Assessing the quality of land system models: moving from *valibration* to *evaludation*

C. Brown[1,2*], J. D. Millington[3], and M. Rounsevell [1,4,5]

[1] Institute of Meteorology and Climate Research, Atmospheric Environmental Research (IMK-IFU), Karlsruhe Institute of Technology, Kreuzeckbahnstraße 19, 82467 Garmisch-Partenkirchen, Germany

[2] Highlands Rewilding Limited, The Old School House, Bunloit, Drumnadrochit IV63 6XG, UK

[3] Department of Geography, King's College London, Aldwych, London, WC2B 4BG, UK

[4] Institute of Geography and Geo-ecology, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

[5] School of Geosciences, University of Edinburgh, Drummond Street, Edinburgh EH8 9XP, UK

Abstract

Reviews suggest that evaluation of land system models is largely inadequate, with undue reliance on a vague concept of validation. Efforts to improve and standardise evaluation practices have so far had limited effect. In this article we examine the issues surrounding land system model evaluation and consider the relevance of the TRACE framework for environmental model documentation. In doing so, we discuss the application of a comprehensive range of evaluation procedures to existing models, and the value of each specific procedure. We develop a tiered checklist for going beyond what seems to be a common practice of 'valibration' (the repeated variation of model parameter values to achieve agreement with data) to achieving 'evaludation' (the rigorous, broad-based assessment of model quality and validity). We propose the Land Use Change – TRACE (LUC-TRACE) model evaluation protocol and argue that engagement with a comprehensive protocol of this kind (even if not this particular one) is valuable in ensuring that land system model results are interpreted appropriately. We also suggest that the main benefit of such formalised structures is to assist the process of critical thinking about model utility, and that the variety of legitimate modelling approaches precludes universal tests of whether a model is 'valid'. Evaludation is therefore a detailed and subjective process requiring the sustained intellectual engagement of model developers and users.

Keywords

Validation, evaluation, CRAFTY, agent-based model, land use change

1. Introduction

In their review of ecological modelling, Augusiak et al. (2014) concluded that the term 'validation' had become so vague as to be "useless for any practical purpose" (p.117). As models became more complex, numerous and influential, Augusiak et al. (2014) found that flawed approaches to validation meant that those models were increasingly likely to mislead the scientists and decision-makers who used them. The same argument can be applied to land system modelling. Models of land use change have attained a crucial position at the science-policy interfaces of the IPCC and IPBES, but recent reviews suggest that model evaluation has not kept pace with model development. Instead, rigorous model evaluation has been the exception rather than the rule, hampered

Correspondence:
Contact C. Brown at calum.brown@kit.edu

Cite this article as:
Brown, C., Millington, J.D., & Rounsevell, M.
Assessing the quality of land system models: moving from *valibration* to *evaludation*
Socio-Environmental Systems Modelling, vol. 5, 18434, 2023, doi:10.18174/sesmo.18434

Socio-Environmental Systems Modelling
An Open-Access Scholarly Journal
http://www.sesmo.org

– as in ecology – by widespread acceptance of inadequate validation practices (Brown et al., 2017; Rosa et al., 2014; van Vliet et al., 2016).

Developing a consistent approach to model evaluation is particularly difficult in land system science because of the range of uses to which models are put. From abstract experimentation to detailed predictions of global land use change over several decades, and from simple linear regression to qualitative, participatory methods, models are too diverse to fit with any single concept of validity (Bianchi & Squazzoni, 2019; Edmonds & ní Aodha, 2019; van Soesbergen, 2016). Furthermore, the central role of people within land systems means that evaluation cannot involve simple comparisons to predictable outcomes. Indeed, because validity can only be established with respect to a model's purpose, there can be no absolute universal measure, but rather a process that must involve model developers and users (Barlas, 1996; Edmonds et al., 2019; Hamilton et al., 2019; Troost & Berger, 2020; Troost et al., 2023).

While recognising these fundamental limitations to assessing the quality of land system models, several papers have defined alternative approaches to model validation (Baldos & Hertel, 2013; Hamilton et al., 2019; Ngo, The An & See, 2012; Troost et al., 2023; van Vliet et al., 2016). These have not yet coalesced into generally accepted protocols, or indeed achieved widespread usage. In this article, we explore and further develop one of these, the TRACE model documentation protocol (Grimm et al., 2014; Schmolke et al., 2010), with particular consideration of land system models and the representation of human behaviour within them. We discuss the relevance of each element in the protocol with reference to existing models and, in the Supplementary Material, to one specific agent-based (behavioural) modelling framework, CRAFTY (Murray-Rust et al., 2014), drawing conclusions about the utility of the evaluation protocol.

Box 1: Terminology

The first challenge in model evaluation relates to terminology. As noted by Augusiak et al. (2014), the term 'validation' has itself become a hindrance, obscuring an enormous range of model evaluation practices that might be assumed to produce a universally comparable, binary outcome (valid or not valid). Formal definitions are rare, and not widely accepted. To try and maximise clarity, we adopt the following definitions of key terms within this manuscript:

Validation: Establishing whether the model is a 'valid' (in the sense of 'accurate' or 'reasonable', but not in the sense of 'true') representation of whatever it is that is being modelled, usually through comparison of model outputs to independent data.

Evaluation: A broad assessment of model quality, potentially including conceptual design, technical implementation, performance and relevance to the application.

Calibration: The use of data (particularly concerning patterns) to inversely parameterise a model.

'Valibration': We propose this term to describe the practice of repeatedly varying model parameter values (effectively 'vibrating' model settings) until agreement with test data (which are not always independent of calibration) is produced, thus conflating validation and calibration. Valibration can in principle be combined with other forms of model evaluation.

'Evaludation': Proposed by Augusiak et al. (2014) as "a fusion of 'evaluation' and 'validation' to describe the entire process of assessing a model's quality and reliability" (p. 117). In this definition, evaludation involves distinct steps of i) data evaluation, ii) conceptual model evaluation, iii) implementation verification, iv) model output verification, v) model analysis and vi) model output corroboration. These steps were subsequently formalised in the TRACE protocol by Grimm et al. (2014), building on the initial TRACE documentation of Schmolke et al. (2010) to create a tool for planning, performing and documenting model evaludation. In this revised form, TRACE stands for 'TRAnsparent and Comprehensive model Evaludation' (Grimm et al., 2014), and was recently recommended as a basis for keeping modelling notebooks by Ayllón et al. (2021).

Land system models: We focus on models of the land system in a broad sense, and so include models of various types dealing with land use change, its drivers and impacts, from a systemic perspective. This definition is purposefully broad and means that our discussion is relevant for models of other systems as well, but is chosen to reflect the emerging identity of land system science, and to acknowledge the particular role of human behaviour in this field. Not all elements of the TRACE protocol are relevant to all models.

2. Current practice and proposed improvements

2.1 Land system model evaluation

Currently, few land system models are rigorously evaluated. In their review of calibration and validation in land-change modelling, Van Vliet et al. (2016) found that 31% of model applications did not report any form of validation. Brown et al. (2017) found an even higher percentage without validation: 55% of the behavioural models they reviewed. Where validation does occur, it usually comprises a single exercise involving testing model fit to data, often data on which the model has already been calibrated (Brown et al., 2017; van Vliet et al., 2016).

This narrow form of evaluation is inadequate for assessing the 'validity' of any model in a general sense. A single comparison to data invites what we have called 'valibration' - the repeated variation of model parameter values until the validation test is passed - or what Grimm et al. (2014) called 'tweaking', in which environmental settings and submodel formulation are tuned alongside parameter values. Such comparisons also favour overfitting and will tend to reward complicated models with more variables. Fitting to data also prioritises replication of particular, usually locational, observations over potentially more general accuracy based on theoretical or empirical information. Perhaps most worryingly, subsequent use of the model, for example in scenario analysis, is likely to require extrapolation beyond the scope of the data used for validation (Rounsevell et al., 2021). Augusiak et al. (2014), building on Oreskes et al. (1994), identify three further shortcomings of validation of this kind: 1) the possibility that agreement with test data is spurious, achieved through inaccurate model design or parameterisation; 2) the likelihood that suitable test data are themselves inaccurate or even unavailable; and 3) the degradation of the complex, debated concept of validation to a misleadingly simple test. Recent papers have suggested further risks: that theory-based models may be under-valued despite being just as useful (and, potentially, accurate) as data-based models (Taghikhah et al., 2021), and that the practice of withholding test data from the training or calibration phase may compromise model robustness (Arsenault et al., 2018). Finally, exclusively technical evaluation procedures are unlikely to be sufficient for the decisions faced by model users, which vary from case to case and user to user (Hamilton et al., 2019).

Many authors have suggested improvements in validation and evaluation practices. Fundamentally, most emphasise tailored assessment of the model with respect to its specific purpose(s) (e.g. Augusiak et al., 2014; Edmonds et al., 2019; Hamilton et al., 2019; Oreskes et al., 1994; Troost & Berger, 2020). These purposes can vary enormously, for instance depending on whether the model is being used by a research scientist, stakeholder or policy maker (Hamilton et al., 2019; Millington et al., 2011). Another important distinction exists between models intended to represent a system in order to improve understanding of its dynamics, and models intended to simulate alternative scenarios of future or other unobservable conditions. While the latter is a major purpose of land system modelling (Roe et al., 2019; Rogelj et al., 2018), it poses specific problems for evaluation because the accuracy of a model's results can only be definitively established if and when the scenario becomes reality (Edmonds & ní Aodha, 2019; Polhill, 2018). Models 'locked into' past land use dynamics by design or valibration are unlikely to accurately represent the diversity of possible future worlds (Brown et al., 2022).

Establishing a model's validity for its intended use therefore requires a variety of approaches, and several examples or protocols have been suggested. Tesfatsion (2017) identifies four aspects of empirical validation that modellers should target simultaneously: input validation, process validation, descriptive output validation and predictive output validation. In land systems science, Troost et al. (2023) define a protocol to identify relevant methods of evaluation depending on model purpose. Newland et al. (2018) propose a multi-objective optimisation method for calibrating and evaluating one class of model (Cellular Automata). In ecology, Augusiak et al. (2014) proposed a set of six 'evaludation' steps (see Box 1) as being essential, and these have since been formalised in the TRACE framework (Grimm et al., 2014, extending the work of Schmolke et al., 2010). Hamilton et al. (2019, 2022) emphasise the role of diverse model users in a process of evaluation that is not only technical in nature, proposing a framework for adaptive learning in model development and application. Others have developed approaches relevant within these steps, such as methods for deciding which model(s) to use in cases of equifinality, where multiple models have similar performance for different reasons (e.g. Williams et al. (2020), who emphasise the value of a diversity of models in such cases).

At some point, the use of complicated evaluation protocols faces practical constraints. But there are also theoretical concerns. In fact, model validity may be impossible to definitively determine whatever the range of methods brought to bear. Steinmann et al. (2020) recommend the adoption of "exploratory modelling, global sensitivity analysis, and robust decision-making" to address this challenge. Several researchers have identified desirable model characteristics as independent of evaluation results; for instance, retaining or increasing behavioural richness in representations of land use decision-making (Arneth et al., 2014; Polhill & Salt, 2017). Related arguments stress the primacy of model validity and utility at aggregate levels, where a diversity of approaches among models allows a more comprehensive understanding to be developed across them (Brown et al., 2021).

Some of these arguments are concerned with avoiding the "natural selection of bad science", in which the advantages that accrue to 'successful' models in competitive research environments causes methodological deterioration even where those successful models themselves are sound (Smaldino & McElreath, 2016). This is a process that some have identified with the dominance of a small number of established Integrated Assessment Models in funding and research on land-based climate change mitigation (Gambhir et al., 2019; Hughes & Paterson, 2017; Low & Schäfer, 2020). Reversing such a process would require a conscious and careful response involving deliberate diversification of modelling approaches, even where some approaches are initially evaluated as inferior. A touch of Feyerabend's epistemological anarchism (Feyerabend, 1993) is apparent here, in the implication that free exploration of alternative and (initially) unjustified approaches may bear more fruit in the long term. This implies that even the most rigorous model evaluation can never be a definitive guide to model utility, and could in some circumstances be worse than no evaluation at all if it leads to concentration of methods and premature rejection of poorly performing models. In the same way that more diverse ecosystems are more robust to perturbations, more diverse model systems may be more robust to changes in knowledge and land system drivers in the future (Brown et al., 2022).

2.2 The special case of agent-based models?

Agent-based models (ABMs) represent system dynamics as emergent from the behaviour and interactions of individual entities, and are favoured by some modellers for their greater apparent representational realism than optimisation-based land system models, which reduce behaviour to aggregate economic responses. They also often have distinct purposes, commonly being used for exploration or understanding of system dynamics rather than reproducing and predicting particular outcomes; attempting to generate and explain observed phenomena rather than to replicate them (Bianchi & Squazzoni, 2019; Boone & Galvin, 2014; Epstein, 1999).

However, ABMs are also viewed with scepticism by many researchers. Giupponi et al. (2022) describe four main common concerns about land system ABMs: 1) they are excessively complicated for anything except specific, small-scale studies; 2) they have not generated transferrable modelling frameworks that can be applied by different researchers to different cases; 3) they are often complex and poorly-described 'black box' models, which are therefore hard to interpret or re-use; and 4) they are difficult to empirically validate. These concerns are widely held, and often expressed in reductionist form with reference to the 'YAAWN syndrome' (Yet Another Agent-based model … Whatever … Nevermind …) (O'Sullivan et al., 2016); originally a call for more coherent development across applications, including through contributions to general theoretical insights, that has in some cases devolved towards an arbitrary dismissal of ABM research.

Not all of these criticisms are reasonable. The CRAFTY framework discussed in this paper (see Supplementary Material) is one example of a relatively simple, generic ABM framework that has been applied around the world, from local to continental scales, coupled to other models, while also being fully open-access, thoroughly documented and widely evaluated using different methods (Brown et al., 2019; Millington et al., 2021; Murray-Rust et al., 2014). Other examples include the Multi-Agent Research and Simulation framework MARS (Clemen et al., 2021; Hüning et al., 2016) and specific integrations between fine-scaled ABMs and coarse-scaled global models (Niamir et al., 2020). Some researchers suggest that ABMs are the subject of residual prejudice that statistical data-based models are more rigorous and scientific than theory-based models (Bianchi & Squazzoni, 2019), and of scientific groupthink in coalescing around more dominant (though intrinsically no more robust or transparent) approaches (Gambhir et al., 2019). At a basic level, land system ABMs operate in the same domain as any other type of land use model, and actually suffer from very similar revealed shortcomings in the practice of model evaluation (Brown et al., 2017; van Vliet et al., 2016).

Nevertheless, ABMs do have specific characteristics that make them unusual, and sometimes qualitatively distinct, subjects for evaluation. To the extent that empirical ABMs are more complicated than other land system models, they are disproportionately harder to calibrate and evaluate because small increases in complicatedness (in the sense of having more components; Sun et al. (2016)) require large increases in supporting data (Srikrishnan & Keller, 2021). These difficulties may be overcome through the use of theoretical 'priors' (Taghikhah et al., 2021) or techniques such as surrogate modelling or Machine Learning to reduce computational costs (Storm et al., 2020; ten Broeke et al., 2021), but are still likely to represent substantial challenges.

More fundamentally, the different uses of ABMs require tailored forms of evaluation. The selection of an ABM approach in itself implies that agency and behaviour are of interest to the modeller, and that evaluation should focus more on these than it would for another type of model with different representational assumptions. Vermeer et al. (2022), in the context of health modelling, recommend a two-step procedure for evaluating high-fidelity ABMs, in which validated individual-level behaviours are drawn from field (observational) data, and then checked for their ability to produce realistic system-level dynamics that explain the phenomenon observed. Millington et al. (2011) argued that structural accuracy (a model specification that appropriately represents processes) might in fact be a more relevant goal for many agent-based modellers than 'mimetic accuracy' (model output that reproduces empirical events). This could be particularly appropriate when ABMs are used with stakeholders, who might intuitively grasp the behaviours these models represent better than they grasp abstract economic or statistical models (e.g. Naivinit et al., 2010). It is therefore possible to actively iterate between qualitative and quantitative forms of evaluation; a particularly powerful, if little used, approach to improve the value of these models because it helps to highlight and evaluate the assumptions that are being used as well as any empirical basis they might have (Grimm & Railsback, 2012; Millington & Wainwright, 2017). Grimm & Railsback (2012) distinguish between 'strong' and 'weak' patterns, which tend to be quantitative and qualitative, respectively, and argue that "multiple weak patterns, observed at different hierarchical levels and scales, can often achieve higher structural realism, with less effort, than focusing only on one strong pattern" (p.300).

Another important use of ABMs is to develop and explore theory, at levels of both system and individual behaviour and, crucially, at their cross-level interface, where agent dynamics translate into system dynamics (Lorscheid et al., 2019). As well as justifying the kind of two-step evaluation proposed by Vermeer et al. (2022), this can make theoretical evaluation more important than fits to data of the kind that might support empirical modelling (Gostoli & Silverman, 2020; Polhill & Salt, 2017). Methods exist to partially disentangle process accuracy from pattern accuracy (Brown et al., 2005), but are constrained by the fact that observational data can only ever describe a single realisation of history, and not the range of outcomes that underlying processes could possibly have produced (Polhill & Salt, 2017), and that these data tend to capture easily-observable outcomes (such as land cover) and not underlying processes (such as human decisions). Edmonds & ní Aodha (2019) identify agent-based modelling as particularly appropriate because it lends itself to the analysis of uncertainties for better understanding of the modelled system; an approach they term Reflexive Possibilistic Modelling.

3. The LUC-TRACE protocol

The TRACE protocol was developed to aid the documentation of ecological and environmental models, and further to provide "a tool for planning, performing, and documenting good modelling practice" (Augusiak et al., 2014; Grimm et al., 2014; Schmolke et al., 2010). Recently, Ayllón et al. (2021) used TRACE to standardise key aspects of model development, evaluation and documentation in modelling notebooks. Here, we build on these papers to propose the Land Use Change – TRACE (LUC-TRACE) model evaludation protocol (Table 1) as a comprehensive checklist for the 'evaludation' of models of land use change, with particular reference to agent-based models. The relevance of the protocol is not restricted to land system models, and indeed varies widely within this class depending on the exact type of model being evaluated, but does reflect the range of evaluation considerations that we believe to be important in land system science.

Table 1: Summary of the 'LUC-TRACE' based on the TRACE protocol outline of (Ayllón et al., 2021) with our additions in bold text (additions concerning model replication are drawn from Essawy et al., 2020). We assign each step of the protocol to a particular tier, with Tier 1 steps being essential, Tier 2 being expected, and Tier 3 being desirable. The right-hand column summarises the extent to which each step has been applied to the CRAFTY framework, with ticks representing complete application, circles representing partial application and crosses representing no application (details in the Supplementary Material). The application of each step is described more fully in the text.

Evaludation step & substeps		Motivating question	Type	Tier	Application to CRAFTY
1. Problem formulation		What is the model intended for? (To allow users to judge how appropriate it is)	Descriptive; free interpretation	1	✓
2. Model description		How does the model work?	Descriptive; free interpretation	1	✓
3. Data evaluation		What is the nature and quality of the data used to inform, develop, calibrate and evaluate the model?	Partially objective	2	O
4. Conceptual model evaluation	**System conceptualisation**	**How is the system conceptualised and are there matches & mismatches with the model?**	**Descriptive; free interpretation**	3	O
	Model design	How is the model conceptualised as a representation of the system?	Descriptive; free interpretation	2	✓
	System conceptualisation represented adequately by that design?	**Explicitly, how do the system & model conceptualisations align?**	**Descriptive; potentially some objective content**	3	O
	Problem relevance, e.g. Ability to handle scenario conditions	**Can the model be used for its intended purpose given the system & model conceptualisations/designs – are aspects of the problem left out?**	**Descriptive; potentially some objective content**	1	✓
5 Implementation verification	Debugging / code testing (unit testing)	Has there been comprehensive testing of individual sections of code to ensure it (only) functions as intended?	Objective; true/false	1	✓
	Software verification/ Testing	Does the model as a whole perform as intended?	Partially objective	1	✓
	Usability tools design	Can the model be used and interpreted correctly given its design and description?	Partially objective	2	O
6. Model output verification	Output verification/ Goodness-of-fit: data used in model development	Does the model fit the data used in its development?	Partially or wholly objective	3	O
	Output verification/ Goodness-of-fit: historical timeseries	Can the model reproduce timeseries?	Partially objective	3	O
	Calibration; Tests on environmental drivers	How was calibration used to achieve fit to data, and what parts/processes did it involve?	Partially objective	1	✓
7. Model analysis and application	Sensitivity & uncertainty analysis	What are the effects of model parameters on outputs?	Should be fully objective, quantitative	1	✓
	Robustness analysis; Simulation experiment	'Reasonableness' of model in known situation; do we understand how the outcomes arise?	Partially objective & quantitative, partly descriptive	2	✗
	Model stochasticity & stability	**What effects do model stochasticity and instability have on the results?**	**Objective**	1	✓

(*Table continued on next page*)

Table 1 (*continued*)

Evaludation step & substeps		Motivating question	Type	Tier	Application to CRAFTY
8. Model output corroboration	Fitting to data “Output corroboration / Validation”	Can the model replicate patterns in independent data, including spatio-temporal, aggregate or otherwise emergent patterns?	Objective but not a binary test	2	O
	Benchmarking against other models	Has there been comparison to independent data representing alternative modelling approaches?	Objective but not intended to assess accuracy	2	✓
9. Participatory/ companion modelling	Participatory model development/selection	Was the model developed or chosen through a process of user engagement?	Descriptive; free interpretation	3	✗
	Details of use in participatory settings	Has the model been used in a participatory setting and what were the outcomes?	Descriptive; free interpretation	3	O
	Communication of results	What methods were used to communicate results, and how well did they work?	Descriptive; free interpretation	3	O
10. Model replication	Repeatability	Does the model produce consistent results across multiple runs?	Objective	1	✓
	Runnability	Does the model produce consistent results on multiple computers?	Objective	2	O
	Reproducibility	Does the model produce consistent results when run by independent researchers?	Partially objective	3	✗
	Replicability	Are consistent results produced by entirely independent studies?	Partially objective	3	✗

3.1 Problem formulation

"The decision-making context in which the model will be used; the types of model clients or stakeholders addressed; a precise specification of the question(s) that should be answered with the model, including a specification of necessary model outputs; and a statement of the domain of applicability of the model, including the extent of acceptable extrapolations." (Grimm et al., 2014)

Problem formulation is an essential basis for model evaluation because it defines the model's purpose. Nevertheless, a tightly-defined problem formulation may be too prescriptive for land use models that are designed to be applied in different places for different reasons. Many land system models (and modelling frameworks in particular) are intended for use by a community of land system scientists and are general in scope, covering a range of social, economic and environmental factors and potentially being relevant to a wide range of questions and spatio-temporal domains. This open-endedness is a design feature in many cases; and one that may legitimately prevent complete or restrictive problem formulations from being developed by the model developer.

The CRAFTY framework was designed with a general problem formulation in mind: the need for a model "to be run over large spatial extents and to be capable of accounting for relevant forms of human behaviour, variations in land use intensities, multifunctional ecosystem service production and the actions of institutions that affect land use change" (Murray-Rust et al., 2014). Other land system models with tighter foci retain fairly broad problem statements. For instance, the Evoland modelling approach used by Guzy et al. (2008) "was designed to investigate alternative futures that may result from different policy approaches in social-ecological systems in the flood plains and riparian forests at the junction of large rivers", and the SLUDGE model documented by Polhill et al. (2008) "was designed to extend existing analytical microeconomic theory to examine relationships between externalities, market mechanisms, and the efficiency of free-market land use patterns" (for further examples and guidance, see Grimm et al., 2020). In general, few problem formulations give precise delimitations of model clients, stakeholders, domains or questions that should be addressed; implicitly, this is left open for model users to make informed decisions about.

We suggest that a relatively open-ended approach, which may define limits of appropriate usage rather than explicit domains, is broadly applicable to land use models. This allows model scope to be defined partly by broad objectives and partly by proscribed uses, but retains flexibility for model users to tailor the model to their own problems as they see fit. We therefore place this step in Tier 1 of our evaludation hierarchy, but relax its more prescriptive requirements.

3.2 Model description

"The model. Provide a detailed written model description. For individual/agent-based and other simulation models, the ODD protocol is recommended as standard format. For complex submodels it should include concise explanations of the underlying rationale. Model users should learn what the model is, how it works, and what guided its design." (Grimm et al., 2014)

A model description is essential for model evaluation, as well as for model use, reproducibility and interpretation. An ODD protocol is often recommended for process-based models, including land system models (Grimm et al., 2010, 2020), with the ODD+D extension of Müller et al. (2013) tailored specifically to ABMs representing human decisions in social-ecological systems. This protocol ensures a degree of standardisation and comparability in model descriptions, and prompts the author to consider a range of essential model elements. At the same time, any standard format implies some inflexibility when applied to very different models, prioritising structure over readability (although Grimm et al. (2020) make suggestions for more narrative and understandable ODDs to address this issue). It is therefore possible that other forms of model description could be more appropriate in some cases, including visual, narrative or interactive descriptions, which are currently very rare, and engagement with the model code itself, ideally through clear links between ODD or other descriptions and corresponding sections of code (Grimm et al., 2020). Of course, model descriptions are difficult to evaluate in their own right, and there can be no guarantee that they are read or understood by model users (we know of nobody who has used an ODD protocol to understand or evaluate the CRAFTY framework). Nevertheless, adoption of ODD has been widespread, suggesting that it is found to be useful as a template for model description (Grimm et al., 2020). While it is designed for process-based simulation models, we suggest that it can be applied to a full range of land system models if only to provide a basis for comparison of model components (with many expected to be absent from statistical land use change models, for example).

3.3 Data evaluation

"*The quality and sources of numerical and qualitative data used to parameterize the model, both directly and inversely via calibration, and of the observed patterns that were used to design the overall model structure. This critical evaluation will allow model users to assess the scope and the uncertainty of the data and knowledge on which the model is based.*" (Grimm et al., 2014)

Data evaluation is often a problem for land system models in general and agent-based models in particular. Data describing fine-grained land uses (as opposed to broad land covers) and, especially, the behaviours that determine their uptake, are hard to find at any quality (Schulze et al., 2017; Verburg et al., 2019). This is one main reason why ABMs are often theory-based, although that is of course a valid approach even where data are available. Where data are used, evaluation is clearly important, but may also be constrained by the original description of those data and the confidence with which they can be interpreted. In their review of ABM evaluation practices, Schulze et al. (2017) also identify a tension between the scale and specificity of available data derived from social surveys, interviews or other intensive forms of elicitation, and the general relevance that many ABMs are intended to have. This potential tension has been directly addressed, for example, by Magliocca et al. (2014), who test the fit of an ABM calibrated on particular data across different case studies, suggesting a role for the model itself in data evaluation. Given these possibilities and challenges, we regard data evaluation as being necessary where and to the extent possible, recognising that this is not always an appropriate step for land system models.

3.4 Conceptual model evaluation

"The simplifying assumptions underlying a model's design, both with regard to empirical knowledge and general, basic principles. This critical evaluation allows model users to understand that model design was not ad hoc but based on carefully scrutinized considerations." (Grimm et al., 2014)

We identify four elements within the scope of conceptual model evaluation: system conceptualisation, conceptual model design, the matches and any mismatches between those, and finally the matches and any mismatches with the particular problem the model will be used to address. System conceptualisation can be described through identification of the simplifying assumptions involved, the primary factors and processes retained, and the theoretical or empirical justification for these choices. As such, the description of system conceptualisation provides a parallel to the description of model design. In further steps, we suggest explicit identification of the ways in which model and system conceptualisations do and do not align with one another, and finally a similar identification of alignments between design choices and particular aspects of the problem being addressed.

As described elsewhere, conceptual model evaluation mainly deals with uncertainty, and Augusiak et al. (2014) suggest the use of Occam's razor as a guiding principle to ensure parsimony. The description of van Vliet et al. (2016) is less prescriptive: "The goal of conceptual modeling is to make the modeler's implicit way of thinking about the system explicit, and thus open to testing, criticism, refinement, and improvement"; an approach that also lends itself to assessment of whether the model is conceptually appropriate for its intended purpose (Brown et al., 2013).

It is important that conceptual model evaluation is not seen as a test that can be passed through appeal to existing authority. Many land use models do not have explicit theoretical groundings (Brown et al., 2017; Groeneveld et al., 2017; Huber et al., 2018), which is a problem if the model's conceptual basis is therefore unclear, but not necessarily otherwise. Theories themselves can be strongly disputed, and models can play a useful role in these disputes. Moreover, representing theories in models can reveal internal contradictions in those theories that preclude any single valid interpretation without reference to some external criteria (Schwarz et al., 2020). It can quite reasonably be argued that any clear conceptual basis is valid if a modeller believes it to be so and wishes to develop a model to explore it further or test outcomes. This may be especially pertinent to

ABMs because they can be used to explore theories in ways that other models cannot – a feature that has been presented as one of their main strengths (Gostoli and Silverman, 2020).

A specific and crucial issue for land use models is the simulation of uncertain futures. In these cases, special attention needs to be paid to the conceptual relevance of the model to the future conditions to which it is being applied. In this sense, the lack of uptake of relevant theories from social science in land use models is a significant problem, as is the near-universal application of the economic optimisation paradigm at large scales, especially in Integrated Assessment Models, without explicit justification of its relevance (Brown et al., 2016; Groeneveld et al., 2017). In both cases, pattern accuracy under present conditions is likely to be favoured over process accuracy under future conditions, creating substantial scope for mismatches with the research questions being addressed (Brown et al., 2022; Steel, 2007). For example, models running Shared Socio-economic Pathways scenarios must somehow represent (or justify a failure to represent) a wide variety of different behavioural, social and cultural factors that are known to be crucial to land system outcomes (Brown et al., 2022; Pedde et al., 2019). These factors differ greatly in future scenarios, but models usually simulate all future conditions using single, fixed model architectures. In this respect, explicit consideration of the fit between model and problem conceptualisation is the most important but most neglected aspect of this evaluation step.

3.5 Implementation verification

"(1) Whether the computer code implementing the model has been thoroughly tested for programming errors, (2) whether the implemented model performs as indicated by the model description, and (3) how the software has been designed and documented to provide necessary usability tools (interfaces, automation of experiments, etc.) and to facilitate future installation, modification, and maintenance." (Grimm et al., 2014)

This step is the most objective and closest to a binary test of model validity. In particular, comprehensive testing of individual sections of code to ensure they (only) function as intended is good programming practice, and relatively straightforward to do through 'Unit Tests'. This is common practice in industry but rarely reported for academic research, where there are increased calls for it to be adopted alongside the rapid growth of pandemic disease modelling post-Covid (Lucas et al., 2020). Nevertheless, systematic tests of model code and performance have been carried out for many land system models, such as CRAFTY (Murray-Rust et al., 2014), the Community Land Model (Hoffman et al., 2005) and the APSIM farm system model (Holzworth et al., 2018), all of which have multiple applications. Because of the importance of this step for model re-use, we identify it as an essential part of evaludation.

3.6 Model output verification

"(1) How well model output matches observations. (2) How much calibration and effects of environmental drivers were involved in obtaining good fits of model output and data." (Grimm et al., 2014)

Model output verification is subject to similar difficulties as input data evaluation (step 3), in that data may play a limited role in model development (and this step explicitly refers to data used in model development, rather than the independent data focused on in output corroboration, section 3.8). As such, this step is applicable to varying degrees within land system modelling. Nevertheless, an important contribution can be made by better descriptions of ABM calibration. New methods can support both the practice and communication of calibration (McCulloch et al., 2022), but at a basic level transparent explanation of how calibration was achieved is beneficial, and in particular how empirical data were used to assess alternative calibration specifications. This would enable clearer understanding of the level of process understanding in different components of the model, which is important given that models are developed and used when understanding is incomplete. Communicating where there is more or less process understanding within the model structure, and therefore less or more valibration was needed, will enable others to assess where greatest uncertainty is in model outputs, and where the focus should be to improve understanding in the future. Similarly, explicit quantitative calibration methods such as those applied by McCulloch et al. (2022) to three different ABMs provide efficient calibration while also revealing the greatest sources of uncertainty.

3.7 Model analysis and application

"(1) How sensitive model output is to changes in model parameters. (2) How well the emergence of model output has been understood." (Grimm et al., 2014)

Sensitivity and uncertainty analyses are some of the most informative checks that can be carried out on land system models, revealing not only model behaviour but also, potentially, aspects of system behaviour itself. They add greatly to model interpretability as well as user confidence. Sensitivity analyses in particular are commonly performed, although generally using partial approaches such as one-at-a-time parameter variations that are known to miss large areas of parameter space (Saltelli et al., 2019). Agent-based models often have some stochastic elements, and in these cases, assessment of the effects of stochasticity on model outputs is also necessary. This can be done using established Monte Carlo sampling approaches, assessments of variance stability and even formalised comparisons to equivalent deterministic models, if available (e.g. Lee et al., 2015; Mohd, 2022).

Nevertheless, uncertainty analyses on ABMs are rare, and improvements in methods and practices are likely to be necessary (McCulloch et al., 2022). For example, Lee et al. (2015) suggest methods to ensure that sample size is appropriate and effects meaningful. Ligmann-Zielinska et al. (2020) provide a roadmap for sensitivity analysis of ABMs that accounts for model design and several common shortcomings to make robust analyses more accessible. Railsback & Grimm (2019) include guidance on analysing uncertainty and its implication for reliability of results in agent and individual-based models, while An et al. (2021) discuss the handling of uncertainty associated with parameters and other components of complex ABMs.

Grimm & Berger (2016) also develop the concept of Robustness Analysis, which forms part of the TRACE protocol. This analysis involves purposefully 'breaking' the model by varying parameters, structure and representation to the point at which the model no longer reproduces an observation, revealing when and why mechanisms included in the model lose their relevance.

3.8 Model output corroboration

"How model predictions compare to independent data and patterns that were not used, and preferably not even known, while the model was developed, parameterized, and verified. By documenting model output corroboration, model users learn about evidence which, in addition to model output verification, indicates that the model is structurally realistic so that its predictions can be trusted to some degree." (Grimm et al., 2014)

The step of model output corroboration may appear to overlap to some extent with model output verification (section 3.6), but involves the model being confronted with independent data rather than calibrated to match data during model development. This step often stands in for the entire process of model evaluation, being regarded as both necessary and sufficient by many model users (Augusiak et al., 2014; Hunka et al., 2013). It can certainly be a useful, informative and somewhat objective step, but is not entirely reliable or relevant in land system studies. This is primarily because different parts of the land system (separated by time and/or space) would not normally be expected to replicate one another's behaviour in the way that different parts of a physical or even biological system might, for the reason that they are realisations of complex social-ecological processes that themselves are known to vary across space and time (Brown et al., 2016; Malek & Verburg, 2020). Observational data can only show one outcome of the underlying processes that an ABM might seek to represent, and this outcome could effectively be an extreme outlier of the theoretical probability distribution. As a result, a model developed using any particular set of empirical or theoretical information would only be expected to match independent sets of information in special cases. Furthermore, 'observational' land system data themselves invariably incorporate implicit assumptions and, effectively, modelling steps; for instance in converting remotely-sensed wavelengths of light into discrete land cover classes. These assumptions and models may not match those adopted by the land system model being evaluated, and mean that independent data cannot be taken as ground truth (Verburg et al., 2011).

Another serious barrier to corroboration is that land system models are often applied to future contexts, making process accuracy and consistency with scenario conditions (as discussed above) far more relevant than fit to historical data. In these cases, binary determinations of model validity are impossible because the future extends beyond the 'validity range' of any existing observations. Instead, the validity of the model for future simulations can be qualitatively tested against the narratives describing those futures, leading to structural changes in the model architecture (Brown et al., 2022).

A type of model corroboration that can be informative here is comparison to spatio-temporal patterns, at aggregate resolutions across different contexts. In being explicitly focused on exploratory comparison rather than arbitrary goodness-of-fit testing, these may provide more complete and more useful information about model performance and relevance to a distinct problem domain. They also allow for multi-criteria analyses and optimisation (Newland et al., 2018) that, while still vulnerable to spurious fits, spread the risk of accidental agreement across multiple tests and/or types of pattern (e.g. not only the location of urban clusters but their existence and size; Rand et al., 2003). This also reduces the scope for cherry-picking single measures in which the model appears to perform well. Multi-criteria analysis can also potentially allow comparison of areas in which process and predictive accuracy can be disentangled (Brown et al., 2005). Methods capable of exploring large areas of parameter space efficiently are therefore very relevant, especially for ABMs with many parameters. A number of robust statistical approaches have been suggested to improve on current practice (Saltelli et al., 2019; Stepanyan et al., 2021). Particularly promising are techniques using machine learning, for instance to create surrogates of ABMs that can be analysed far more quickly than the model itself (Angione et al., 2022; ten Broeke et al., 2021).

The ability to pattern match does not necessarily make it a good idea, however. ABMs are often intended explicitly for exploration of individual and social behaviours, and not for pattern-matching. Several authors have argued that prediction is a minor goal in land system science, and one that is unduly restrictive in prioritising the replication of single outcomes (Edmonds, 2017; Epstein, 2008; Williams et al., 2020). While useful debates about the ideal role for prediction continue (e.g. Polhill, 2018), it is certainly true that many if not most land system models have purposes that cannot be assessed through tests of predictive accuracy alone.

Because data-based corroboration of model outputs is so difficult, the associated practice of model benchmarking can be particularly useful, and we include this as a distinct component of our LUC-TRACE protocol. Model benchmarking involves the comparison of a model's outputs to those of alternative models, along with comparison of model inputs and structure to allow meaningful conclusions to be drawn. Benchmarking projects already exist for models of land and other systems (e.g. ILAMB; Collier et al., 2018), but benchmarking is not common practice for small-scale or behavioural models. Formalised comparisons can allow both individual model behaviour and comparative differences to be attributed to particular features, or even to aspects of the system.

3.9 Participatory/companion modelling

Records of participatory or companion modelling are new additions of ours to the TRACE protocol, although Schulze et al. (2017) discuss the role of participatory modelling in the evaludation framework. We suggest individual components relating to participatory model development or selection (details of whether and how the model was developed or chosen with user engagement), usage in participatory settings (records of such usage including its design and outcomes) and communication of results (how results were presented and how feedback at this point was captured and used). As argued by, for example, Hamilton et al. (2019), the prominence of environmental models in political and societal decision processes, and the wide range of purposes they are used for in these contexts, means that model selection and evaluation must actively involve a range of stakeholders. Land system models and ABMs in particular have additional interest in relating directly to human systems, and so stakeholders are likely to have important perspectives on model design and use.

There is strong evidence to suggest that participatory modelling can improve models themselves and stakeholder understanding of their outputs (Burton et al., 2018; Hamilton et al., 2019; Holman et al., 2016; Millington et al., 2011). Despite the value of participatory approaches, they are rarely used at any stage of model design (Millington & Wainwright, 2017; Sohl & Claggett, 2013). Model complicatedness and complexity can be barriers, as can basic characteristics like the treatment of geographical space (e.g. in terms of discrete units) that are strong abstractions of the world as perceived by many stakeholders (Barnaud et al., 2013; Sun et al., 2016). Although ABMs may be more suitable for participatory modelling due to their representation of potentially relatable behaviours, their use in this way remains under-developed (O'Sullivan et al., 2016). Potential ways forward are provided by strategies for better and more comprehensive engagement that address model design, selection and evaluation in general (Hamilton et al., 2022).

3.10 Model replication

Replication is an essential check in scientific progress in general, but one that requires a formidable amount of work in modelling. Replication can take the form of independent use of the model to check reproducibility of existing results, or, at its extreme, independent re-development of the model itself. Because both of these are so challenging, time-demanding and generally unrewarded, they are almost never done. As a result, suspicions that many published results may be erroneous in some way are widespread, and are borne out by evidence in some cases (Polhill et al., 2005; Zhang & Robinson, 2021).

In the context of the TRACE protocol, a further issue for replication is that the protocol is generally intended to describe new models that are unlikely to have been available long enough for replication to occur. Nevertheless, efforts to facilitate replication are crucial and can be made at this stage. Like evaluation in general, replication can be hampered by poor model descriptions and lack of open-access data and code, which other parts of this evaludation protocol address. We also propose that models be situated within the taxonomy of reproducibility of Essawy et al. (2020), who define four key steps: 1) Repeatability, achieved when a model produces consistent results when run in its original form; 2) Runnability, achieved when consistent results are produced on a different computer; 3) Reproducibility, achieved when an independent researcher produces consistent results in their own computational environment; and 4) Replicability, achieved when independent studies produce consistent results in their own ways. We propose that modellers detail any steps taken to facilitate or implement the above steps during model development, as well as detailing the full computational environment in which published results were generated. Examples also exist of modellers actively testing different computing environments and recording performance, as Hoffman et al. (2005) did for the Community Land Model, and as Essawy et al. (2020) did while exploring methods for simplifying the process of maximising replicability.

4. Discussion and Conclusion

The LUC-TRACE 'evaludation' framework we propose builds on the existing TRACE protocol (Ayllón et al., 2021; Grimm et al., 2014; Schmolke et al., 2010) in recognising that there are many ways of evaluating land system models, all of which have some value, some of which are essential, but none of which are sufficient on their own. Combining several methods of evaluation therefore provides a particularly robust approach, and the tiered checklist we suggest in Table 1 is intended to highlight the most important methods to use in a land system modelling context. It is also intended to streamline the process of evaludation and illustrate the value even of a brief and partial document that covers key points.

It is nonetheless necessary to acknowledge practical problems associated with implementing these steps. Land system models are often complicated and complex, and so methods for their evaluation are not simple. It may be difficult for readers and users to understand what has been done even in the best of circumstances, and this places an explanatory burden on model developers that is not usually supported by academic or funding systems. Recent moves towards transparency and open science are important steps, but even with greater support, creating (for the model developer) and reading (for the model user) large evaluation documents may simply be too time-consuming and tedious. It is also important to recognise that the ultimate documentation remains the model code itself, which describes exactly what the model does in a standardised form. Structured frameworks like TRACE can be invaluable in providing a reference manual of model evaludation, and possibly also an interpretive bridge between narrative descriptions and underlying code, but are not intended to be universally applicable or sufficient.

There are further challenges involved in balancing rigorous evaluation against a preference among (applied) funding bodies and many model users for apparently certain results that answer specific questions, even when models are unable to do so reliably. The consistent use of established models to address questions to which they might not be suited, rather than to invest in the development of alternative models, has been cited as evidence of this preference (Gambhir et al., 2019; Low & Schäfer, 2020). There are also inevitable difficulties in evaluating any models that deal with qualitative, complex and contentious issues. Registered reports, of the kind suggested for empirical science, could be valuable for modelling studies, ensuring that full descriptions of models were available and removing the need for particular results to justify publication (Center for Open Science, 2022). The development of models from general empirical, conceptual or theoretical foundations can also ensure a basic level of acceptability. Methods could include the use of 'Digital Twins' that pair models with their real-world counterparts through ongoing data exchange, allowing iterative calibration and evaluation to occur (Lenfers et al., 2021), agreement on particularly important ‘high-level’ processes to be included in models of particular systems (Brown & Rounsevell, 2021; Urban et al., 2021), or simply fuller use of existing theories and qualitative information (Groeneveld et al., 2017).

Ultimately of course, modelling has an important role to play as an imaginative practice to be followed where it might lead. In this case, model utility is dependent on communication and interpretability, and model evaluation is a process of understanding model behaviour – and therefore, hopefully, system behaviour itself. Users are likely to be faced with a choice among models that are neither valid nor invalid in a strict sense, with utilities (or ‘fitnesses’ as Hamilton et al. (2022) term them) that cannot be comprehensively predetermined. Instead, models have as much or as little validity as users give them when they apply and interpret them, and evaludation is useful to the extent that it supports these processes.

References

An, L., Grimm, V., Sullivan, A., Turner, B. L., II, Malleson, N., Heppenstall, A., Vincenot, C., Robinson, D., Ye, X., Liu, J., Lindkvist, E., & Tang, W. (2021). Challenges, tasks, and opportunities in modeling agent-based complex systems. Ecological Modelling, 457, 109685. https://doi.org/10.1016/j.ecolmodel.2021.109685

Angione, C., Silverman, E., & Yaneske, E. (2022). Using machine learning as a surrogate model for agent-based simulations. PloS One, 17(2), e0263150. https://doi.org/10.1371/journal.pone.0263150

Arneth, A., Brown, C., & Rounsevell, M. D. A. (2014). Global models of human decision-making for land-based mitigation and adaptation assessment. Nature Climate Change, 4(7), 550–557. https://doi.org/10.1038/nclimate2250

Arsenault, R., Brissette, F., & Martel, J.-L. (2018). The hazards of split-sample validation in hydrological model calibration. Journal of Hydrology, 566, 346–362. https://doi.org/10.1016/j.jhydrol.2018.09.027

Augusiak, J., Van den Brink, P. J., & 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. https://doi.org/10.1016/j.ecolmodel.2013.11.009

Ayllón, D., Railsback, S. F., Gallagher, C., Augusiak, J., Baveco, H., Berger, U., Charles, S., Martin, R., Focks, A., Galic, N., Liu, C., van Loon, E. E., Nabe-Nielsen, J., Piou, C., Polhill, J. G., Preuss, T. G., Radchuk, V., Schmolke, A., Stadnicka-Michalak, J., … Grimm, V. (2021). Keeping modelling notebooks with TRACE: Good for you and good for environmental research and management support. Environmental Modelling & Software, 136, 104932. https://doi.org/10.1016/j.envsoft.2020.104932

Baldos, C., & Hertel, T. W. (2013). Looking back to move forward on model validation: insights from a global model of agricultural land use Related content Climate adaptation as mitigation: the case of agricultural investments. Environmental Research Letters: ERL [Web Site], 8, 34024. https://doi.org/10.1088/1748-9326/8/3/034024

Barlas, Y. (1996). Formal aspects of model validity and validation in system dynamics. System Dynamics Review, 12(3), 183–210. https://doi.org/10.1002/(sici)1099-1727(199623)12:3<183::aid-sdr103>3.0.co;2-4

Barnaud, C., Le Page, C., Dumrongrojwatthana, P., & Trébuil, G. (2013). Spatial representations are not neutral: Lessons from a participatory agent-based modelling process in a land-use conflict. Environmental Modelling & Software, 45, 150–159. https://doi.org/10.1016/j.envsoft.2011.11.016

Bianchi, F., & Squazzoni, F. (2019). Modelling and social science. In Modelling Transitions (pp. 60–74). https://doi.org/10.4324/9780429056574-5

Boone, R. B., & Galvin, K. A. (2014). Simulation as an approach to social-ecological integration, with an emphasis on agent-based modeling. In Understanding Society and Natural Resources: Forging New Strands of Integration Across the Social Sciences (pp. 179–202). Springer Netherlands. https://doi.org/10.1007/978-94-017-8959-2_9

Brown, C., Alexander, P., Holzhauer, S., & Rounsevell, M. D. A. (2017). Behavioral models of climate change adaptation and mitigation in land-based sectors. Wiley Interdisciplinary Reviews: Climate Change, 8 (2), e448. https://doi.org/10.1002/wcc.448

Brown, C., Brown, K., & Rounsevell, M. (2016). A philosophical case for process-based modelling of land use change. Modeling Earth Systems and Environment, 2(2), 50. https://doi.org/10.1007/s40808-016-0102-1

Brown, C., Holman, I., & Rounsevell, M. (2021). How modelling paradigms affect simulated future land use change. Earth System Dynamics, 12, 211–231. https://doi.org/10.5194/esd-12-211-2021

Brown, C., & Rounsevell, M. (2021). How can social–ecological system models simulate the emergence of social–ecological crises? People and Nature, 3(1), 88–103. https://doi.org/10.1002/pan3.10167

Brown, C., Seo, B., Alexander, P., Burton, V., Chacón-Montalván, E. A., Dunford, R., Merkle, M., Harrison, P. A., Prestele, R., Robinson, E. L., & Rounsevell, M. (2022). Agent-based modeling of alternative futures in the British land use system. Earth’s Future, 10(11), e2022EF002905. https://doi.org/10.1029/2022ef002905

Brown, C., Seo, B., & Rounsevell, M. (2019). Societal breakdown as an emergent property of large-scale behavioural models of land use change. Earth System Dynamics Discussions, May, 1–49. https://doi.org/10.5194/esd-2019-24

Brown, D. G., Page, S., Riolo, R., Zellner, M., & Rand, W. (2005). Path dependence and the validation of agent-based spatial models of land use. International Journal of Geographical Information Science: IJGIS, 19(2), 153–174. https://doi.org/10.1080/13658810410001713399

Brown, D. G., Verburg, P. H., Pontius, R. G., & Lange, M. D. (2013). Opportunities to improve impact, integration, and evaluation of land change models. Current Opinion in Environmental Sustainability 5 (5), 452–457. https://doi.org/10.1016/j.cosust.2013.07.012

Burton, V., Metzger, M. J., Brown, C., & Moseley, D. (2018). Green Gold to Wild Woodlands; understanding stakeholder visions for woodland expansion in Scotland. Landscape Ecology, 34, 1693-1713. https://doi.org/10.1007/s10980-018-0674-4

Center for Open Science. (2022). Registered Reports. https://www.cos.io/initiatives/registered-reports

Clemen, T., Lenfers, U. A., Dybulla, J., Ferreira, S. M., Kiker, G. A., Martens, C., & Scheiter, S. (2021). A cross-scale modeling framework for decision support on elephant management in Kruger National Park, South Africa. Ecological Informatics, 62, 101266. https://doi.org/10.1016/j.ecoinf.2021.101266

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., Mu, M., & Randerson, J. T. (2018). The international land model benchmarking (ILAMB) system: Design, theory, and implementation. Journal of Advances in Modeling Earth Systems, 10(11), 2731–2754. https://doi.org/10.1029/2018ms001354

Edmonds, B. (2017). Different Modelling Purposes. In B. Edmonds & R. Meyer (Eds.), Simulating Social Complexity: A Handbook (pp. 39–58). Springer International Publishing. https://doi.org/10.1007/978-3-319-66948-9_4

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(3), 6. https://doi.org/10.18564/jasss.3993

Edmonds, B., & ní Aodha, L. (2019). Using Agent-Based Modelling to Inform Policy – What Could Possibly Go Wrong? Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11463 LNAI, 1–16. https://doi.org/10.1007/978-3-030-22270-3_1

Epstein, J. M. (1999). Agent-based computational models and generative social science. Complexity, 4(5), 41–60. https://onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1099-0526(199905/06)4:5%3C41::AID-CPLX9%3E3.0.CO;2-F

Epstein, J. M. (2008). Why Model? Journal of Artificial Societies and Social Simulation, 11(4)12.

Essawy, B. T., Goodall, J. L., Voce, D., Morsy, M. M., Sadler, J. M., Choi, Y. D., Tarboton, D. G., & Malik, T. (2020). A taxonomy for reproducible and replicable research in environmental modelling. Environmental Modelling & Software, 134, 104753. https://doi.org/10.1016/j.envsoft.2020.104753

Feyerabend, P. (1993). Against Method. Verso. https://play.google.com/store/books/details?id=8y-FVtrKeSYC

Gambhir, A., Butnar, I., Li, P. H., Smith, P., & Strachan, N. (2019). A review of criticisms of integrated assessment models and proposed approaches to address these, through the lens of BECCs. Energies, 12(9), 1–21. https://doi.org/10.3390/en12091747

Gary Polhill, J., Parker, D., Brown, D., & Grimm, V. (2008). Using the ODD protocol for describing three agent-based social simulation models of land-use change. JASSS. https://www.jasss.org/11/2/3.html Journal of Artificial Societies and Social Simulation, 11(2)3.

Giupponi, C., Ausseil, A.-G., Balbi, S., Cian, F., Fekete, A., Gain, A. K., Essenfelder, A. H., Martínez-López, J., Mojtahed, V., Norf, C., & Others. (2022). Integrated modelling of social-ecological systems for climate change adaptation. Socio-Environmental Systems Modelling, 3, 18161–18161. https://doi.org/10.18174/sesmo.18161

Gostoli, U., & Silverman, E. (2020). Sound behavioural theories, not data, is what makes computational models useful. Review of Artificial Societies and Social Simulation. https://rofasss.org/2020/04/22/sound-behavioural-theories/

Grimm, V., Augusiak, J., Focks, A., Frank, B. M., Gabsi, F., Johnston, A. S. A., Liu, C., Martin, B. T., Meli, M., Radchuk, V., Thorbek, P., & Railsback, S. F. (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. (2016). Robustness analysis: Deconstructing computational models for ecological theory and applications. Ecological Modelling, 326, 162–167. https://doi.org/10.1016/j.ecolmodel.2015.07.018

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: A review and first update. Ecological Modelling, 221(23), 2760–2768. https://doi.org/10.1016/J.ECOLMODEL.2010.08.019

Grimm, V., & Railsback, S. F. (2012). Pattern-oriented modelling: a 'multi-scope' for predictive systems ecology. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 367(1586), 298–310. https://doi.org/10.1098/rstb.2011.0180

Grimm, V., Railsback, S. F., Vincenot, C. E., Berger, U., Gallagher, C., DeAngelis, D. L., Edmonds, B., Ge, J., Giske, J., Groeneveld, J., Johnston, A. S. A., Milles, A., Nabe-Nielsen, J., Polhill, J. G., Radchuk, V., Rohwäder, M.-S., Stillman, R. A., Thiele, J. C., & Ayllón, D. (2020). 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(2), 7. https://doi.org/10.18564/jasss.4259

Groeneveld, J., Müller, B., Buchmann, C. M., Dressler, G., Guo, C., Hase, N., Hoffmann, F., John, F., Klassert, C., Lauf, T., Liebelt, V., Nolzen, H., Pannicke, N., Schulze, J., Weise, H., & Schwarz, N. (2017). Theoretical foundations of human decision-making in agent-based land use models – A review. Environmental Modelling & Software, 87, 39–48. https://doi.org/10.1016/J.ENVSOFT.2016.10.008

Guzy, M., Smith, C., Bolte, J., Hulse, D., & Gregory, S. (2008). Policy Research Using Agent-Based Modeling to Assess Future Impacts of Urban Expansion into Farmlands and Forests. Ecology and Society, 13(1) 37. https://doi.org/10.5751/ES-02388-130137

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

Hamilton, S. H., Pollino, C. A., Stratford, D. S., Fu, B., & Jakeman, A. J. (2022). Fit-for-purpose environmental modeling: Targeting the intersection of usability, reliability and feasibility. Environmental Modelling & Software, 148, 105278. https://doi.org/10.1016/j.envsoft.2021.105278

Hoffman, F. M., Vertenstein, M., Kitabata, H., & White, J. B. (2005). Vectorizing the Community Land Model. The International Journal of High Performance Computing Applications, 19(3), 247–260. https://doi.org/10.1177/1094342005056113

Holman, I. P., Harrison, P. A., & Metzger, M. J. (2016). Cross-sectoral impacts of climate and socio-economic change in Scotland: implications for adaptation policy. Regional Environmental Change, 16(1), 97–109. https://doi.org/10.1007/s10113-014-0679-8

Holzworth, D., Huth, N. I., Fainges, J., Brown, H., Zurcher, E., Cichota, R., Verrall, S., Herrmann, N. I., Zheng, B., & Snow, V. (2018). APSIM Next Generation: Overcoming challenges in modernising a farming systems model. Environmental Modelling & Software, 103, 43–51. https://doi.org/10.1016/j.envsoft.2018.02.002

Huber, R., Bakker, M., Balmann, A., Berger, T., Bithell, M., Brown, C., Grêt-Regamey, A., Xiong, H., Le, Q. B., Mack, G., Meyfroidt, P., Millington, J., Müller, B., Polhill, J. G., Sun, Z., Seidl, R., Troost, C., & Finger, R. (2018). Representation of decision-making in European agricultural agent-based models. Agricultural Systems, 167, 143–160. https://doi.org/10.1016/j.agsy.2018.09.007

Hughes, H. R., & Paterson, M. (2017). Narrowing the climate field: The symbolic power of authors in the IPCC's assessment of mitigation. The Review of Policy Research, 34(6), 744–766. https://doi.org/10.1111/ropr.12255

Hüning, C., Adebahr, M., Thiel-Clemen, T., Dalski, J., Lenfers, U., & Grundmann, L. (2016). Modeling & simulation as a service with the massive multi-agent system MARS. Proceedings of the Agent-Directed Simulation Symposium, 1, 1–8. https://dl.acm.org/doi/10.5555/2972193.2972194

Hunka, A. D., Meli, M., Thit, A., Palmqvist, A., Thorbek, P., & Forbes, V. E. (2013). Stakeholders' Perspective on Ecological Modeling in Environmental Risk Assessment of Pesticides: Challenges and Opportunities. Risk Analysis: An Official Publication of the Society for Risk Analysis, 33(1), 68–79. https://doi.org/10.1111/j.1539-6924.2012.01835.x

Lee, J. S., Filatova, T., Ligmann-Zielinska, A., Hassani-Mahmooei, B., Stonedahl, F., Lorscheid, I., Voinov, A., Polhill, G., Sun, Z., & Parker, D. C. (2015). The complexities of agent-based modeling output analysis. Journal of Artificial Societies and Social Simulation, 18(4)4. https://doi.org/10.18564/jasss.2897

Lenfers, U. A., Ahmady-Moghaddam, N., Glake, D., Ocker, F., Osterholz, D., Ströbele, J., & Clemen, T. (2021). Improving Model Predictions—Integration of Real-Time Sensor Data into a Running Simulation of an Agent-Based Model. Sustainability: Science Practice and Policy, 13(13), 7000. https://doi.org/10.3390/su13137000

Ligmann-Zielinska, A., Siebers, P.-O., Magliocca, N., Parker, D. C., Grimm, V., Du, J., Cenek, M., Radchuk, V., Arbab, N. N., Li, S., Berger, U., Paudel, R., Robinson, D. T., Jankowski, P., An, L., & Ye, X. (2020). 'one size does not fit all': A roadmap of purpose-driven mixed-method pathways for sensitivity analysis of agent-based models. Journal of Artificial Societies and Social Simulation, 23(1), 6. https://doi.org/10.18564/jasss.4201

Lorscheid, I., Berger, U., Grimm, V., & Meyer, M. (2019). From cases to general principles: A call for theory development through agent-based modeling. Ecological Modelling, 393, 153–156. https://doi.org/10.1016/j.ecolmodel.2018.10.006

Low, S., & Schäfer, S. (2020). Is bio-energy carbon capture and storage (BECCS) feasible? The contested authority of integrated assessment modeling. Energy Research and Social Science, 60, 101326. https://doi.org/10.1016/j.erss.2019.101326

Magliocca, N. R., Brown, D. G., & Ellis, E. C. (2014). Cross-site comparison of land-use decision-making and its consequences across land systems with a generalized agent-based model. PloS One, 9(1), e86179. https://doi.org/10.1371/journal.pone.0086179

Malek, Ž., & Verburg, P. H. (2020). Mapping global patterns of land use decision-making. Global Environmental Change: Human and Policy Dimensions, 65, 102170. https://doi.org/10.1016/j.gloenvcha.2020.102170

McCulloch, J., Ge, J., Ward, J. A., Heppenstall, A., Polhill, J. G., & Malleson, N. (2022). Calibrating agent-based models using uncertainty quantification methods. Journal of Artificial Societies and Social Simulation, 25 (2), 1. https://doi.org/10.18564/jasss.4791

Millington, J. D. A., Demeritt, D., & Romero-Calcerrada, R. (2011). Participatory evaluation of agent-based land-use models. Journal of Land Use Science, 6(2–3), 195–210. https://doi.org/10.1080/1747423X.2011.558595

Millington, J. D. A., Katerinchuk, V., Bicudo da Silva, R. F., de Castro Victoria, D., & Batistella, M. (2021). Modelling drivers of Brazilian agricultural change in a telecoupled world. Environmental Modelling & Software, 105024. https://doi.org/10.1016/j.envsoft.2021.105024

Millington, J. D. A., & Wainwright, J. (2017). Mixed qualitative-simulation methods: Understanding geography through thick and thin. Progress in Human Geography, 41(1), 68–88. https://doi.org/10.1177/0309132515627021

Mohd, M. H. (2022). Revisiting discrepancies between stochastic agent-based and deterministic models. Community Ecology, 23(3), 453–468. https://doi.org/10.1007/s42974-022-00118-2

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 & Software, 48, 37–48. https://doi.org/10.1016/j.envsoft.2013.06.003

Murray-Rust, D., Brown, C., van Vliet, J., Alam, S. J., Robinson, D. T., Verburg, P. H., & Rounsevell, M. (2014). Combining agent functional types, capitals and services to model land use dynamics. Environmental Modelling & Software, 59, 187–201. https://doi.org/10.1016/j.envsoft.2014.05.019

Naivinit, W., Le Page, C., Trébuil, G., & Gajaseni, N. (2010). Participatory agent-based modeling and simulation of rice production and labor migrations in Northeast Thailand. Environmental Modelling & Software, 25(11), 1345–1358. https://doi.org/10.1016/j.envsoft.2010.01.012

Newland, C. P., Maier, H. R., Zecchin, A. C., Newman, J. P., & van Delden, H. (2018). Multi-objective optimisation framework for calibration of Cellular Automata land-use models. Environmental Modelling and Software, 100, 175–200. https://doi.org/10.1016/j.envsoft.2017.11.012

Ngo, The An, & See, L. (2012). Calibration and Validation of Agent-Based Models of Land Cover Change. In A. J. Heppenstall, A. T. Crooks, L. M. See, & M. Batty (Eds.), Agent-Based Models of Geographical Systems (pp. 181–197). Springer Netherlands. https://doi.org/10.1007/978-90-481-8927-4_10

Niamir, L., Ivanova, O., & Filatova, T. (2020). Economy-wide impacts of behavioral climate change mitigation: Linking agent-based and computable general equilibrium models. Environmental Modelling & Software, 134, 104839. https://doi.org/10.1016/j.envsoft.2020.104839

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the Earth sciences. Science, 263(5147), 641–646. https://doi.org/10.1126/science.263.5147.641

O'Sullivan, D., Evans, T., Manson, S., Metcalf, S., Ligmann-Zielinska, A., & Bone, C. (2016). Strategic directions for agent-based modeling: avoiding the YAAWN syndrome. Journal of Land Use Science, 11(2), 177–187. https://doi.org/10.1080/1747423X.2015.1030463

Pedde, S., Kok, K., Hölscher, K., Frantzeskaki, N., Holman, I., Dunford, R., Smith, A., & Jäger, J. (2019). Advancing the use of scenarios to understand society's capacity to achieve the 1.5 degree target. Global Environmental Change: Human and Policy Dimensions, 56, 75–85. https://doi.org/10.1016/J.GLOENVCHA.2019.03.010

Polhill, G. (2018). Why the social simulation community should tackle prediction. Review of Artificial Societies and Social Simulation. https://rofasss.org/2018/08/17/gp/

Polhill, G., Izquierdo, L. R., & Gotts, N. M. (2005). The Ghost in the Model (and Other Effects of Floating Point Arithmetic). Journal of Artificial Societies and Social Simulation, 8(1), 1–24. http://jasss.soc.surrey.ac.uk/8/1/5.html

Polhill, Gary, & Salt, D. (2017). The Importance of ontological structure: Why validation by 'fit-to-data' is insufficient. In Understanding Complex Systems (pp. 141–172). Springer Verlag. https://doi.org/10.1007/978-3-319-66948-9_8

Railsback, S. F., & Grimm, V. (2019). Agent-Based and Individual-Based Modeling: A Practical Introduction. 2nd Edition. Princeton University Press. https://play.google.com/store/books/details?id=VQOjinaRG9cC

Rand, W., Brown, D. G., Page, S. E., Riolo, R., Fernandez, L. E., & Zellner, M. (2003). Statistical validation of spatial patterns in agent-based models. Proceedings of Agent Based Simulation 2003, Vol. 4. https://pdfs.semanticscholar.org/c2da/2de18efa2300bf3cdfdf86b51963480d1804.pdf

Roe, S., Streck, C., Obersteiner, M., Frank, S., Griscom, B., Drouet, L., Fricko, O., Gusti, M., Harris, N., Hasegawa, T., Hausfather, Z., Havlík, P., House, J., Nabuurs, G. J., Popp, A., Sánchez, M. J. S., Sanderman, J., Smith, P., Stehfest, E., & Lawrence, D. (2019). Contribution of the land sector to a 1.5 °C world. Nature Climate Change, 9(11), 817–828. https://doi.org/10.1038/s41558-019-0591-9

Rogelj, J., Shindell, D., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., Handa, C., Kheshgi, H., Kobayashi, S., Kriegler, E., Mundaca, L., Séférian, R., & Vilariño, M. V. (2018). Mitigation Pathways Compatible with 1.5°C in the Context of Sustainable Development. In V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, P. R. Skea, J. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, E. Lonnoy, T. Maycock, V. Tignor, & T. Waterfield (Eds.), Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, (p. 82). IPCC Secretariat. https://www.ipcc.ch/site/assets/uploads/sites/2/2019/02/SR15_Chapter2_Low_Res.pdf

Rosa, I. M. D., Ahmed, S. E., & Ewers, R. M. (2014). The transparency, reliability and utility of tropical rainforest land-use and land-cover change models. Global Change Biology, 20(6), 1707–1722. https://doi.org/10.1111/gcb.12523

Rounsevell, M., Arneth, A., Brown, C., Cheung, W. W. L., Gimenez, O., Holman, I., Leadley, P., Luján, C., Mahevas, S., Maréchaux, I., Pélissier, R., Verburg, P. H., Vieilledent, G., Wintle, B. A., & Shin, Y.-J. (2021). Identifying uncertainties in scenarios and models of socio-ecological systems in support of decision-making. One Earth, 4(7), 967–985. https://doi.org/10.1016/j.oneear.2021.06.003

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., & Wu, Q. (2019). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. Environmental Modelling & Software, 114, 29–39. https://doi.org/10.1016/j.envsoft.2019.01.012

Schmolke, A., Thorbek, P., DeAngelis, D. L., & Grimm, V. (2010). Ecological models supporting environmental decision making: a strategy for the future. Trends in Ecology & Evolution, 25(8), 479–486. https://doi.org/10.1016/J.TREE.2010.05.001

Schulze, J., Müller, B., Groeneveld, J., & Grimm, V. (2017). Agent-Based Modelling of Social-Ecological Systems: Achievements, Challenges, and a Way Forward. Journal of Artificial Societies and Social Simulation, 20(2), 8. https://doi.org/10.18564/jasss.3423

Schwarz, N., Dressler, G., Frank, K., Jager, W., Janssen, M., Müller, B., Schlüter, M., Wijermans, N., & Groeneveld, J. (2020). Formalising theories of human decision-making for agent-based modelling of social-ecological systems: practical lessons learned and ways forward. Socio-Environmental Systems Modelling, 2, 16340. https://doi.org/10.18174/sesmo.2020a16340

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. Royal Society Open Science, 3(9), 160384. https://doi.org/10.1098/rsos.160384

Sohl, T. L., & Claggett, P. R. (2013). Clarity versus complexity: Land-use modeling as a practical tool fordecision-makers. In Journal of Environmental Management, 129, 235–243. https://doi.org/10.1016/j.jenvman.2013.07.027

Srikrishnan, V., & Keller, K. (2021). Small increases in agent-based model complexity can result in large increases in required calibration data. Environmental Modelling & Software, 138, 104978. https://doi.org/10.1016/j.envsoft.2021.104978

Steel, D. (2007). Across the boundaries: Extrapolation in biology and social science. Oxford, UK: Oxford University Press.

Steinmann, P., Wang, J. R., van Voorn, G. A. K., & Kwakkel, J. H. (2020). Don't try to predict COVID-19. If you must, use Deep Uncertainty methods. Review of Artificial Societies and Social Simulation. https://rofasss.org/2020/04/17/deep-uncertainty/

Stepanyan, D., Grethe, H., Zimmermann, G., Siddig, K., Deppermann, A., Feuerbacher, A., Luckmann, J., Valin, H., Nishizawa, T., Ermolieva, T., & Havlik, P. (2021). Multiple rotations of Gaussian quadratures: An efficient method for uncertainty analyses in large-scale simulation models. Environmental Modelling & Software, 136, 104929. https://doi.org/10.1016/j.envsoft.2020.104929

Storm, H., Baylis, K., & Heckelei, T. (2020). Machine learning in agricultural and applied economics. European Review of Agricultural Economics, 47(3), 849–892. https://doi.org/10.1093/erae/jbz033

Sun, Z., Lorscheid, I., Millington, J. D., Lauf, S., Magliocca, N. R., Groeneveld, J., Balbi, S., Nolzen, H., Müller, B., Schulze, J., & Buchmann, C. M. (2016). Simple or complicated agent-based models? A complicated issue. Environmental Modelling & Software, 86, 56–67. https://doi.org/10.1016/j.envsoft.2016.09.006

Taghikhah, F., Filatova, T., & Voinov, A. (2021). Where does theory have it right? A comparison of theory-driven and empirical agent based models. Journal of Artificial Societies and Social Simulation, 24(2), 4. https://doi.org/10.18564/jasss.4573

ten Broeke, G., van Voorn, G., Ligtenberg, A., & Molenaar, J. (2021). The use of surrogate models to Analyse agent-based models. Journal of Artificial Societies and Social Simulation, 24(2), 3. https://doi.org/10.18564/jasss.4530

Tesfatsion, L. (2017). Modeling economic systems as locally-constructive sequential games. Journal of Economic Methodology, 24(4), 384–409. https://doi.org/10.1080/1350178X.2017.1382068

Troost, C., & Berger, T. (2020). Formalising validation? Towards criteria for valid conclusions from agent-based simulation. In A. van Griensven Jiri Nossent Daniel P. Ames (Ed.), 10th International Congress on Environmental Modelling and Software.

Troost, C., Huber, R., Bell, A. R., van Delden, H., Filatova, T., Le, Q. B., Lippe, M., Niamir, L., Polhill, J. G., Sun, Z., & Berger, T. (2023). How to keep it adequate: A protocol for ensuring validity in agent-based simulation. Environmental Modelling & Software, 159, 105559. https://doi.org/10.1016/j.envsoft.2022.105559

Urban, M. C., Travis, J. M. J., Zurell, D., Thompson, P. L., Synes, N. W., Scarpa, A., Peres-Neto, P. R., Malchow, A.-K., James, P. M. A., Gravel, D., De Meester, L., Brown, C., Bocedi, G., Albert, C. H., Gonzalez, A., & Hendry, A. P. (2021). Coding for Life: Designing a Platform for Projecting and Protecting Global Biodiversity. Bioscience 72, 1. https://doi.org/10.1093/biosci/biab099

van Soesbergen, A. (2016). A Review of Land-Use Change Models. Cambridge, UK: UNEP World Conservation Monitoring Centre

van Vliet, J., Bregt, A. K., Brown, D. G., van Delden, H., Heckbert, S., & Verburg, P. H. (2016). A review of current calibration and validation practices in land-change modeling. Environmental Modelling & Software, 82, 174–182. https://doi.org/10.1016/j.envsoft.2016.04.017

Verburg, P. H., Alexander, P., Evans, T., Magliocca, N. R., Malek, Z., Rounsevell, M. D. A., & van Vliet, J. (2019). Beyond land cover change: towards a new generation of land use models. Current Opinion in Environmental Sustainability, 38, 77–85. https://doi.org/10.1016/j.cosust.2019.05.002

Verburg, P. H., Neumann, K., & Nol, L. (2011). Challenges in using land use and land cover data for global change studies. Global Change Biology, 17(2), 974–989. https://doi.org/10.1111/j.1365-2486.2010.02307.x

Vermeer, W. H., Smith, J. D., Wilensky, U., & Brown, C. H. (2022). High-Fidelity Agent-Based Modeling to Support Prevention Decision-Making: an Open Science Approach. Prevention Science: The Official Journal of the Society for Prevention Research, 23(5), 832–843. https://doi.org/10.1007/s11121-021-01319-3

Williams, T. G., Guikema, S. D., Brown, D. G., & Agrawal, A. (2020). Assessing model equifinality for robust policy analysis in complex socio-environmental systems. Environmental Modelling & Software, 104831. https://doi.org/10.1016/j.envsoft.2020.104831

Zhang, J., & Robinson, D. T. (2021). Replication of an agent-based model using the Replication Standard. Environmental Modelling & Software, 139, 105016. https://doi.org/10.1016/j.envsoft.2021.105016