Artem Serdyuk,
PhD-3 student, Department of Sociology,
National University of Kyiv-Mohyla Academy
a.serdiuk@ukma.edu.ua
https://orcid.org/0000-0002-0060-1311

VALIDATION STRATEGIES FOR AGENT-BASED MODELS IN SOCIAL SCIENCE: INSIGHTS FROM MODELING OF MIGRATION CAUSED BY THE FULL-SCALE RUSSIAN INVASION OF UKRAINE

Agent-based modeling (ABM) is a cornerstone of generative social science, which seeks to construct artificial societies where macro-level phenomena emerge from the interactions of individual agents (Epstein & Axtell, 1996). In these models, agents are autonomous entities with decision-making rules that govern their interactions with each other and their environment.

ABMs possess defining characteristics that distinguish them from other modeling approaches. Epstein (2008) enumerates the following: heterogeneity, autonomy and bounded rationality of agents, spatiality and local interactions between them, and non-equilibrium dynamics of the studied systems.

Key purposes of ABM might include explaining phenomena, generating hypotheses, bounding possible outcomes, enhancing stakeholders' communication, and training and education through counterfactual simulations. A notable aspect of ABMs is their capacity to model systems even when those systems are not fully understood.

Despite their advantages, ABMs have well-known limitations. Collecting the necessary data to calibrate and validate models can be challenging (Lee et al., 2015). Also, comparing model outcomes to reality becomes challenging when real-world data is scarce or incomplete. It is particularly problematic when modeling poorly understood systems (Gilbert et al., 2018).

Epstein (2008) emphasizes that all researchers operate with implicit or explicit models. Unlike implicit models, ABMs have explicitly defined rules, making them testable and subject to validation. Gilbert emphasizes that a model's validity is often determined by whether it generates valuable scientific work rather than by strict empirical verification.

Validation in ABM can take multiple forms:

- Comparing model output with empirical data: when available, models can be calibrated and tested against real-world observations (Gilbert et al., 2018).
- Theoretical consistency: a model should align with existing theoretical knowledge about the studied system. If empirical validation is impossible, consistency with established theories can provide confidence in the model's validity (Collins et al., 2024).
- Expert validation: engaging domain experts can help assess whether the model's assumptions and results align with real-world knowledge (Gilbert).
- Sensitivity analysis: examining how small changes in input parameters affect outcomes can reveal the robustness of a model's conclusions (Lee et al., 2015).

Collins et al. (2024) emphasize that the choice of validation method depends on the model's purpose and data availability (see Table 1).

To apply these validation methods, this study presents an agent-based model designed to simulate individual decision-making processes in an environment exposed to war atrocities. Its purpose is to study the war-induced migration patterns and test the influence of specific push and pull factors on refugees' decision to leave or return home. The author developed the model using the NetLogo 6.4 environment. The main flow of the model is shown in Figure 1.

Table 1. **Preferred Methods of ABM Validation**

Prerequisites	Preferred Validation Methods
Real-world data is available,	Researchers can use empirical validation or bootstrapping to compare the model with observed outcomes
Causal relationships need to be established	One can use causal analysis to assess internal logic
Robustness is the primary concern	One should use data analytics or sensitivity analysis
If no empirical data exists, but there are plausible theoretical models	One should use docking to compare with existing models. Role-playing is an option to check the plausibility
Interpretability is key	Use visualization to ensure results align with theoretical expectations
The primary goal is to understand the complexity of the system, observe emerging phenomena, and communicate them to others	Visualization should also be a method of choice

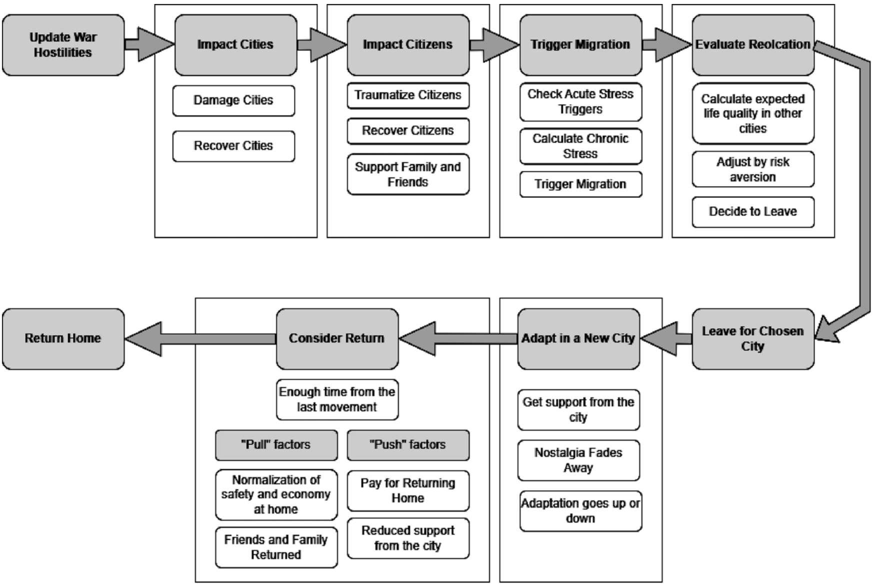

Fig. 1. War-Induced Migration Model Process Overview

The model implements most of the ABM's key features. Heterogeneity is implemented by individual agents with distinct attributes generated randomly. Agents act independently, making relocation and return decisions without any centralized management. The environment comprises cities with different war exposure, recovery capabilities, infrastructure, and population. The spatial dimension of the model is shown in Figure 2.

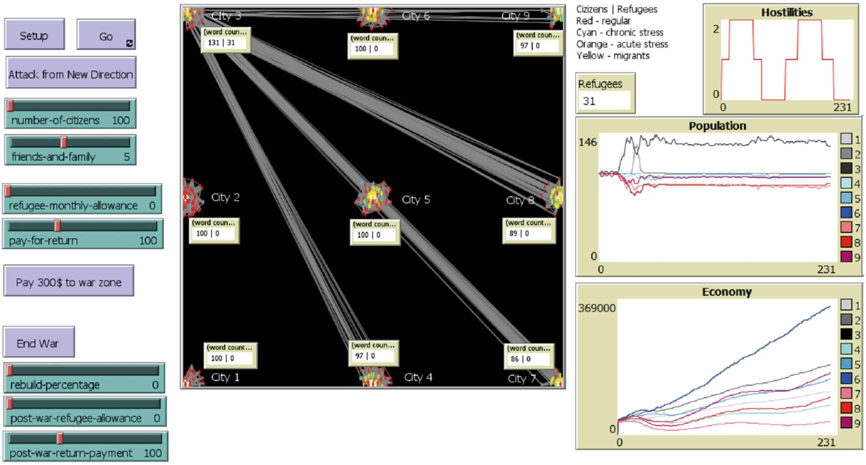

Fig. 2. Spatial dimension of the model of war-induced migration

The calibration and validation strategy of the model is limited to the availability of empirical migration data in absolute numbers compatible with those generated by the model to make empirical validation possible. Although much quantitative research has been done about Ukrainian war migrants, this data can hardly be compared to the data produced by the model, especially considering factors that make refugees return home.

Docking, in its turn, requires robust social and psychological models to explain the underlying mechanisms of migration decisions. Due to the challenges of empirical validation, we prioritize visualization as the primary validation method, allowing us to compare emerging migration patterns with observed trends. For comparison, we have chosen studies of war-induced migration from Ukraine held by the Centre for Economic Studies in 2023, 2024, and 2025. An example of comparing data from the CES report (Mykhailyshyna et al., 2023, p. 18) to the data generated by the model is shown in Figure 3.

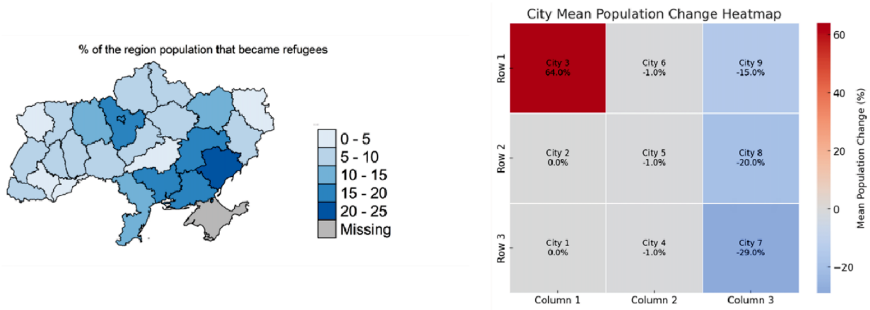

Fig. 3. Population Left the Regions (Left) Compared to the Population Loss Simulated by the Model (Right)

As shown in Figure 2, survey and model data are not directly comparable. However, the migration pattern is similar in both data generated by the model and collected in the survey—the highest percentage left from the areas directly affected by the full-scale Russian invasion. The further the region is from the front line, the less population loss there are in the survey and the simulation. This comparison shows that simulation produces plausible migration patterns based on the agents' decisions and interactions.

Also, visualization is a plausible way to validate an agent-based model of this kind. Our further research will be dedicated to other ways to verify and calibrate the model from the present data and combine the agent-based learning approach with other research approaches to strengthen its weak points.

References

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.

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

Gilbert, N., et al. (2018). Computational modelling of public policy: Reflections on practice. *Journal of Artificial Societies and Social Simulation*, 21 (1), 14. https://doi.org/10.18564/jasss.3669.

Lee, J.-S., et al. (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.

Mykhailyshyna, D., Samoiliuk, M., & Tomilina, M. (2023, August 29). *Refugees from Ukraine: Who are they, how many are there and how to return them? Final report*. Centre for Economic Strategy. https://ces.org.ua/wp-content/uploads/2023/09/refugees.pdf.