Наукові записки НаУКМА
Соціологія 8 (2025), 4–15

NaUKMA Research Papers
Sociology 8 (2025), 4–15

СТАТТІ **ARTICLES**

DOI: 10.18523/2617-9067.2025.8.4-15
UDC 303.214.3:519.872

Artem Serdyuk
https://orcid.org/0000-0002-0060-1311

CHOOSING VALIDATION METHODS FOR AGENT-BASED MODELS: A PRACTICAL FRAMEWORK

Abstract

Agent-based modeling (ABM) has emerged as a computational approach for studying complex social phenomena, yet validation practices remain inconsistent. This article develops a systematic framework for ABM validation method selection. Through computational analysis of 76 scenarios across model purposes, observability levels, and development stages, we identify three validation method categories: universal (visualization, expert review, structural checks), context-dependent (varying with data availability), and specialized (requiring specific conditions). We demonstrate how system observability and the development stage fundamentally constrain validation approaches. Application to Adam and Gaudou's bushfire evacuation model illustrates the framework's utility. Our findings show that effective validation of an agent-based model requires systematic contextual assessment, providing researchers with tools for better ABM validation decisions.

Keywords: agent-based modeling, model validation, computational social science, generative social science, social complexity, complex systems, uncertainty, Ukrainian society.

Introduction

Agent-based modeling (ABM) is a computational method for studying complex social phenomena where large-scale patterns emerge from small-scale interactions. ABM emerged from cellular automata research at the Santa Fe Institute in the late 1980s and has been applied across economics, sociology, political science, and public policy. The field developed through four phases: theoretical foundations (1970s–1990s) with models like Schelling's segregation (Schelling, 1971); connecting simulations to real data (1990s–2000s) (Dean et al., 2000; Epstein & Axtell, 1996); methodological standards (2010s) (Grimm et al., 2006); and machine learning integration (2020s) (Epstein, 2023). Currently, agent-based modeling addresses climate change, migration, social polarization, inequality, urban mobility, and disease spread. COVID-19 accelerated ABM use for epidemic modeling with behavioral and economic factors (Lorig et al., 2021; Squazzoni et al., 2020).

Ukrainian ABM research began in the early 2010s with practical applications: language dynamics (Paniotto & Hrushetskyi, 2013; Vengrina, 2012), electoral processes (Pugachova, 2019), and policy analysis (Pugachova, 2021). Recent work addresses war-related challenges like forced migration and organizational adaptation, including universities' responses to the russian invasion (Serdyuk, 2025), showing ABM's value when traditional data collection is impossible, but policy insights are urgent.

ABM faces validation challenges. Its generative nature requires different validation than statistical models (Humphreys, 2004). Different models' purposes require different validation standards, yet no established frameworks exist. ABM validation practices vary considerably across different fields, and resource constraints often force researchers to make trade-offs without systematic frameworks to guide them. Both Ukrainian and international research demonstrate this problem: validation practices remain diverse and unstandardized, reflecting individual choices rather than any coordinated standards.

This article develops a framework for selecting ABM validation methods. We analyzed 76 scenarios to show how validation needs vary by context, identified constraints on method choice, and provided practical guidelines for matching validation approaches to research situations.

Definition of Agent-Based Modeling

Agent-based modeling (ABM) is a computational approach in which autonomous, heterogeneous "agents" – individuals, households, birds, cells, or other entities – follow simple rules and interact within an explicit environment such as physical space, social networks, or institutional settings (Epstein & Axtell, 1996). Because every agent's states and actions are computed step-by-step, researchers can observe how familiar macro-patterns such as segregation, price bubbles, or norm compliance emerge from numerous micro-interactions rather than being imposed from above.

Agent-based models of social processes share key characteristics (Epstein, 2008). They represent heterogeneous agents with unique attributes such as wealth, preferences, or social connections, who act autonomously according to decision rules rather than centralized control. Agents exist in explicit spatial environments where proximity shapes local interactions. Their decisions reflect bounded rationality, relying on limited information and simple heuristics instead of perfect optimization. ABMs focus on dynamic, non-equilibrium processes, enabling the study of evolving systems, tipping points, and large-scale social change.

Despite their advantages, ABMs face well-known implementation challenges:

- **Simplicity-realism trade-offs**. Models must balance interpretability against capturing essential dynamics (Collins et al., 2024). This balance proves challenging, with models often suffering from oversimplification or excessive, unmanageable complexity.
- **Uncertainty in decision rules**. Social behaviors involve heuristics, bounded rationality, and incomplete information, making the parametrization of agent decision-making challenging (Simon, 1996).
- **Computational demands**. As ABMs scale, they require substantial computational resources and generate vast datasets requiring sophisticated analysis techniques (Lee et al., 2015).
- **Validation difficulties**. ABMs present unique validation challenges produced by their generative epistemology and technical complexity.

Agent-based modeling is a cornerstone of generative social science, which seeks to construct artificial societies where macro-level phenomena emerge from the interactions of individual agents. This bottom-up approach enables researchers to investigate the dynamics of social processes, revealing emergent patterns that may be difficult to predict using traditional statistical or equation-based methods (Epstein, 2008).

ABMs can model even poorly understood systems. Simon argues that simulation remains valuable despite incomplete knowledge, as it abstracts essential properties (Simon, 1996). While real-world validation of such a model might be problematic due to complexity, its value is obvious and lies in explanatory power, not accurate depiction (Ahrweiler & Gilbert, 2005).

ABM Validation Methods

Traditionally, models are used to make predictions. However, as Epstein argues, ABMs are often better suited for other scientific purposes beyond prediction, especially in dynamic and non-equilibrium environments where traditional methods struggle (Epstein, 2008). Key purposes of ABM might include:

- **Explaining phenomena** rather than predicting – like plate tectonics explaining earthquakes without forecasting occurrences, ABMs illuminate underlying mechanisms (Epstein, 2008).
- **Testing theory robustness** by simulating extreme/impossible conditions (Epstein, 2008).
- **Generating hypotheses** by simulating alternative processes and testing plausibility (Epstein, 2008).
- **Bounding outcomes** by identifying feasible/impossible scenarios under given assumptions (Gilbert et al., 2018).
- **Enhancing communication** through intuitive system representations for stakeholders (Collins et al., 2024).
- **Training and education** via counterfactual simulations (Epstein, 2008).

Model validation determines whether a model adequately represents its target system for its intended purpose (Collins et al., 2024), ensuring sufficient accuracy for its application domain (Sargent, 2010). However, computational models serve as mediating instruments between theory and reality, requiring validation that assesses this mediating role rather than direct correspondence (Morrison & Morgan, 1999).

While some distinguish verification (testing model performance) from validation (testing system representation), these concepts overlap in practice (Gräbner, 2018). We treat them as equivalent activities for establishing model credibility.

Each modeling paradigm has distinct philosophical foundations shaping its validation approach. **Statistical models** prioritize observable patterns through goodness-of-fit measures and significance tests (Pearl, 2009). **Mathematical models** assume equations represent natural laws, using curve fitting and analytical benchmarks. **Network models** focus on structural validation where position determines behavior (Wasserman & Faust, 1994). **Game-theoretic models** use equilibrium analysis from rational choice axioms (Reiss, 2011). **System dynamics** emphasizes feedback loops through structure-behavior tests (Barlas & Carpenter, 1990). **Machine learning** focuses on prediction accuracy alone, ignoring mechanisms (Prosperi, 2020).

Agent-based models follow generative epistemology, asking “Can you grow it?” not “Can you explain it statistically?” (Epstein & Axtell, 1996). ABMs must show that proposed mechanisms generate observed patterns, requiring methods that evaluate emergence and micro-to-macro dynamics. Gilbert notes that model validity “is often determined by whether it generates useful scientific work rather than strict empirical verification” (Gilbert et al., 2018).

This creates distinctive validation challenges: **disciplinary fragmentation** (methods accepted in one field may be rejected in another); **observability constraints** (many systems like historical changes cannot be observed) (Sargent, 2010); **data limitations** (missing behavioral details, no comparison metrics); **computational complexity** (thousands of variables requiring sophisticated analysis) (Lee et al., 2015); **stakeholder communication** (complex uncertainty is difficult to interpret); and **emergent properties** (patterns irreducible to components).

These constraints require choosing validation methods that maximize credibility within available resources. Table 1 presents common ABM validation methods with their applications, strengths, and limitations.

Model credibility develops gradually through multiple tests across the development lifecycle rather than a single post-development test (Sargent, 2010). This iterative process requires validation activities at all stages, with changes at any stage potentially affecting the conceptual or computerized model, necessitating repeated cycles until satisfactory results are achieved.

Computational experiment

A computational experiment conducted in the summer of 2025 assessed which validation methods work best for different ABM scenarios. The experiment examined combinations of six model purposes (explaining phenomena, validating theory, generating hypotheses, bounding outcomes, communicating findings, and training), four data availability levels (rich data, partial/noisy data, proxy data only, and no data), and four development stages (conceptual, implementation, refinement, and final validation). This factorial design produced 96 scenarios. Twenty unrealistic combinations where certain purposes cannot be achieved at specific development stages were excluded, leaving 76 feasible scenarios.

Each scenario’s 12 validation methods were scored 0–2 based on systematic interpretation of validation literature (Collins et al., 2024; Lee et al., 2015; Sargent, 2010) and logical constraints. A score of 0 means the method cannot be applied in that context, 1 means usable but not ideal, and 2 means recommended. For example, empirical validation receives a score of 0 when no data exists (impossible to apply), while sensitivity analysis receives a score of 2 during refinement (ideal for exploring parameters).[1]

To analyze how validation strategies differ across scenarios, Euclidean distances between method scores were calculated:

$$Distance(i,j) = \sqrt{\Sigma \ \left(V_i^k - V_j^k\right)^2},$$

where k represents one of twelve validation methods, i and j are the corresponding scenarios, and V_i^k and V_j^k are the validation scores of method k for those scenarios.

These distances show how similar or different validation approaches are between contexts. The maximum observed distance (8.5) occurs between scenarios with different validation strategies. For example, a “rich data, final stage” scenario can use nearly all validation methods. In contrast, a “no data, conceptual stage” scenario can only use universal methods like expert review and visualization. Distances ranging from 0 to 8.5 confirm that validation requirements vary substantially depending on model purpose, available data, and development stage. This variation underscores why one-size-fits-all validation approaches cannot work for ABMs.

[1] Complete scoring matrices, computational scripts, and detailed results are available at https://doi.org/10.5281/zenodo.15633195

Table 1. **ABM Validation Methods**

Method	Category	Description	Validation Outcome	Strengths	Weaknesses
Data Analytics	Data Validation	Analyzes input/output datasets for consistency and trends	Ensure data quality and reliability	Detects patterns and inconsistencies; supports transparency	Requires specialized skills; costly and time-consuming
Sampling	Input Validation	Explores parameter space using experimental designs	Achieve comprehensive parameter exploration	Ensures parameter coverage; identifies influential inputs	Computationally expensive; needs careful design
Bootstrapping	Input Validation	Resamples data to generate empirical distributions	Establish statistical confidence with limited data	Enables validation with limited data; extrapolates population information	Not widely trusted; multiplicity problems
Structural Validation	Conceptual Model Validation	Evaluates model architecture and theoretical foundations	Ensure theoretical consistency	Maintains conceptual fidelity and internal consistency	Time-consuming; requires deep understanding; expert disagreements
Expert and Community Validation	Model Analysis	Engages experts/stakeholders to assess plausibility	Achieve expert consensus on credibility	Provides real-world insight across validation stages	Subjective; participant biases
Role-Playing and Interactive Validation	Process Validation	Participatory simulation with humans as agents	Verify behavioral realism	Tests agent behaviors; validates assumptions; improves understanding	Difficult setup; limited to human systems
Inverse Generative Social Science	Process/Conceptual Validation	Creates alternative models to explore equifinality	Establish mechanism uniqueness and robustness	Provides model comparisons; tests uniqueness	Requires a priori behavior specification; computationally intensive
Causal Analysis	Process Validation	Explores event chains producing behaviors	Verify logical consistency and mechanisms	Confirms logic; identifies errors; reproducible	Requires structured data; design biases
Visualization	Model Analysis	Uses graphs/animations to interpret results	Enable error detection and communication	Easily interpretable; detects errors	Subjective interpretation; depends on visualization quality
Empirical Validation	Descriptive Output Validation	Statistically compares outputs with real data	Establish model-reality correspondence	Direct evidence of fit; reproducible	Depends on data quality; impossible for non-observable systems
Docking (Model-to-Model Comparison)	Model Output Corroboration	Compares outputs with established models	Achieve consistency across approaches	Ensures consistency; easy implementation	Groupthink risk if models share assumptions
Sensitivity Analysis	Model Analysis	Tests the effect of parameter changes on outputs	Identify influential parameters	Explores what-if scenarios; finds key parameters	Computationally expensive; may miss interactions

Cluster analysis revealed three method categories based on coverage patterns and performance consistency across scenarios. Figure 1 demonstrates distinct categories based on applicability frequency (coverage percentage) and consistency patterns (standard deviation).[2]

Universal Methods demonstrate broad coverage with consistent applicability and low variability. Expert review and visualization show high applicability across all scenarios, while structural checks maintain near-universal presence. These methods are effective regardless of context.

Context-dependent methods like data analytics and causal analysis show moderate to high coverage but notable variability in applicability. Data analytics is essential when data are available but unusable otherwise; causal analysis is appropriate when models are executable and data exist. This

[2] Detailed clustering analysis scripts and full results available at https://doi.org/10.5281/zenodo.15633195

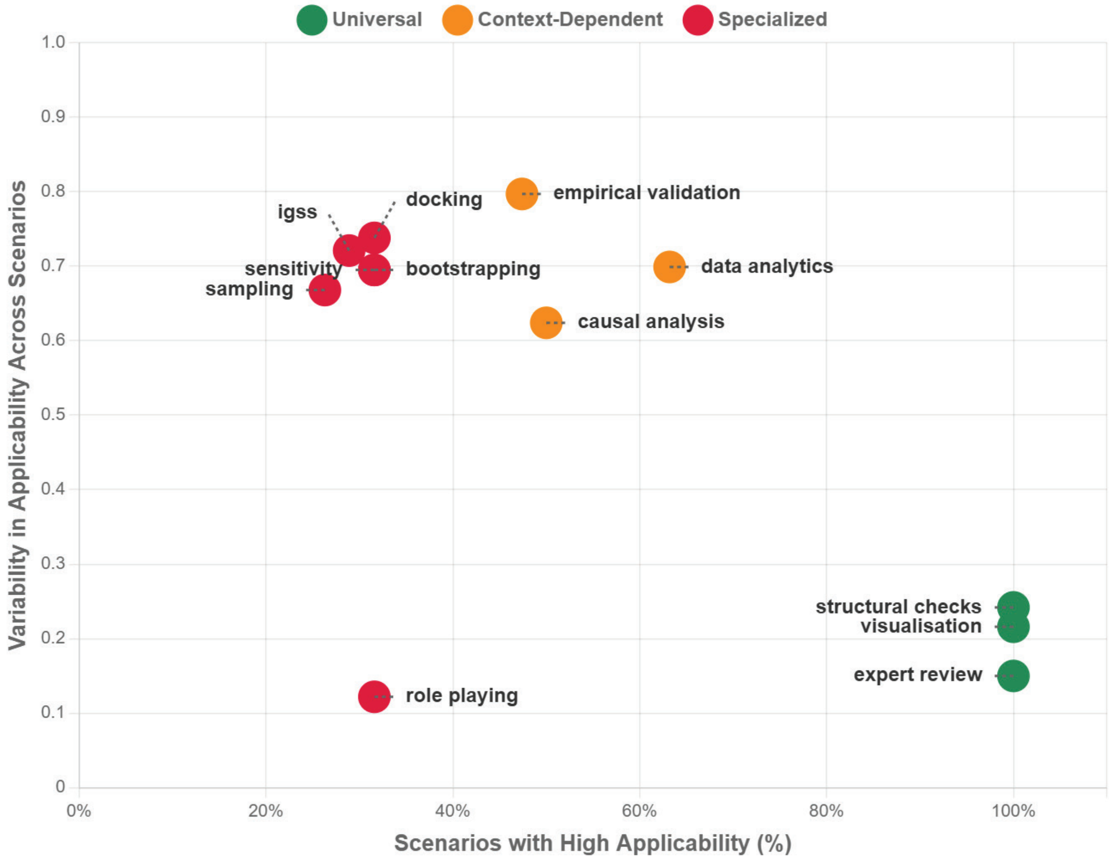


Fig. 1. Validation Methods Grouped by Applicability Frequency and Consistency of Use.
Horizontal axis: applicability frequency, vertical axis: standard deviation of methods' applicability across validation scenarios. Cluster 1 (green) – universal methods, cluster 2 (amber) – context-dependent methods, and cluster 3 (red) – specialized methods.

variability confirms their dependence on specific validation contexts rather than universal suitability.

Specialized Methods include six techniques with lower coverage and high variability: empirical validation ranges from essential (data-rich cases) to inapplicable (unobservable cases). IGSS, model docking, sensitivity analysis, bootstrapping, and sampling require specific conditions, which make them either effective or unusable depending on scenario characteristics. Role-playing forms a distinct pattern: low coverage and low variability; it is applicable only in stakeholder-driven contexts but consistently relevant there.

This analysis reveals no methods in the "medium coverage, low variability" zone. Validation methods are either broadly usable or specialized, with little in between. The pattern supports a constraint-based interpretation: contextual conditions, not researcher preferences, determine method applicability.

ABM Validation Method Selection Framework

To understand what drives validation method selection, we analyzed eight boundary scenarios (the most extreme cases showing the broadest range of validation approaches). We used a greedy max-min selection algorithm that picked scenarios as different from each other as possible, ensuring they represent truly different validation approaches rather than slight variations. We compared these scenarios by counting how many validation methods each could use, noting which methods were applicable (scoring above 0) and which were ideal (scoring above 1.5).

The analysis showed significant differences in method availability. The most flexible scenarios could use 10–12 validation methods; the most limited could only use 4–6 methods. These scenarios formed distinct groups with clear boundaries rather than a smooth range. When we compared purpose, data availability, and development stage between high-availability and low-availability scenarios, we found two main constraints:

The Observability Constraint. Scenarios with rich data could use all 12 methods; scenarios with no observable data could only use three universal methods (visualization, expert review, structural checks). This is a hard limit: historical events, hypothetical scenarios, and things we cannot

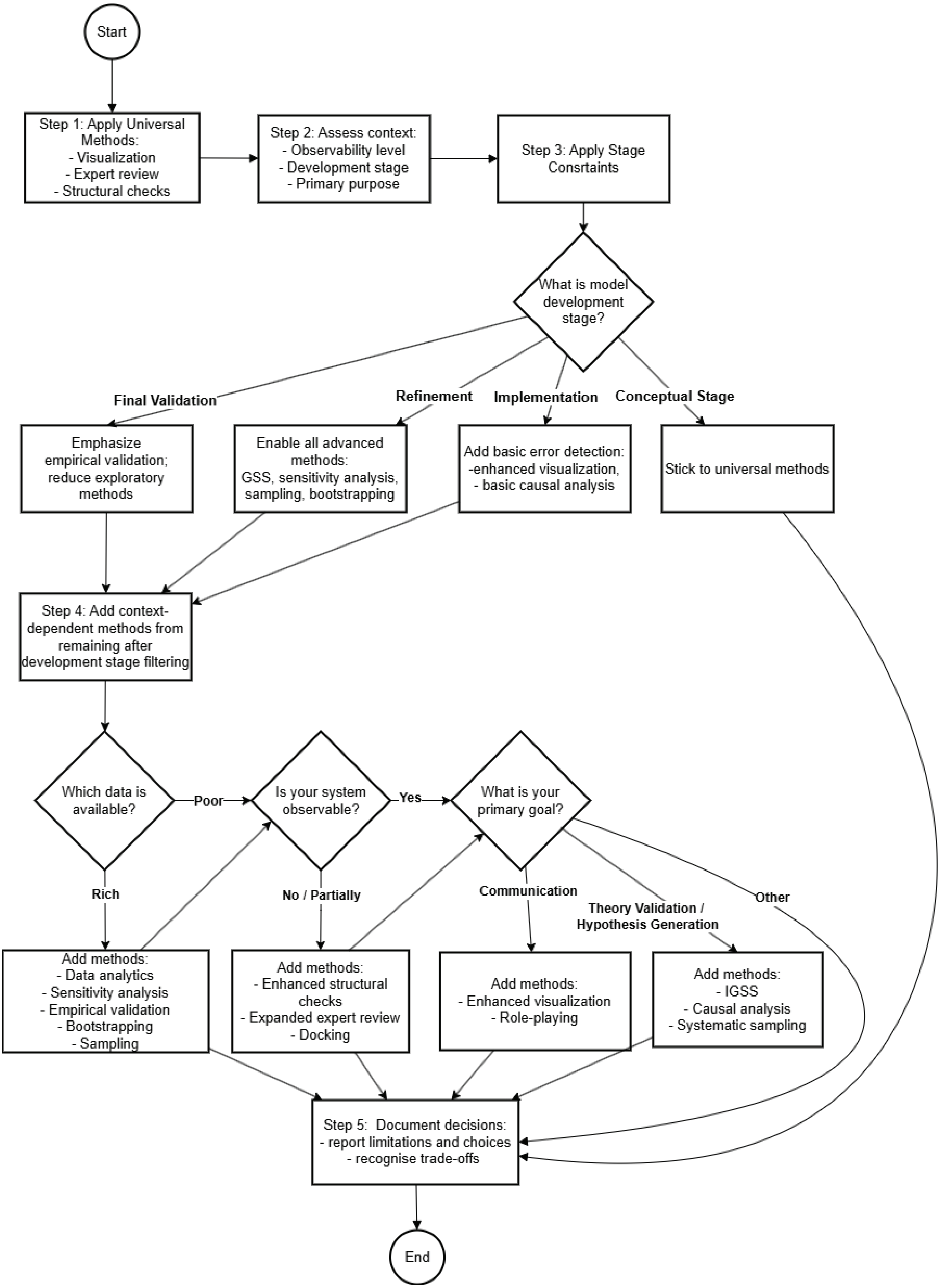


Fig. 2. ABM Validation Method Selection Framework

Table 2. **Framework for Selection of ABM Validation Method**

	Step	Action	Key Decisions
1	**Assess context**	Evaluate the situation across three dimensions	**System observability**: rich data, partial/noisy, proxies only, or none **Development stage**: conceptual, implementation, refinement, or final **Primary purpose**: theory validation, hypothesis generation, bounding outcomes, explanation, communication, or training
2	**Apply Universal Methods**	Always include core validation approaches	**Visualization** (error detection and communication) **Expert review** (domain expertise verification) **Structural checks** (theoretical consistency)
3	**Apply Stage Constraint**	Filter methods by development stage	**Conceptual**: Universal methods only **Implementation**: Add basic error detection (enhanced visualization, basic causal analysis) **Refinement**: Enable all advanced methods (IGSS, sensitivity analysis, sampling, bootstrapping) **Final**: Emphasize empirical validation and documentation; reduce exploratory methods
4	**Add Strategy-Specific Methods**	Select additional methods based on context	**If rich data is available**: • Add empirical suite: Data Analytics – Sensitivity Analysis – Empirical Validation • Include Bootstrapping for uncertainty quantification • Use comprehensive sampling **For limited/no observability:** • Focus on theoretical validation via Structural Checks • Emphasize Expert Review • Use Docking when comparable models exist **For communication/training goals**: • Enhance visualization with interactive elements • Conduct Role-Playing exercises • Keep technical methods minimal **For theory validation/hypothesis generation**: • Use IGSS if data/resources permit • Emphasize Causal Analysis for mechanism verification • Use systematic sampling to explore theoretical space
5	**Recognize Trade-offs**	Document limitations and choices	Document the approaches used and why Be transparent about constraints Consider if the purpose-stage combination is atypical Acknowledge that some methods will be impossible Prepare different presentations for scientific and stakeholder audiences

measure cannot be validated with data, regardless of researchers' desires or resources.

The Development Stage Constraint. Early, conceptual-stage scenarios could only use theoretical validation methods. Implementation-stage scenarios allowed basic empirical testing. Only refinement and final-stage scenarios could use advanced methods like sensitivity analysis and inverse generative social science, showing how validation options grow as models develop.

The model's purpose worked differently. While the purpose influenced which methods researchers should focus on, it did not limit which methods they could use.

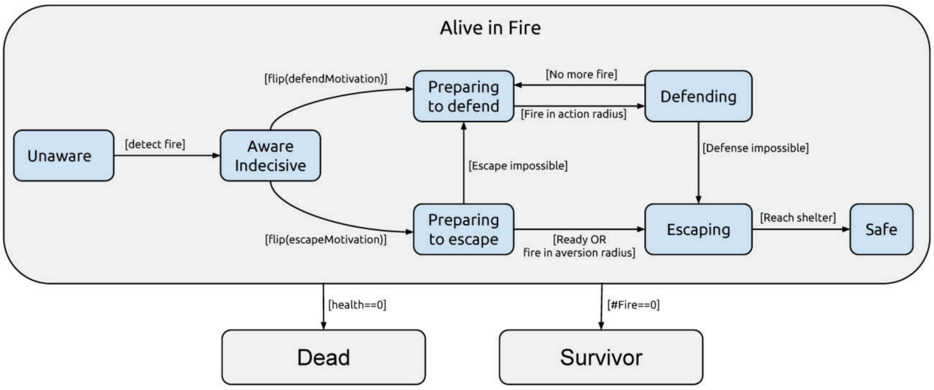
Based on these findings, we developed a practical framework following this constraint logic: first assess the context, then apply universal methods, then add more methods based on available data, development stage, and research purpose. Figure 2 shows the framework visually, and Table 2 provides step-by-step guidance.

Table 2 details the specific actions and method recommendations for each pathway.

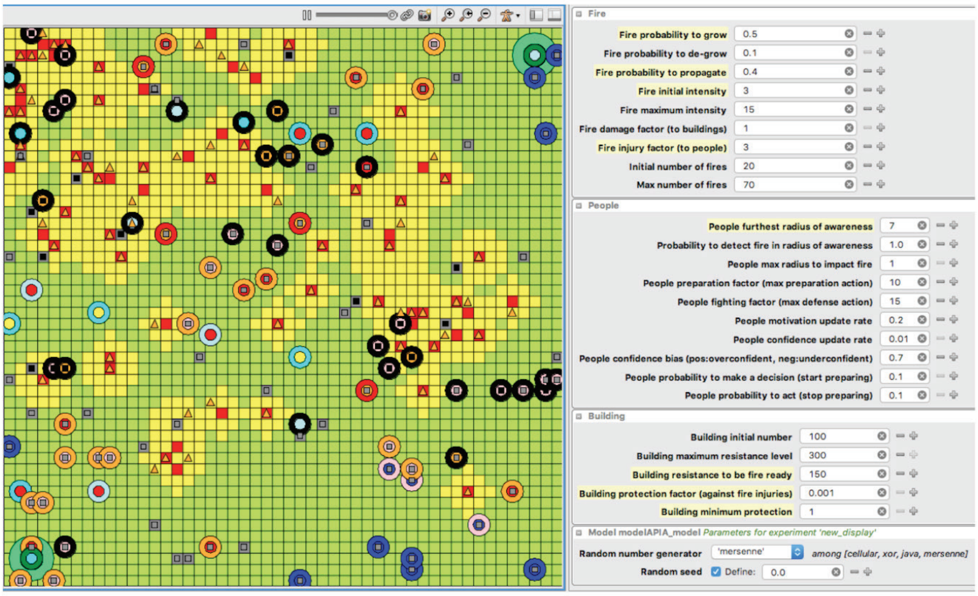
Choosing Validation Methods in Practice: A Case Study

To demonstrate our validation framework's practical application, we analyze how Adam and Gaudou approached validation for their bushfire evacuation model.

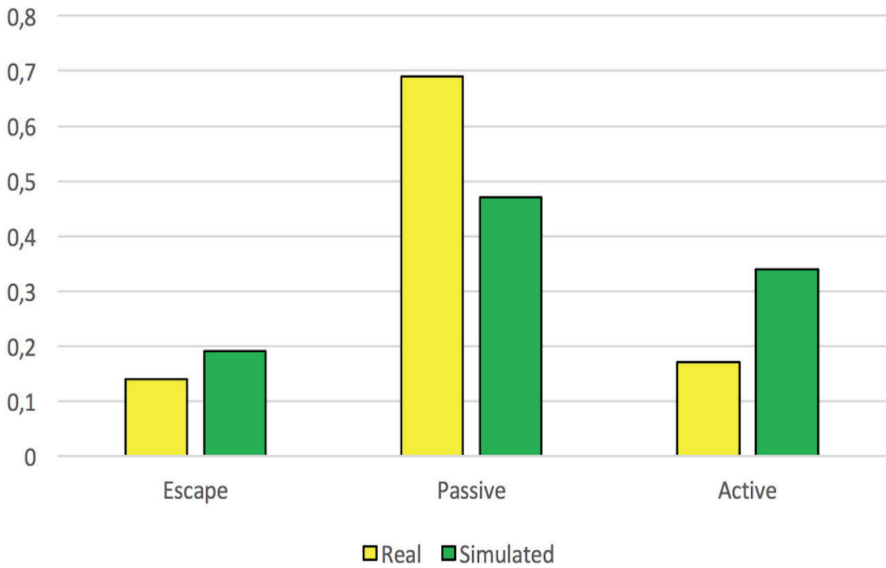
Adam and Gaudou's disaster behavior study shows how ABM integrates qualitative and quantitative data for policy analysis of interventions that cannot be tested in reality (Adam & Gaudou, 2017). They modeled resident behavior during the 2009 Black Saturday bushfires in Victoria, Australia, to explain why actual crisis behavior differ from emergency planners' expectations. Using finite-state machines, they tracked residents through states (Unaware, Aware, Indecisive, Preparing to Defend, Defending, Preparing to Escape, Escaping, Safe, Dead, and Survivor) based on subjective perceptions. Their key finding: gaps between objective danger and subjective perceptions explain seemingly irrational crisis behavior.



(a) State-Transition Diagram of Resident Behavior



(b) GAMA Simulation Interface with Parameter Settings



(c) Comparison of Real vs. Simulated Mortality Causes

Fig. 3. An integrated overview of the fire-evacuation model
(adapted from Adam & Gaudou, 2017)

The model includes all essential ABM features: heterogeneous agents with varied attributes, autonomous decision-making, spatial grid environment, local agent-fire-building interactions, bounded rationality through subjective evaluations, and dynamic state transitions (Epstein, 2006). Implemented in GAMA, the model combines state transitions, decision logic, spatial representation, and parameter controls to generate outputs comparable to empirical data (Fig. 3).

The model provides emergency planners with key insights into unexpected resident behaviors, showing how perception gaps drive crisis decisions. However, validation is challenging because human thoughts and emotions during disasters cannot be directly observed. Only outcomes are documented.

The authors pursued two validation objectives: achieving correctness by comparing simulated trajectories with observed behaviors, and demonstrating explanatory value by showing how perception gaps affect survival. They prioritized explaining mechanisms over precise prediction, which is suitable for highlighting human factors in disaster response.

According to our framework, the Adam and Gaudou model is a **partially observable system** with moderate data availability. Though the 2009 bushfires produced extensive records (100 witness statements, 86 police hearings, death statistics), the core phenomena – cognitive processes and subjective perceptions during crisis – remain unobservable. The model serves an **explanatory purpose**, revealing behavioral mechanisms rather than making predictions. It reached the **final validation stage**, requiring empirical assessment of the complete implementation.

This configuration places the model in a **hybrid validation approach**, combining data-driven and theory-based elements from our analysis. Despite partial observability limiting some methods, rich qualitative data and the final development stage enable more validation options than purely theoretical scenarios. Our analysis indicates that partially observable systems with qualitative data support 6–8 validation methods, between the 4–6 methods for theory-based validation and 10–12 for fully observable systems. Table 3 represents our ABM validation method selection framework applied to the developed model.

The analysis shows that Adam and Gaudou successfully implemented universal validation methods through systematic qualitative analysis of

Table 3. **Validation Method Assessment Using Practical Validation Framework**

	Framework Step	Recommended Actions	Authors Implementation
1	**Context Assessment**	Evaluate observability, development stage, and primary purpose	Correctly identified as partially observable, explanatory purpose, final stage
2	**Universal Methods Application**		
	Visualization	Visualize in all contexts	Implemented via the GAMA platform interface
	Expert Review	Involve experts in all contexts	Not Implemented
	Structural Checks	Check the implementation of theoretical mechanisms in all contexts	Implemented via alignment with Lazarus' stress theory and cognitive bias literature
3	**Apply Stage Constraints** (final validation stage emphasizes empirical validation, communication, and documenting)		
	Filter out exploratory methods	IGSS, Sampling, Bootstrapping	Correctly avoided
	Keep in consideration	Causal Analysis, Data Analytics, Sensitivity Analysis	Selected subset
	Consider high-priority	Empirical Validation, Role-playing	Selected a subset but missed some priorities
4	**Strategy-Specific Methods Selection** (for explanatory models with partial data, with adjustment for the final validation stage)		
	Causal Analysis	Verify mechanisms when feasible	Implemented via discrepancy analysis between objective/subjective factors
	Empirical Validation	Limited to available data	Implemented via death statistics comparison
	Data Analytics	Extract patterns from data	Systematic qualitative analysis of 100+ interviews
	Sensitivity Analysis	Critical for explanatory models	Not Implemented
	Docking	Compare with other models	Not feasible (no comparable models existed)
	Role-playing	Stakeholder engagement	Not Implemented
5	**Trade-off Recognition**	Document limitations and choices	Acknowledged the equifinality problem but deferred critical analyses

interview data and a solid theoretical foundation in psychological frameworks. Their use of structural validation, visualization, causal analysis, and empirical validation provides a solid foundation. They followed our proposed framework by emphasizing empirical methods at the final stage and avoiding exploratory methods, but missed three key priorities:

No Expert Review. The authors incorporated psychological expertise but lacked validation from emergency managers and fire survivors, critical for a model meant to inform emergency practice. Domain experts could confirm whether the identified behavioral patterns match real-world experience.

Missing Role-Playing Validation. For a model focused on subjective perceptions during crises, role-playing exercises would be valuable. Participants could directly experience perception gaps and validate emotional responses, providing insights beyond interview data. Though the authors plan future game development, role-playing during model development would have added unique value.

Postponed Sensitivity Analysis. This is essential for the explanation of the mechanisms, not just addressing equifinality. Without it, researchers cannot confirm whether the modeled objective-subjective discrepancy reliably explains outcomes or cannot understand how parameter uncertainty affects findings. Basic sensitivity analysis would strengthen their conclusions.

The Adam and Gaudou case shows how validation must balance contextual constraints with scientific rigor. Their transparent discussion of limitations and systematic qualitative approach makes their conclusions persuasive, and adding expert review, sensitivity analysis, and role-playing validation would improve credibility while maintaining their focus on explaining behavioral mechanisms.

Conclusion

This study addresses fundamental challenges in agent-based model validation by establishing ABM's distinctive theoretical foundations and developing a practical framework for validation method selection.

Through computational analysis of 76 validation scenarios, we demonstrated that the appropriateness of validation methods depends systematically on model purpose, system observability, and development stage. Our distance-based analysis revealed that scenarios require fundamentally different validation approaches, with feasible methods ranging from 4 to 12, depending on context. We identified three validation method categories based on applicability patterns: universal methods (visualization, expert review, structural checks) applicable across all contexts; context-dependent methods (data analytics, causal analysis) varying with data availability; and specialized methods (empirical validation, IGSS, sensitivity analysis, bootstrapping, sampling, model docking) requiring specific conditions.

Our practical framework guides researchers through systematic validation method selection: beginning with context assessment, ensuring universal methods application, adding strategy-specific methods based on constraints, and acknowledging necessary trade-offs. Application to Adam and Gaudou's bushfire model demonstrated the framework's utility in identifying both successful strategies and critical gaps.

Our findings show that adequate ABM validation requires systematic consideration of contextual constraints rather than standardized protocols. Future work should extend this framework to emerging ABM applications incorporating machine learning and large language models.

Acknowledgements

The author thanks Olena Pugachova for inspiration in the field of agent-based modeling and computational social sciences, and for ideas and constructive comments during the work on this paper.

References

Adam, C., & Gaudou, B. (2017). Modelling Human Behaviours in Disasters from Interviews: Application to Melbourne Bushfires. *Journal of Artificial Societies and Social Simulation, 20*(3), 12. https://doi.org/10.18564/jasss.3395

Ahrweiler, P., & Gilbert, N. (2005). Caffè Nero: The Evaluation of Social Simulation. *Journal of Artificial Societies and Social Simulation, 8*(4), 1–14. https://www.jasss.org/8/4/14.html

Barlas, Y., & Carpenter, S. (1990). Philosophical roots of model validation: Two paradigms. *System Dynamics Review, 6*(2), 148–166.

Collins, A., Koehler, M., & Lynch, C. (2024). Methods That Support the Validation of Agent-Based Models: An Overview and Discussion. *Journal of Artificial Societies and Social Simulation, 27*(1), 11. https://doi.org/10.18564/jasss.5258

Dean, J. S., Gumerman, G. J., Epstein, J. M., Axtell, R. L., Swedlund, A. C., Parker, M. T., & McCarroll, S. (2000). Understanding Anasazi culture change through agent-based modeling. In T. A. Kohler & G. J. Gumerman (Eds.), *Dynamics in Human and Primate Societies: Agent-Based Modelling of Social and Spatial Processes* (pp. 179–205). Oxford University Press.

Epstein, J. M. (2006). *Generative social science: Studies in agent-based computational modeling.* Princeton University Press.

Epstein, J. M. (2008). Why Model? *Journal of Artificial Societies and Social Simulation, 11*(4), 12. https://www.jasss.org/11/4/12.html

Epstein, J. M. (2023). Inverse generative social science: Backward to the future. *Journal of Artificial Societies and Social Simulation, 26*(2), 9. https://doi.org/10.18564/jasss.5083

Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom up.* MIT Press.

Gilbert, N., Ahrweiler, P., Barbrook-Johnson, P., Narasimhan, K. P., & Wilkinson, H. (2018). Computational Modelling of Public Policy: Reflections on Practice. *Journal of Artificial Societies and Social Simulation, 21*(1), 14. http://jasss.soc.surrey.ac.uk/21/1/14.html. http://dx.doi.org/10.18564/jasss.3669

Gräbner, C. (2017). How to relate models to reality? An epistemological framework for the validation and verification of computational models. *Journal of Artificial Societies and Social Simulation, 21*(3), 8. https://www.jasss.org/21/3/8.html

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S. K., Huse, G., Huth, A., Jepsen, J. U., Jørgensen, C., Mooij, W. M., Müller, B., Pe'er, G., Piou, C., Railsback, S. F., Robbins, A. M. … DeAngelis, D. L. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling, 198*(1–2), 115–126. https://doi.org/10.1016/j.ecolmodel.2006.04.023

Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese, 169*(3), 615–626. https://doi.org/10.1007/s11229-008-9435-2

Lee, J.-S., Filatova, T., Ligmann-Zielinska, A., Hassani-Mahmooei, B., Stonedahl, F., Lorscheid, I., Voinov, A., Polhill, J. 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

Lorig, F., Johansson, E., & Davidsson, P. (2021). Agent-based social simulation of the Covid-19 pandemic: A systematic review. *Journal of Artificial Societies and Social Simulation, 24*(3), Article 5. https://doi.org/10.18564/jasss.4601

Morrison, M., & Morgan, M. S. (1999). Models as mediating instruments. In M. S. Morgan & M. Morrison (Eds.), *Models as mediators: Perspectives on natural and social science* (pp. 10–37). Cambridge University Press.

Paniotto, V. I., & Hrushetskyi, A. M. (2013). Chy shche ne vmerlo modeliuvannia? Istoriia sotsialnoho modeliuvannia v Ukraini ta ahentno-oriientovanyi pidkhid na prykladi prohnozuvannia movnoi sytuatsii v Ukraini. In Ye. I. Holovakha & T. Ya. Liubova (Eds.), *Suchasni metody zboru i analizu danykh v sotsiolohii: Materialy VI Mizhnarodnykh sotsiolohichnykh chytan pamiati N. V. Paninoi* (pp. 23–47). Institute of Sociology of the National Academy of Sciences of Ukraine [in Ukrainian].

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.

Prosperi, M., Guo, Y., Sperrin, M., Koopman, J. S., Min, J. S., He, X., Rich, S., Wang, M., Buchan, I. E., & Bian, J. (2020). Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence, 2*(7), 369–375. https://doi.org/10.1038/s42256-020-0197-y

Pugachova, O. (2019). Agent-based Simulation of Revenue Policy. *NaUKMA Research Papers. Sociology*, *2*, 16–22. https://doi.org/10.18523/2617-9067.2019.2.16-22 [in Ukrainian].

Pugachova, O. (2021). Modelling COVID-19 epidemic and its social consequences. *NaUKMA Research Papers. Sociology*, *4*, 18–27. https://doi.org/10.18523/2617-9067.2021.4.18-27 [in Ukrainian].

Reiss, J. (2011). A plea for (good) simulations: Nudging economics toward an experimental science. *Simulation & Gaming, 42*(2), 243–264.

Sargent, R. G. (2010). Verification and validation of simulation models. *Proceedings of the 2010 Winter Simulation Conference* (pp. 166–183). Baltimore, MD, USA. https://doi.org/10.1109/WSC.2010.5679166

Schelling, T. C. (1971). Dynamic models of segregation. *The Journal of Mathematical Sociology, 1*(2), 143–186. https://doi.org/10.1080/0022250X.1971.9989794

Serdyuk, A. (2025). Policy responses and organizational adaptation of Ukrainian universities during the full-scale Russian invasion: An agent-based approach. In T. Nahorniak & I. Zaichenko (Eds.), *Mizhdystsyplinarni ekspertyzy dlia vidnovlennia i rozvytku Ukrainy: zbirnyk materialiv mizhnarodnoi naukovo-praktychnoi konferentsii (m. Kyiv, 5 chervnia 2025 r.)* (pp. 216–221). National University of Kyiv-Mohyla Academy. https://doi.org/10.5281/zenodo.16421646

Simon, H. A. (1996). *The Sciences of the Artificial* (3rd ed.). MIT Press.

Squazzoni, F., Polhill, J. G., Edmonds, B., Ahrweiler, P., Antosz, P., Scholz, G., Chappin, É., Borit, M., Verhagen, H., Giardini, F., & Gilbert, N. (2020). Computational models that matter during a global pandemic outbreak: A call to action. *Journal of Artificial Societies and Social Simulation, 23*(2), Article 10. https://doi.org/10.18564/jasss.4298

Vengrina, I. (2012). Language dynamics via simulation methods. *NaUKMA Research Papers. Sociology*, *135*, 19–24. https://ekmair.ukma.edu.ua/server/api/core/bitstreams/abe50615-8f4b-4259-bbd6-0009431e9a8d/content [in Ukrainian].

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.

Артем Сердюк

ВИБІР МЕТОДІВ ВАЛІДАЦІЇ ДЛЯ АГЕНТНИХ МОДЕЛЕЙ: ПРАКТИЧНІ ПІДХОДИ

Агентне моделювання виникло як обчислювальний підхід для вивчення складних соціальних явищ через дослідження взаємодії агентів на мікрорівні, проте практика валідації моделей залишається непослідовною.

У статті запропоновано практичний підхід для вибору методів валідації. За допомогою обчислювального експерименту автор проаналізував 76 сценаріїв валідації, які охоплюють шість цілей моделювання (пояснення явищ, валідація теорії, генерація гіпотез, обмеження результатів, комунікація висновків, навчання), чотири рівні спостережуваності (багаті дані, часткові або зашумлені дані, лише проксі-дані, відсутність спостережуваності) та чотири етапи розроблення (концептуальний, імплементація, вдосконалення, фінальна валідація).

Кластерний аналіз виявив три категорії методів валідації: універсальні методи (візуалізація, експертне оцінювання, структурні перевірки), які використовуються широко і незалежно від контексту; контекстно-залежні методи (аналітика даних, каузальний аналіз), застосування яких залежить від наявності даних; та спеціалізовані методи (емпірична валідація, IGSS, аналіз чутливості, бутстрепінг, семплінг, докінг), які потребують специфічних умов. Аналіз крайових сценаріїв виявив два основні обмеження, які впливають на вибір методів валідації: обмеження спостережуваності та обмеження етапу розроблення. На основі цих висновків розроблено практичний підхід вибору методу валідації агентної моделі на основі контексту: 1) оцінювання контексту; 2) застосування універсальних методів у всіх випадках; 3) вибір контекстно-залежних методів за етапом розроблення; 4) додавання спеціалізованих методів на основі контексту; 5) визнання компромісів та документування обмежень.

Застосування підходу, запропонованого Керол Адам (Carole Adam) і Бенуа Году (Benoit Gaudou) до моделі евакуації під час лісових пожеж, продемонструвало його цінність. Аналіз виявив як успішні стратегії валідації, використані авторами (систематичний якісний аналіз, структурна валідація), так і прогалини в їхньому підході (бракує експертного оцінювання, рольових ігор, відкладений аналіз чутливості).

Результати дослідження свідчать, що ефективна валідація моделей агентного моделювання вимагає систематичного оцінювання контекстуальних обмежень, що допомагає дослідникам обирати найбільш відповідні методи валідації розроблених агентних моделей.

Ключові слова: агентне моделювання, валідація моделей, обчислювальні соціальні науки, генеративні соціальні науки, соціальна складність, складні системи, невизначеність, українське суспільство.

Подано / Submitted: 08.06.2025
Схвалено до публікації / Accepted: 19.08.2025
Оприлюднено / Published: 11.12.2025

Сердюк Артем — аспірант кафедри соціології Національного університету «Києво-Могилянська академія»

Serdyuk Artem — PhD student at the Department of Sociology of the National University of Kyiv-Mohyla Academy

https://orcid.org/0000-0002-0060-1311

a.serdiuk@ukma.edu.ua