Which processes dampen or decrease it? Then, secondly, explain which of these processes are represented in your model and to what extent. This clear separation of assumptions about real-world processes and model description is essential for an understandable explanation of assumptions. All too often this is mixed up and model and real-world entities are not clearly distinguished.

For a concise and understandable description of the model itself, we suggest describing them concisely along the parts of the Overview section of the ODD and ODD+D protocol (Müller et al. 2013, Grimm et al. 2020b), which includes enumeration of the model's "Entities, state variables, and scales" and "Process overview and scheduling" at a compact and concise level. To put these in terms that may resonate beyond the modeling community: describe the different pieces and key variables (quantities that can rise and fall in the model), the processes that can change them, and the way these are sequenced in the model. Many of these pieces may themselves come from prior models or existing theory, so that links to their earlier homes in the literature are critical reflections of the iterative modeling cycle. Where the systems of interest are complex and words alone are insufficient to communicate assumptions effectively, visualizations may be helpful (as in the Visual ODD; Szangolies et al. 2024). For example, system diagrams such as causal loop maps (with visual connections communicating system scales, processes and feedbacks, as well as system boundaries and the movement of quantities across them; e.g., Kim 1999) can enhance (but not replace) descriptions in text.

Across both of these bundlings of assumptions (about observations and about system processes), authors of model findings should reflect both conscious and unconscious biases (e.g., as primed through disciplinary training) and be explicit about the concepts that shape your lens as a modeler. In what ways might sets of processes or observations be outside of your frame of reference? Which factors did you leave out in your model that you discussed explicitly, or that are included in similar models, and why did you leave them out?

This is a long set of things to ask for, and for complicated models or complex systems, a description of everything can become long and cumbersome. We highlight our emphasis with PAVE on communication (rather than only description), and that the PAVE summary might not need all model processes in detail. Similarly, it may be helpful to communicate processes the model does ∗not∗ include, that a layman might otherwise expect. Importantly, as part of PAVEing a place for your modeling work, the description of assumptions should give a simple, accessible narrative of processes that plausibly are capable together of informing your purpose. Whether they ∗actually∗ do or not is your next task to show.

V—Validity: How well can you demonstrate the appropriateness of your set of assumptions (A) to inform your purpose (P)?

Modelers are often asked to "show me that your model works!" Fair enough—models are by definition simplified representations. How can we be sure that the factors we omit or simplify are not inappropriately biasing the model's answer to our question?

The words "validity" and in particular "validation" are challenging to use in this context, as they are understood in so many different and sometimes even contradictory ways, that it has been recommended to avoid them altogether (Augusiak et al. 2014). For our purposes, we use "validity" to mean: to what degree is the model a good fit for its purpose. Importantly, validity is a measure of how well a model informs a question in a particular context, and not an enduring property of the model itself (Troost et al. 2023, Lim 2024).

Showing the reader how well your assumptions match your problem may take many forms. Many disciplines have established methods to demonstrate one or more dimensions of validity. For PAVE's purpose of communicating to nonmodelers or across disciplinary boundaries, it will be key to take a step back from disciplinary habits and also communicate implicit conventions or tacit knowledge embedded in these methods in a simple way. We direct the interested readers to a list of key approaches in appendix 1, and focus here on the goals these validation efforts have in common. Troost and colleagues (2023) provide a comprehensive protocol to find an adequate concept for demonstrating validity (or "fit-for-purpose") for any modeling context. Distilling this concept into a simple communicable format leads to three essential parts for the V section in PAVE.

First, show that the assumptions on model structure (what theory you have used, what you have left out) are adequate for the purpose that is the main outcomes are not driven by simplifications you have made, but by core defensible assumptions or reliable data. This should involve theoretical arguments using shared understanding of real-world mechanisms (theory) as well as sensitivity analysis and robustness checks. Communicate the variation and uncertainty that characterize the system you are modeling. The literature on communicating uncertainty in technical knowledge is quite large and beyond the scope of our short communication; highlights include Brugnach and colleagues (2007), Fischoff and Davis (2014), Guillaume and colleagues (2017), and Doyle and colleagues (2019).

Second, explain whether it is meaningful or possible to compare model outputs to empirical data and if yes, do so! This must involve checking on representativeness of data, constancy of relationships, and identifiability of estimated parameters (Troost et al. 2023).

Third, draw from what you have shown to put the model in its place. Force yourself to qualify the "fit-for-purpose" with clear language: "It can be used for…, as long as, …" "It should not be used for…" "It will become invalid if … changes." Tell a user or a reader, "this is how well you should expect this model to fit this specific purpose." Do not expect nonmodelers to draw these conclusions themselves. Especially, if prediction is the purpose, be clear regarding what kind of precision and conditionality is possible with the model: Unconditional or probabilistic prediction with estimated accuracy (like a weather forecast or forecast of election results)? Conditional projection of long-term statistical averages (like climate projections)? What-if scenario analysis? Vague exploration of possible outcomes? (Marchau et al. 2019).

Consider that your validity demonstration should match the fidelity with which you run and interpret your experiment and communicate whether experiment conditions fall within or outside of space included in your validity check. Without any such qualification, we have no way of putting any experimental or exploratory results (E) into context.

E—Exploration: Given (A) and (V), what exciting new thing can you show?

You have done it. You have stated your purpose (P), described the assumptions (A) you bring to bear on it in your model, and demonstrated the fit of those assumptions to the problem with some demonstration of validity (V). With this trust and shared understanding, you can now pivot to showing something new and exciting, by trying to answer your original research question, and perhaps even more. Maybe you have asked a question about forest fire and climate, described a model of forest cover that appropriately includes rainfall and burning, and showed how well that model reproduces recent histories of fire and forest loss. Perhaps now you might show what is expected for the coming decades, given what we know about that forest and rainfall expectations; maybe instead you will show how the last few years might have been different if we had tried a different approach to brush management; or you might demonstrate a different theory or model that reproduces some or all of the past history even better. Whatever it is that you are doing, you are constructing modeling experiments that—like experiments in any other branch of science (Peck 2004, Lorscheid et al. 2012), and heeding best practices in exploratory model analysis and its uncertainty (Kwakkel and Pruyt 2013, Kwakkel 2017)—should rigorously examine the response of some outcome to careful variation in model inputs or conditions. And whatever such experiments are, your efforts at demonstrating and communicating validity (V) make them easier to understand and interpret.

Importantly, though (E) comes at the end of the word PAVE, it is not likely the end of our modeling process. Recalling our iterative modeling cycle (figure 1), model exploration helps both to meet the model's purpose (P), as well as to better understand how the modelled system works (i.e., how the answers that are obtained with the model are emerging). Exploration may reveal issues that challenge the assembled assumptions beyond what validity exercises suggested, or may lead one to revisit the very purpose (P) of the whole exercise. Wherever it might go next, the current PAVE snapshot should make clear where (and why) that is.

Box 1. A toy example of a PAVE summary for … an analog thermometer

In this example, we provide a PAVE summary for the exciting finding of who had the best-heated soup. PAVEing real modeling findings might take up more space; here, we hope to highlight both the what and the why of P, the building blocks that make up A, and the critical distinction between activities we undertake for V versus E.

P (Purpose)—To find out whose soup is juuuuuust right, by constructing a measure of temperature with sufficient fidelity to identify soup in the optimal temperature range for consumption.

A (Assumptions)—We develop an instrument for measuring soup temperature by assuming (1) mercury expands roughly linearly with temperature and (2) incompressible fluids in a tube will flow up the tube as they expand. Thus, markings made along a clear tube with mercury inside can be used to indicate points in temperature associated with the current mercury level. We draw on previous research identifying the optimal temperature range for hot beverages like soup, which has been shown to be between 54 and 71°C (Abraham and Diller 2019).

V (Validity)—To calibrate our thermometer, we drew on two independently verified temperature-driven events—the freezing of water at atmospheric pressure (0°C) and the boiling of water at atmospheric pressure (100°C). We placed our thermometer near a block of ice and warmed the area until the water began to melt, marking the mercury level as 0°C. We then let the thermometer sit in the water and continued to warm it until it boiled, marking the mercury level as 100°C. We then assumed linearity in between these points.

We then validated our calibrated thermometer with an independently known point. We had our cousin Martin put the thermometer underneath his tongue for a minute, giving a measure of about 37°C, which is what was expected.

We do not have further reference points below 0°C or above 100°C, and do not know whether our thermometer will respond similarly outside of the range 0–100°C. However, since our optimal range is between 54 and 71°C, within the functioning range of our thermometer, we are able to use this model to determine ∗whether∗ soup is within this range or not, while not having confidence at comparing beverages outside the range of 0–100°C (e.g., to see which among several might be closer to optimal).

E (Exploration)—After rinsing the thermometer, we took measures of the temperature of everyone's bowl of soup. Mahdi's was 89°C, Violet's was 43°C, whereas Chip's was juuuust right at 68°C (within the range of 54–71°C identified by past research).

Please see appendix 2 for further examples of PAVE summaries in practice.

Discussion

We have presented the PAVE pathway as a distillation of the key elements of protocols such as TRACE, presented at an "Overview" scale as defined in the ODD, to streamline the use and discussion of model findings among modeling and nonmodeling professionals. In appendix 2, we present several short example PAVE uses. We plan to collect more examples with support from the Standards Working Group of the Open Modeling Foundation (OMF, www.openmodelingfoundation.org/), in which all authors are members. OMF has the mission to develop a body of standards and best practices among virtually all communities of modeling scientists. We suggested in our opening words at least two distinct groups in the audience for this contribution, asking questions such as the following:

(a) I'm not a modeler—what should I look for when I read a modeling study, use modeling results, or join a participatory modeling process?
(b) I'm a modeler—how should I communicate my contribution to nonmodelers?

For the group of nonmodelers, we hope that this document may serve as a quick reference for the critical pieces necessary to put a modeling contribution in context. The modeling process is an iterative cycle of theory building and theory testing, but any one set of model findings can be described by a PAVE snapshot—a clear purpose or question (P), the set of assumptions applied to inform that purpose or question (A), a demonstration of how valid that set of assumptions is for informing the purpose or question (V), and finally an exploration of something new (E). Importantly, as we discuss further below, we cannot talk about what the model shows us ∗until∗ we have understood what the model is doing, and how.

For the group of modelers, we similarly hope that PAVE can help to distill key pieces of your modeling process and documentation into a narrative that nonmodelers can engage with and challenge, with less of the burden of modeler-specific knowledge that protocols like ODD and TRACE demand. In this, we are proposing that PAVE may better function as a boundary object—a tool allowing for cooperation among actors with different interests and understandings of a process (Star 2010)—than other protocols that have come before.

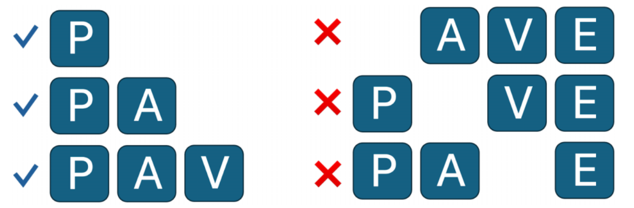

Figure 3. Acceptable and unacceptable partially PAVEd spaces.

Partially PAVEd spaces

A further strength of PAVE as a snapshot of the modeling process is the recognition that it is a chain of processes starting always with a purpose. There are many ways in which a part of this chain can stand on its own as a contribution, but it must be unbroken from the left—no gaps (figure 3). For instance , consider the following:

(a) drawing on literature review to identify key questions for future modeling work (P);
(b) outlining an appropriate conceptual model to respond to a research question (P-A);
(c) demonstrating the possibility space of a model implemented to respond to a specific question (P-A-V).

This manuscript itself is a partially PAVEd space. We present our problem of finding a way to reliably communicate models between those who did and did not build them (P), and present the PAVE pathway as a set of assumptions about the way to move forward (A). We hope to observe how PAVE enables improved communication in different modeling groups, both including and not including us, as well as in the reading and writing of modeling studies, as ways of validating this model moving forward.

The key to remember is that PAVE is a chain, and once an element is missing, nothing to the right of it can be treated as meaningful. For instance, consider the following:

(a) A-V-E: Without a defined question, it is not possible to know whether a set of assumptions (conceptual model) is appropriate, nor how to interpret any findings of the model;
(b) P-V-E: Without knowing what the underlying model assumptions are, it is not possible to make sense of any outputs from the model;
(c) P-A-E: Without knowing how well the assumptions of the model allow an interrogation of the question (i.e., what model outputs are possible in response to what inputs), it is not possible to put an experimental result in proper context.

PAVEing beyond models and simulation

PAVE has applicability beyond describing modeling and simulation applications. More broadly, instruments for data collection/generation—questionnaires, experiments, or games—typically invoke some set of assumptions (theory or model) to generate the outcome measure(s) of interest. PAVE could also make developing, talking about, and understanding conceptual models easier, with V and E marking out key "woulds"—what would help demonstrate the fit of assumptions to the purpose, and what new exploration would such a model enable? The ODD protocol has been applied in just this way in psychology to make a certain conceptual model explicit, even with no intention of implementing a computational model (Caldwell-Harris 2019). Similarly, Secchi and colleagues (2024) recommend to formulate theories in sustainability science using a clear interface (what we have called here a boundary object) for communication. PAVE follows the same rationale, in a way that is less technical than the ODD.

Very commonly, these instruments encourage respondents to state or otherwise reveal some preference, belief, or behavior in a hypothetical (instrument) setting that is assumed to be a construct for a real-world outcome of interest—assumptions that require clear explanation and some degree of validation before they can be of use in making inferences. Like computational models, these instruments are of greater value when the results of any experimental or exploratory work E are put in context through P, A, and V, and we hope PAVE can be of use in the "credibility revolution" that is building out the space for construct and external validity alongside the better-measured internal validity in the experimental and empirical social sciences (Jiménez-Buedo and Russo 2021, Esterling et al. 2023).

PAVEing beyond this manuscript

Good practices emerge from evaluation and adaptation. Much as the ODD (Grimm et al. 2006) grew (Grimm et al. 2010) and then grew again (ODD+D; Muller et al. 2013) and again (Grimm et al. 2020b), we see this as PAVE's first step. PAVE's next steps will include structured experiments—(i) challenging modeler teams to develop PAVE summaries for model findings and evaluating their role in improving shared understanding across multi-actor teams and (ii) challenging nonmodelers to use PAVE principles to shift their ability to participate in model discourse. Across all such experiments, careful characterization of the models in focus will help us to identify where PAVE is more or less effective, and where there is scope for iteration and update.

Conclusions

Gaps in communication persist among those who build models and those who use or otherwise wish to understand them, a problem only partially addressed through participatory modeling processes. Clear protocols for describing model theory, logic, and process have been developed and refined over recent decades, with the ODD and TRACE protocols as leading examples for documentation standards that provide everything modelers need to implement or extend a model (see Jakeman et al. 2024 for a recent overview of standards and Good Modeling Practice). However, models and the modeling process are complicated, with documentation often being long and dense, so that the very strengths of protocols like ODD and TRACE for modelers can limit their utility as communication tools beyond the modeling community. We propose the PAVE pathway as a streamlined narrative for describing findings from a model that draws on the strengths of ODD and TRACE while functioning more as a boundary object to enable communication and cooperation across groups with different goals and understandings. For nonmodelers, we intend for PAVE to structure a narrative of everything they should ask and be aware of as they interpret a specific set of findings from a model. For modelers, we intend PAVE to make clear the basic descriptive needs for a contribution based on their modeling work, to be able to communicate effectively to those who did not build it, but might make use of it.

Author contributions

Andrew Reid Bell (Conceptualization, Writing – original draft, Writing – review & editing), Christian Troost (Conceptualization, Writing – original draft, Writing – review & editing), Tatiana Filatova (Writing – original draft, Writing – review & editing), Christian Vincenot (Writing – original draft, Writing – review & editing), Cheng-Zhi Qin (Writing – original draft, Writing – review & editing), Sebastian Fiedler (Writing – original draft, Writing – review & editing), Carsten Lemmen (Writing – original draft, Writing – review & editing), Iseult Lynch (Writing – original draft, Writing – review & editing), Maria Pierce (Writing – original draft, Writing – review & editing), Bruce Edmonds (Writing – original draft, Writing – review & editing), Eliot McIntire (Writing – original draft, Writing – review & editing), Hsiao-Hsuan Wang (Writing – original draft, Writing – review & editing), Tomasz E. Koralewski (Writing – original draft, Writing – review & editing), William E. Grant (Writing – original draft, Writing – review & editing), Ming Wang (Writing – original draft, Writing – review & editing), and Volker Grimm (Conceptualization, Writing – original draft, Writing – review & editing).

Appendix 1. Examples of validation approaches

In a general sense, "validity" has many dimensions so that assessing whether a model "works well enough" is a multi-criteria assessment. Lim (2024) provides a comprehensive typology of these dimensions, including the following:

(a) **Face validity**—Does the model appear to be reproducing patterns of interest?
(b) **Construct validity**—Does the modeled outcome have a meaningful link to the real-world outcome of interest?
(c) **Internal validity**—Can changes in the modeled outcome be meaningfully attributed to changes in model parameters?
(d) **External validity**—Do the input parameter ranges and observed outcomes map well onto specific real-world contexts of interest?

Different dimensions of validity will be more relevant in different exercises, and different models will perform differently across dimensions of validity. Thus, demonstrating validity (V)—the fit of a set of assumptions (A) to inform a specific question or purpose (P)—can mean many different things. Examples of common efforts to demonstrate problem-specific validity include the following:

(a) **Matching patterns in outcomes of interest**. Observing whether expected relationships appear in model outputs—for example, "when less rain falls, we see more fires and more severe fires." Although this seems trivial, it is a first step for building trust in the internal validity of the model, sometimes called a "sanity test." A challenge here is that there are commonly a large number of different models that are able to generate the pattern in consideration (known as *equifinality*); this can be a problem if those same models might generate very different patterns under other conditions. The pattern-oriented modeling (POM) approach (Grimm et al. 2005) addresses this by choosing larger numbers of related but different patterns to match simultaneously (e.g., matching migration rates but also labor sector participation and demographic distribution). Each pattern is used as a filter to reject unsuitable models, so that ultimately fewer models will match all of the different patterns well, raising confidence that the mechanisms included as assumptions are good fits to the question raised.
(b) **Sensitivity analysis**. Systematic sweeps across allowable values for model parameters—either one at a time, or using one of several approaches to varying all at the same time to observe interactions (Brugnach et al. 2007, Thiele et al. 2014)—to show that (1) model results are not overly sensitive to parameter uncertainty or values and (2) those parameters which are most sensitive are likely to be about those processes that dominate the model's dynamics and response.
(c) **Calibration.** Fitting a model to best reproduce a particular range of outcomes by identifying parameter settings that minimize error between observations for those outcomes and model predictions. In cases where your observations can *meaningfully* be separated into independent subsets, a combined calibration and *validation* exercise may be performed by calibrating (or training) the model on one set, and validating (or testing) on the other—that is, using part of the data to estimate what parameters fit the model best, but then measuring that fit based on the prediction error in the other part of the dataset (Windrum et al. 2007). There are further clever variations on this general approach. For example, Approximate Bayesian Computation (ABC) finds good calibration parameters by gradually narrowing the range of possible values to explore, based on the ranges seen in strong-performing calibration runs (van der Vaart et al. 2015). The "leave out part of the data for testing" approach to validating can be explored iteratively, breaking the dataset into k parts and exploring the features of models tested on each of the k parts of the data after calibrating on the remaining k-1 parts (known as k-fold cross-validation; Fushiki 2011).

These are examples drawn from a spectrum of approaches to demonstrate model validity with varying fidelity. Different approaches for calibration and validation are reviewed in Windrum and colleagues (2007) and Fagiolo and colleagues (2007), with a wonderful walkthrough of what demonstrating validity can mean for agent-based models found in Troost and colleagues (2023).

Appendix 2. PAVE summaries of existing studies

The following pages give three simple examples of explaining model findings (one game experiment, one agent-based model, and one example of explaining a communication protocol, the PAVE pathway itself) using the PAVE pathway as simple one-page narratives. These short summaries do not contain all of the details included in the full published works, but instead, shorter overviews that make the linkages clear across the PAVE elements:

(a) **P (Purpose)**—What are we asking in this work?
(b) **A (Assumptions)**—What theory or model(s) and other assumptions are we assembling to inform our purpose?
(c) **V (Validity)**—What indications do we have that the assembled assumptions are appropriate to inform our purpose?
(d) **E (Exploration)**—With our estimate of validity known, what new interesting things are we learning from the work?

These examples are not meant to be prescriptive of how to use PAVE in your work, but rather, to demonstrate how the PAVE components may be included in published work to create a usable narrative. Some PAVE descriptions may lend themselves well to bullet points and "highlights," whereas others may work better as prose. Some may function well with text narratives, but where helpful, visualizations (such as in the assumptions of A2.2 below) may help to communicate structure or findings.

As a modeler, simple narrative summaries like these may be helpful as you plan your experiment, or plan your finished product. As a nonmodeler participating in a modeling project or reading a modeling study, short narrative summaries like these that you develop yourself or request from your colleagues might make helpful aids as you consider what model findings could possibly mean. As instructors or as students of modeling or the use of model findings, they may similarly enhance thinking or discussion around modeling or guide the development of research proposals.

Wherever you are in your modeling, analyzing, or reading, we hope PAVE is a helpful lens to consider what you are learning or doing.

A2.1 PAVE summary for "Understanding pastoralist adaptations to drought via games and choice experiments—field testing among Borana communities" (Bell et al. 2025)

P (Purpose)—How does dry forage harvesting help to coordinate use of pasture resources in pastoral systems facing uncertain and variable rainfall?

A (Assumptions)—We developed a game experiment called "GreenReserve" in which six to eight players make herding decisions in a shared 8 × 8 grid of pasture cells. Players identify what part of this grid is "shared reserve" in which all may

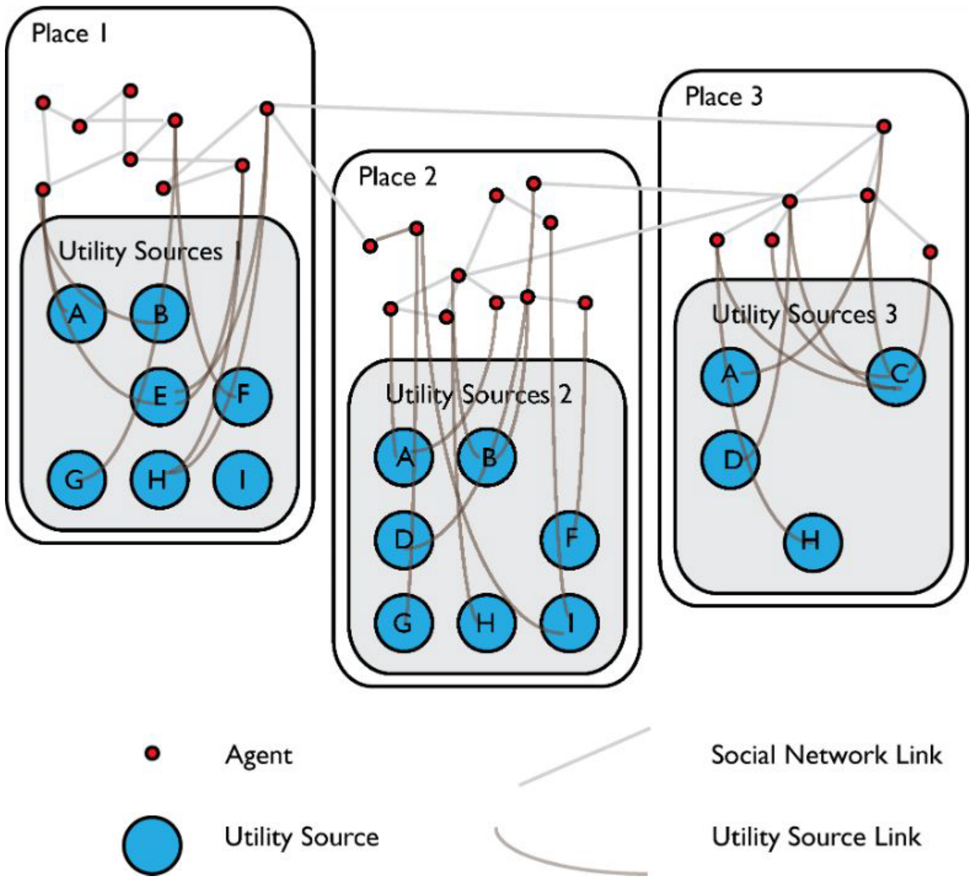

Figure A1. Agents drawing on utility sources and connected through social networks within and across places in MIDAS.

use the land subject to unanimous agreement of all players, what part of the grid is "private" or homestead land for each individual player, and what part of the grid is open and unrestricted. Players then make decisions about how to apportion scarce household effort between moving household animals around the pasture space to graze and harvesting surplus fresh forage to set aside in place as dry forage. Pasture grass growth is proportional both to how much grass is present, and how much rain has fallen, which is stochastic and only partially knowable to players. Communication and coordination among players is allowed at all times.

V (Validity)—We have several indicators of construct validity in GreenReserve. First, we conducted a community debrief in all communities after our game play, with participants noting the many ways in which the structure of the game matched the dilemmas that they face. One particular example excerpted from the manuscript: *"This game represents our daily lives. We have a community reserve and during the rainy season no one is allowed to herd near it. We also have helpers just like in the game. When it rained in a certain area, we discussed among ourselves and decided how we would move. When there was no rain, we returned to our own private reserves or near a water source or the community land to graze."* Second, we conducted a discrete choice experiment (where respondents demonstrate preferences by choosing among sets of differing, hypothetical options) of herd effort and structure before and after game play, and observed a significant shift in attention after game play on herd losses, suggesting an alignment of in-game herding activities with real-life activities. Lastly, our pilot overlapped with a failed short rainy season, and we observed significant shifts in herding management across groups as the drought persisted, again indicating a linkage between in-game decisions and real-life herding decisions.

E (Exploration)—Our small study did not include different experimental treatments for the game (every participant played the same, single game condition) but we are able to learn some new things from heterogeneity analysis (identifying differences in participant characteristics that explain variation in game outcomes), some of which also contributes to our demonstration of validity. We observed that game groups that included more women had fewer conflicts (where the group disagreed about whether some players should use the shared reserve or not), and also used less of the open grass area and invested less in grass harvesting. We observed that those who revealed in a discrete choice experiment (DCE) that they had concerns about bad-year herd losses tended to invest less in grass harvest, but used more of the shared reserve and ended the game with more surviving animals. We also observed that those who revealed in the DCE that they were more willing to commit labor to their herds were more likely to harvest grass for later use.

A2.2 PAVE summary for "Migration towards Bangladesh coastlines projected to increase with sea-level rise through 2100" (Bell et al. 2021)

P (Purpose)—What will be the effect of increased sea level rise (SLR) on migration toward the coast in Bangladesh over the coming decades?

A (Assumptions)—This application of MIDAS (Migration, Intensification, and Diversification as Adaptive Strategies) models

Bangladesh as several thousand agents living across Bangladesh's 64 districts, occupying a portfolio of "utility sources" (think jobs, assets, access to nature—things that people derive value from) in the place that they live. Agents share resources and information with each other across social networks (figure A1), and make regular decisions about what the best portfolio of opportunities is for them; when the best portfolio of opportunities is in another place, agents migrate. Agents consider the costs of accessing new things alongside possible incomes, so that these decisions embed properties of push (losses in opportunities in the current place), pull (available opportunities elsewhere), and mooring (immovable opportunities in the current place) of push-pull-mooring (PPM) theory.

Agents are boundedly rational, choosing from across a small number of options, comparing their expectations for each option into the future using a prospect-theory-based utility maximizing decision. Opportunities in this MIDAS application are drawn from three waves of the Bangladesh Income and Expenditure Survey (BIES, 2005–2015), distilled down to 13 different sectors with four different levels per sector, and resolved for each district.

Flood shocks are estimated using the Global Tides and Surge Reanalysis, and measured as the likelihood of exceeding flood depths from 1 to 5 m, by district by year under three different Representative Concentration Pathways (RCPs)—2.6, 4.5, and 8.5. Damages from these floods are estimated using best available historical data, an econometric estimation of the damage in wages experienced by both agricultural and nonagricultural sectors in Bangladesh in the immediate period and year following the 1998 "flood of the century" flooding event. This function for wage damages is combined with the flood shock prediction, and then subtracted from the wages predicted by our BIES model to provide an estimate of flood-impacted wages over time.

V (Validity)—We calibrated MIDAS to reproduce relative interdistrict flow rates (i.e., proportion of moves between districts i and j, across all i and j), weighted by population in places i and j, and averaged across the period 2002–2011, for which data were available from the Bangladesh Sample Vital Registration Study, explaining 26% of observed variation. We employed an Approximate Bayesian Computation (ABC) approach over approximately 50 parameters. In an ABC approach, parameter values are drawn randomly from a defined range for a large number of simulations; the parameter values from a small number of high-performing simulation runs are used to refine the appropriate range for each parameter in a new round of simulations. This process of simulation and refining is repeated over several rounds of simulation until no further improvement in calibration performance is observed. Migration rates are highly correlated in space and time, making it impractical to separate out this dataset for independent calibration and validation.

The calibrated application thus offers our best estimate of expected migration response to changes in wage resulting from flooding, based on migration response to wages during the early 2000s and assuming (i) flood response would be similar to the response to 1998 flood, and (ii) economic opportunities out into future decades would be similar to those available from 2005 to 2015. It would be inappropriate to apply outputs from this model application under conditions in future where these assumptions do not hold (e.g., economic impacts of sea level change do not mirror past examples of flooding, or economic opportunities deviate from conditions observed over 2005–2015).

E (Exploration)—We compared overall migration rates toward the coast in Bangladesh under different climate forcing RCPs and different degrees of SLR. We found that migration toward the coast persisted even under strong flooding, as our predicted wage damages were not sufficient to diminish coastal opportunities. We also found that providing credit to agents did not always lead them to move out of risky places, instead leading them in some cases to invest in place and further moor themselves to their location.

A2.3 PAVE summary for "The PAVE pathway: how to talk about models with people who built them, and people who didn't" (current paper)

P (Purpose)—To provide an effective method to communicate models between modelers and nonmodelers, and to allow nonmodelers to interpret the utility of the model findings.

A (Assumptions)—The key assumption of the PAVE (Purpose, Assumptions, Validity, and Exploration) pathway is that the gap in model communication between modelers and nonmodelers can be bridged. The existing protocols for model documentation were designed for modelers, and thus those protocols may be ineffective in communicating models between modelers and nonmodelers. We assume that modelers can unambiguously reinterpret their language so that the information provided to nonmodelers is understandable, sufficient, and compact.

V (Validity)—To be determined.

(A first signal of the validity of PAVE as a construct for communicating model findings might be the successful publication of this manuscript, surviving a peer-review process including those who might consider themselves modelers as well as nonmodelers. But peer review is blind, and the unknowable cannot stand as validation.

Instead, we will look toward the outcomes of future PAVE applications, across modeling domains and across various modeler and nonmodeler partnerships, for signals of the degree to, and generalizability with which PAVE communicates model findings and builds shared understandings across actors.)

E (Exploration)—In the absence of any demonstration of validity (V), the PAVE pathway dictates that it is not possible to present exploration (E).

References

Abraham J, Diller K. 2019. A review of hot beverage temperatures—satisfying consumer preference and safety. *Journal of Food Science* 84: 2011–2014. https://doi.org/10.1111/1750-3841.14699

Arnejo Z, Gaudou B, Saqalli M, Bantayan N. 2025. Communicating agent-based models to stakeholders: A scoping review. *Journal of Artificial Societies and Social Simulation* 28: 2. https://doi.org/10.18564/JASSS.5623

Augusiak J, Van den Brink PJ, Grimm V. 2014. Merging validation and evaluation of ecological models to "evaludation": A review of terminology and a practical approach. *Ecological Modelling* 280: 117–128.

Ayllón D, et al. 2021. Keeping modelling notebooks with TRACE: Good for you and good for environmental research and management support. *Environmental Modelling and Software* 136: 104932. https://doi.org/10.1016/j.envsoft.2020.104932

Bell AR, et al. 2021. Migration towards Bangladesh coastlines projected to increase with sea-level rise through 2100. *Environmental Research Letters* 16: 024045. https://doi.org/10.1088/1748-9326/abdc5b

Bell AR, Rakotonarivo SO, Zhang W, De Petris C, Kipchumba A, Meinzen-Dick RS. 2025. Understanding pastoralist adaptations to drought via games and choice experiments: Field testing among Borana communities. *Ecology and Society* 30: 25.

Borowski I, Hare M. 2007. Exploring the gap between water managers and researchers: Difficulties of model-based tools to support practical water management. *Water Resources Management* 21: 1049–1074. https://doi.org/10.1007/s11269-006-9098-z

Boschetti F, Hughes M, Jones C, Lozano-Montes H. 2018. On decision makers' perceptions of what an ecological computer model is, what it does, and its impact on limiting model acceptance. *Sustainability (Switzerland)* 10: 2767. https://doi.org/10.3390/su10082767

Brugnach M, Tagg A, Keil F, de Lange WJ. 2007. Uncertainty matters: Computer models at the science-policy interface. *Water Resources Management* 21: 1075–1090. https://doi.org/10.1007/s11269-006-9099-y

Caldwell-Harris CL. 2019. Language learning following immigration: Modeling choices and challenges. Pages 95–123 in Diallo SY, Wildman WJ, Shults FL, Tolk A, eds. *Human Simulation: Perspectives, Insights, and Applications*. Springer.

Daly AJ, De Visscher L, Baetens JM, De Baets B. 2022. Quo vadis, agent-based modelling tools? *Environmental Modelling & Software* 157: 105514.

Doyle EEH, Johnston DM, Smith R, Paton D. 2019. Communicating model uncertainty for natural hazards: A qualitative systematic thematic review. *International Journal of Disaster Risk Reduction* 33: 449–476. https://doi.org/10.1016/j.ijdrr.2018.10.023

Edmonds B, Le Page C, Bithell M, Chattoe-Brown E, Grimm V, Meyer R, Montañola-Sales C, Ormerod P, Root H, Squazzoni F. 2019. Different modelling purposes. *Journal of Artificial Societies and Social Simulation* 22: 6. http://jasss.soc.surrey.ac.uk/22/3/6.html.

Elsawah S, Bakhanova E, Hämäläinen RP, Voinov A. 2023. *A Competency Framework for Participatory Modeling, Group Decision and Negotiation*. Springer. https://doi.org/10.1007/s10726-023-09818-0

Esterling KM, Brady D, Schwitzgebel E. 2023. The necessity of construct and external validity for generalized causal claims: A critical review of the literature on quantitative causal inference. *Political Science Association, and the WZB Talks Series* 18: 1–5.

Fagiolo G, Moneta A, Windrum P. 2007. A critical guide to empirical validation of agent-based models in economics: Methodologies, procedures, and open problems. *Computational Economics* 30: 195–226. https://doi.org/10.1007/s10614-007-9104-4

Filatova T, et al. 2025. AGENTBLOCKS : A community platform for sharing, comparing, and improving reusable building blocks for (agent-based) models. *Journal of Artificial Societies and Social Simulation* 28: 11.

Fischhoff B, Davis AL. 2014. Communicating scientific uncertainty. *Proceedings of the National Academy of Sciences of the United States of America* 111: 13664–13671. https://doi.org/10.1073/pnas.1317504111

Frisch R. 1933. Editor's note. *Econometrica* 1: 1–4. www.jstor.org/stable/1912224

Fushiki T. 2011. Estimation of prediction error by using K-fold cross-validation. *Statistics and Computing* 21: 137–146. https://doi.org/10.1007/s11222-009-9153-8

Gray S, Paolisso M, Jordan R, Gray S. 2017. *Environmental Modeling with Stakeholders Theory, Methods, and Applications* 392. Springer Cham. https://doi.org/10.1007/978-3-319-25053-3

Grimm V, et al. 2006. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling* 198: 115–126.

Grimm V, et al. 2014. Towards better modelling and decision support : Documenting model development, testing, and analysis using TRACE. *Ecological Modelling* 280: 129–139. https://doi.org/10.1016/j.ecolmodel.2014.01.018

Grimm V, Berger U, DeAngelis DL, Polhill JG, Giske J, Railsback SF. 2010. The ODD protocol: A review and first update. *Ecological Modelling* 221: 2760–2768. https://doi.org/10.1016/j.ecolmodel.2010.08.019

Grimm V, Johnston ASA, Thulke HH, Forbes VE, Thorbek P. 2020a. Three questions to ask before using model outputs for decision support. *Nature Communications* 11: 10–12. https://doi.org/10.1038/s41467-020-17785-2

Grimm V, et al. 2020b. The ODD protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism. *Journal of Artificial Societies and Social Simulation* 23: 7.

Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, Railsback SF, Thulke H-H, Weiner J, Wiegand T, DeAngelis DL. 2005. Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science* 310: 987–991. https://doi.org/10.1126/science.1116681

Guillaume JHA, Helgeson C, Elsawah S, Jakeman AJ, Kummu M. 2017. Toward best practice framing of uncertainty in scientific publications: A review of Water Resources Research abstracts. *Water Resources Research* 53: 6744–6762. https://doi.org/10.1002/2017WR020609

Hall DM, Lazarus ED, Swannack TM. 2014. Strategies for communicating systems models. *Environmental Modelling and Software* 55: 70–76. https://doi.org/10.1016/j.envsoft.2014.01.00

Herman BD, Mckay SK, Altman S, Richards NS, Reif M, Piercy CD, Swannack TM. 2019. Unpacking the black box: Demystifying ecological models through interactive workshops and hands-on learning. *Frontiers in Environmental Science* 7: 1–14. https://doi.org/10.3389/fenvs.2019.00122

Hutton EWH, Piper MD, Tucker GE. 2020. The Basic Model Interface 2. 0: A standard interface for coupling numerical models in the geosciences. *The Journal of Open Source Software* 5: 10–11. https://doi.org/10.21105/joss.02317

Jakeman AJ, Elsawah S, Wang H-H, Hamilton S, Melsen L, Grimm V. 2024. Towards normalizing good practice across the whole modeling cycle: Its instrumentation and future research topics. *Socio-Environmental Systems Modelling* 6: 18755. https://doi.org/10.18174/sesmo.18755

Janssen JAEB, Hoekstra AY, de Kok JL, Schielen RMJ. 2009. Delineating the model-stakeholder gap: Framing perceptions to analyse the information requirement in river management. *Water Resources Management* 23: 1423–1445. https://doi.org/10.1007/s11269-008-9334-9

Jiménez-Buedo M, Russo F. 2021. Experimental practices and objectivity in the social sciences: Re-embedding construct validity in the internal–external validity distinction. *Synthese* 199: 9549–9579. https://doi.org/10.1007/s11229-021-03215-3

Jones HFE, Özkundakci D, Hunt S, Giles H, Jenkins B. 2020. Bridging the gap: A strategic framework for implementing best practice guidelines in environmental modelling. *Environmental Science and Policy* 114: 533–541. https://doi.org/10.1016/j.envsci.2020.09.030

Kim D. 1999. *Introduction to Systems Thinking*. Pegasus Communications.

Kolokathis PD, et al. 2024. Easy-MODA: Simplifying standardised registration of scientific simulation workflows through MODA template guidelines powered by the Enalos Cloud Platform. *Computational and Structural Biotechnology Journal* 25: 256–268. https://doi.org/10.1016/j.csbj.2024.10.018

Kwakkel JH, Pruyt E. 2013. Exploratory modeling and analysis, an approach for model-based foresight under deep uncertainty. *Tech-*

nological Forecasting and Social Change 80: 419–431. https://doi.org/10.1016/j.techfore.2012.10.005

Kwakkel JH. 2017. The exploratory modeling workbench: An open source toolkit for exploratory modeling, scenario discovery, and (multi-objective) robust decision making. *Environmental Modelling and Software* 96: 239–250. https://doi.org/10.1016/j.envsoft.2017.06.054

Lim WM. 2024. A typology of validity: Content, face, convergent, discriminant, nomological and predictive validity. *Journal of Trade Science* 12: 155–179. https://doi.org/10.1108/jts-03-2024-0016

Loblay V, Freebairn L, Occhipinti JA. 2023. Conceptualising the value of simulation modelling for public engagement with policy: A critical literature review. *Health Research Policy and Systems* 21: 1–14. https://doi.org/10.1186/s12961-023-01069-4

Lorscheid I, Heine BO, Meyer M. 2012. Opening the "black box" of simulations: Increased transparency and effective communication through the systematic design of experiments. *Computational and Mathematical Organization Theory* 18: 22–62.

Mannix DH, Birkenholtz TL, Abrams DB, Cullen C. 2022. Uncertain waters: Participatory groundwater modelling in Chicago's suburbs. *Geoforum* 132: 182–194. https://doi.org/10.1016/j.geoforum.2021.09.006

Marchau Vincent AWJ, Walker Warren E, Bloemen Pieter JTM, Popper Steven W. eds. 2019. *Decision Making under Deep Uncertainty*. Springer.

Martens JP. 2020. Theories, models, and effects-oh my! differentiating similar constructs. *Transformative Dialogues: Teaching and Learning Journal* 13: 64–78.

Monks T, Currie CSM, Onggo BS, Robinson S, Kunc M, Taylor SJE. 2019. Strengthening the reporting of empirical simulation studies: Introducing the STRESS guidelines. *Journal of Simulation* 7778: 1–13. https://doi.org/10.1080/17477778.2018.1442155

Müller B, Bohn F, Dreßler G, Groeneveld J, Klassert C, Martin R, Schlüter M, Schulze J, Weise H, Schwarz N. 2013. Describing human decisions in agent-based models—ODD+D, an extension of the ODD protocol. *Environmental Modelling and Software* 48: 37–48. https://doi.org/10.1016/j.envsoft.2013.06.003

Peck SL. 2004. Simulation as experiment: A philosophical reassessment for biological modeling. *Trends in Ecology and Evolution* 19: 530–534.

Polhill GJ, Parker DC, Brown DG, Grimm V. 2008. Using the ODD protocol for describing three agent-based social simulation models of land use change. *Journal of Artificial Societies and Social Simulation* 11: 3. <0:italic >http://jasss.soc.surrey.ac.uk/11/2/3.html</0:italic>

Rosello C, Guillaume JHA, Taylor P, Cuddy SM, Pollino CA, Jakeman AJ. 2025. Towards good practice in engaging users in evaluation of computer model software: Introducing the critical appraisal approach (CAA). *Environmental Modelling and Software* 190: 106469. https://doi.org/10.1016/j.envsoft.2025.106469

Schmolke A, Thorbek P, DeAngelis DL, Grimm V. 2010. Ecological models supporting environmental decision making: A strategy for the future. *Trends in Ecology and Evolution* 25: 479–486. https://doi.org/10.1016/j.tree.2010.05.001

Secchi D, Grimm V, Herath DB, Homberg F. 2024. Modeling and theorizing with agent-based sustainable development. *Environmental Modelling and Software* 171: 105891.

Star SL. 2010. This is not a boundary object: Reflections on the origin of a concept. *Science Technology and Human Values* 35: 601–617. https://doi.org/10.1177/0162243910377624

Szangolies L, Rohwäder M, Ahmed H, Jahanmiri F, Wagner A, Soutoveiga R, Grimm V, Cara A. 2024. Visual ODD : A standardised visualisation illustrating the narrative of agent-based models. *Journal of Artificial Societies and Social Simulation* 27: 1. https://doi.org/10.18564/jasss.5450

Thiele JC, Kurth W, Grimm V. 2014. Facilitating parameter estimation and sensitivity analysis of agent-based models: A cookbook using NetLogo and R. *Journal of Artificial Societies and Social Simulation* 17: 11.

Troost C, et al. 2023. How to keep it adequate: A protocol for ensuring validity in agent-based simulation. *Environmental Modelling and Software* 159: 105559. https://doi.org/10.1016/j.envsoft.2022.105559

van der Vaart E, Beaumont MA, Johnston ASA, Sibly RM. 2015. Calibration and evaluation of individual-based models using approximate Bayesian computation. *Ecological Modelling* 312: 182–190. https://doi.org/10.1016/j.ecolmodel.2015.05.020

Vincenot CE. 2018. How new concepts become universal scientific approaches: Insights from citation network analysis of agent-based complex systems science. *Proceedings of the Royal Society B: Biological Sciences* 285: 20172360.

Voinov A, Kolagani N, McCall MK, Glynn PD, Kragt ME, Ostermann FO, Pierce SA, Ramu P. 2016. Modelling with stakeholders—next generation. *Environmental Modelling and Software* 77: 196–220. https://doi.org/10.1016/j.envsoft.2015.11.016

Will M, Dressler G, Kreuer D, Thulke HH, Grêt-Regamey A, Müller B. 2021. How to make socio-environmental modelling more useful to support policy and management? *People and Nature* 3: 560–572. https://doi.org/10.1002/pan3.10207

Windrum P, Fagiolo G, Moneta A. 2007. Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation* 10: 8. https://doi.org/Article

Zellner ML. 2024. Participatory modeling for collaborative landscape and environmental planning: From potential to realization. *Landscape and Urban Planning* 247: 105063. https://doi.org/10.1016/j.landurbplan.2024.105063

Zurell D, et al. 2020. A standard protocol for reporting species distribution models. *Ecography* 43: 1261–1277. https://doi.org/10.1111/ecog.04960

Received: June 5, 2025. **Revised:** January 12, 2026. **Accepted:** January 22, 2026