Combined machine learning and agent-based modeling for studying environmental migration dynamics

Kelsea Best [a], Jonathan Gilligan [a]

[a] Earth and Environmental Sciences, Vanderbilt University, 5726 Stevenson Center, Nashville, TN 37240 USA

Abstract

The decision to migrate is influenced by many factors including livelihood opportunities, social norms, individual preferences, and characteristics of both origin and potential destination locations. At the same time, environmental impacts may interact with migration decisions across multiple scales and in non-trivial ways. In this context, environmental migration is a notoriously complex phenomenon within coupled human-environmental systems. Because of the complexity of environmental migration, we propose a combination of advanced methods based on both empirical data and modeling, where a single approach may be more limited. In this work, we use both machine learning and agent-based modeling to investigate dynamics of environmental migration in coastal Bangladeshi communities. To begin, we use random forest models to identify salient predictors of migration within a large household survey consisting of data from more than 3,000 individuals from approximately 1,695 households across 9 sites in southwestern Bangladesh. Building upon insights from the random forest results, we use hazards models to further explore the ways that salient variables impact a household's migration outcomes. From these data-driven analyses as well as relevant literature and theory, we construct an original agent-based model to simulate migration from a hypothetical origin location. The model consists of individual, household, and community level characteristics and processes, allowing for representation of the multi-level dimensions of environmental migration. The model is parameterized with empirical data from southwestern Bangladesh, and uncertain parameters are calibrated using a support vector machine classification algorithm. Model results are validated using a pattern-oriented approach, and we show that the agent-based model is successfully able to reproduce relevant patterns of migration at the household and community level. Finally, our agent-based model is used as a testing bed to assess different methods of decision-making around migration at the household level. With this approach, we are able to show that community-level inequality in land ownership as well as job availability are critical factors that influence environmental migration dynamics. This important insight stemmed from this multi-step and multi-method approach. Where a single method may provide insights into environmental migration, we therefore show how the use of multiple methods in combination and provide a deeper understanding of dynamics at play.