Ben-Gurion University of the Negev
The Faculty of Natural Sciences
The Department of Life Sciences

A multiscale cellular automata–Markov framework for modelling shifting cultivation and agroforestry landscapes

Thesis submitted in partial fulfillment of the requirements
for the Master of Sciences degree

Yuval Bloch

Under the supervision of: **Prof. Shai Pilosof**

June 2026 **סיון תשפ"ו**

Ben-Gurion University of the Negev
The Faculty of Natural Sciences
The Department of Life Sciences

June 2026 **סיון תשפ"ו**

A multiscale cellular automata–Markov framework for modelling shifting cultivation and agroforestry landscapes

Thesis submitted in partial fulfillment of the requirements
for the Master of Sciences degree

Yuval Bloch

Under the supervision of: **Prof. Shai Pilosof**

Signature of student: _____ Date: 08/12/2026

Signature of supervisor: _____ Date: 12 August 2026

Signature of chair of graduate studies committee: _____ Date: _____

* Please add a signature line for each supervisor, when there are more than one supervisor

A multiscale cellular automata–Markov framework for modelling shifting cultivation and agroforestry landscapes

Yuval Bloch

Master of Sciences Thesis

Ben-Gurion University of the Negev

2026

Abstract

Land-use change is a major driver of environmental degradation and public health risk worldwide, contributing to biodiversity loss, soil degradation, and the emergence of zoonotic diseases. Although numerous models can project future land-use change across diverse settings, most have been developed for urban or industrial agricultural landscapes characterized by broad-scale, long-term transitions. Our ability to model smallholder shifting cultivation and agroforestry systems — which operate across multiple spatial and temporal scales — remains comparatively limited.

This gap matters because rural communities in tropical, low-income regions depend on these agroecosystems for their livelihoods, often in landscapes that overlap with major biodiversity hotspots. These communities face acute vulnerability to land degradation and exposure to disease, making a deeper understanding of their land-use dynamics essential for both human well-being and ecological conservation.

To address this, we developed a modeling framework that fits landscape-level composition and configuration indices rather than relying on cell-by-cell spatial agreement. To capture the multiscale structure of these landscapes, we incorporated fractal geometry into the modeling approach. We calibrated the model against observed land-use maps in the Sava region of northeastern Madagascar, then projected the equilibrium landscapes expected under a gradient of agroeconomic scenarios representing increasing reliance on vanilla agroforestry, and compared their composition and configuration. As a proof of concept, we coupled the projected landscapes to a tick population model.

Calibration showed that our land-use indices could constrain only two configuration parameters — the spatial stochasticity of cultivation and the spatial stochasticity of natural change — indicating that the limitation lay in the sensitivity of the indices rather than in the complexity of the model. The configuration indices responded to differences in vanilla-cultivation practice only once vanilla dominated the landscape, whereas the tick model

responded to these differences whenever any vanilla was present, highlighting a limitation of configuration indices relative to a biological model. In the coupled model, tick exposure changed non-monotonically, and the effect of spatial preference was stage-dependent.

Together, these results demonstrate the value of the framework for informing land-management and public health policy in smallholder tropical systems. Beyond tick exposure, the approach extends to questions of land degradation, water management, and biodiversity conservation. More broadly, the study clarifies the relationship between fractal landscape properties and dynamic agricultural systems, advancing our understanding of shifting cultivation and agroforestry landscapes worldwide.

Acknowledgements

I would like to express my deepest gratitude to the Department of Life Sciences at Ben-Gurion University of the Negev for providing a supportive and stimulating environment throughout my Master's studies. Special thanks to the administrative staff for their endless patience and invaluable help in navigating the submission process.

To my advisor, Dr. Shai Pilosof: thank you for granting me the freedom to pursue my ideas, even when they sounded unfeasible, and for providing the precise guidance needed to make them feasible after all. Thank you for teaching me how to conduct rigorous research, for helping me recognize and develop good ideas, and for standing by me with unwavering support whenever I lost confidence in myself or my work.

To the collaborators on the Madagascar team: thank you for generously sharing your hard-won data with me. I can barely imagine the challenges of collecting it over such a long period in such a demanding environment. Thank you also for the many fruitful discussions, from which I learned so much, and for the chance to be part of an international research team where I was always made to feel welcome, valued, and equal.

To my colleagues in the Ecological Complexity Lab: thank you for accepting me completely, quirks and all. Thank you for the insightful discussions during lab meetings that meaningfully improved my work, the lunchtime conversations that reminded me how wonderful it is to be a nerd, and all the social events and emotional support we shared along the way.

I would also like to extend my sincere thanks to the teachers whose courses and shared wisdom opened entirely new directions for my research. A special thank-you goes to Professor Yaron Ziv. While your course on Spatial Ecology gave me an excellent introduction to the spatial modeling techniques I used, I am especially grateful for introducing me to the concept of fractal dimensionality, which became a vital pillar of this thesis.

Finally, to my friends and family: the intensity of this research, unfolding against the continuous backdrop of the ongoing war, pushed me to my emotional limits. Thank you for being there when I was exhausted, when I lost faith in myself, and when I felt I simply could not go on. A deeply special thank-you to Lizi Ribchinsky and Rachel Golan, with whom I could always share life's hardships and safely admit my truest exhaustion. And to Pazit Frankel — the time we spent together on my writing retreats to Mitzpe Ramon is part of what made those trips so deeply inspiring. And, of course, to my parents, who taught me to be curious and to love both math and nature.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Land use represents the fundamental interface between humanity and nature. Every human activity, whether in dense urban centers or remote rural communities, involves occupying and altering the environment to meet specific needs. This environment, in turn, critically impacts human well-being. Our food sources, health, exposure to disease, and vulnerability to natural disasters are intrinsically linked to our surroundings and how we modify them. Land use change is a primary driver of global environmental crises. It is the main cause of 85% of endangered-species cases [1], drives the emergence of 27% of vector-borne zoonotic diseases [2], is the leading cause of soil loss [3], and accounts for almost a quarter of greenhouse-gas emissions [4]. Predicting how different patterns of land-use change will unfold and affect such processes is crucial for developing management strategies that minimize negative consequences for both the environment and human populations.

Understanding the complex impacts of land-use change remains a significant challenge. The immense spatiotemporal scale of these processes makes field experiments time-consuming, expensive, and often infeasible. Therefore, scientists frequently use computer simulation models. These models forecast how land-use strategies under different economic, cultural, regulatory, and environmental regimes will affect critical indicators, including biodiversity, disease emergence, and soil degradation.

To predict these indicators, simulations of land-use change must account for two properties of land-use maps: (1) land-use composition, which is the proportion of different land uses on a map, and (2) land-use configuration, which is the spatial arrangement of land-use categories on the map. These two components are modeled with different frameworks. The most common approach for quantifying changes in land use composition is the Markov chain. Transition rates between categories are derived from past observations and used to project future proportions of each class. Future land-use configurations are typically modeled using Cellular Automata (CA). CA models divide land into grid cells that change state based on their neighbors' states. Repeatedly applying this process allows the model to capture complex, nonlinear spatial dynamics. The two approaches can also be combined into a single CA–Markov framework — a rapidly growing field with diverse applications and methodological extensions [5, 6]. For example, Hamad [5] examined how economic sanctions influenced deforestation and land degradation in a remote region of Iraq. They used maps from four time points to estimate land-transition rates for each category under both sanction and non-sanction periods. Coupled with a CA model, they

projected alternative futures with and without sanctions, effectively linking geopolitical events to ecological processes. Similarly, Zare [7] combined simulated land-use maps with projected climate trends and a soil degradation model to estimate future soil loss in multiple scenarios in the Kasilian basin in Iran. Extending these methods, Huang and Liao [8] replaced predefined CA rules with an artificial neural network to predict land-use change in the rapidly urbanizing Yangtze River basin in China, where population displacement has accelerated expansion. They then evaluated how projected land-use changes altered habitat connectivity in several animal groups.

These examples highlight the utility and versatility of spatial CA–Markov land-use change simulations across diverse contexts (agricultural and urban), drivers (development, sanctions, and population displacement), and objectives (connectivity and land degradation). However, despite their success, CA–Markov models are used mainly in directional land-use change. Directional land-use systems are those in which cells that change from one land use to another have a low probability of returning to their first state, such as modern agriculture and urban systems. **To date, there have been only a few attempts to model smallholder shifting cultivation agroecosystems.** This oversight is striking, given that hundreds of millions of farmers in the least developed tropical regions depend on these systems for their subsistence [9, 10]. In addition, these farms are often located within biodiversity hotspots and critical carbon sinks, which are highly sensitive to environmental degradation. The intricate mosaic of natural and anthropogenic land uses in these regions intensifies human–nature interactions.

Modeling cyclical agroecosystems, such as shifting cultivation, is inherently challenging. Unlike conventional land-use modeling frameworks based on pixel-by-pixel comparisons, these systems operate across multiple temporal and spatial scales. One temporal scale captures long-term transitions driven by ecological, demographic, and economic processes. A second temporal scale reflects short-term, cyclical land-use changes such as cultivation, degradation, and recovery. These changes occur over much shorter periods and are more strongly influenced by stochastic processes. As a result, the land use observed in a given spatial unit at a single point in time provides only limited insight into the long-term dynamics of land-use change [11].

This limitation is well illustrated by *tavy*, a traditional shifting cultivation system in remote areas of eastern Madagascar. Within a *tavy* cycle, individual land parcels may rapidly transition between cultivation and different stages of natural succession. At the same time, other parcels may become permanently degraded. To compensate for declining local land productivity, households may clear new forest areas, driving a slower but more persistent process of village relocation upslope in search of new arable land [12]. When comparing static land-use maps, transitions of individual parcels between stages of the *tavy* cycle reveal little about this broader, long-term transition of settlement patterns and land-use frontiers. Similar challenges arise when broader structural processes reshape shifting cultivation systems. These include transitions toward cash-crop production that reduce reliance on shifting cultivation [13], as well as changes in regulatory regimes that permit or restrict the inclusion of specific land areas in the shifting cultivation cycle [14].

In many least-developed regions, these challenges are exacerbated by complex dynamics across multiple spatial scales. Traditional land ownership and inheritance practices often result in highly fragmented, small land plots [15]. Combined with degradation patterns from shifting

cultivation, variable soils, and steep topography, this fragmentation produces extreme heterogeneity at the scale of tens of meters [16]. Conversely, land-sharing traditions [12] and communal fencing practices [11] create landscape structures that only emerge at much broader scales. Capturing both extremes simultaneously presents a significant methodological hurdle, known as the "middle-number problem" [17]. The problem is that averaging fine-scale data to capture high-level trends erases critical heterogeneity, while attempting to model broad-scale structures from the bottom up remains computationally intractable. Even before modeling, the challenge already manifests in mapping: classifying smallholder shifting cultivation from satellite imagery typically requires complex models trained on high-resolution spatiotemporal remote sensing time series [18].

In such complex systems, pixel-level agreement at small scales is often unachievable. Instead, models are fitted and evaluated based on different indices. Different indices capture different aspects of the land-use map, and choosing among them shapes both the perceived predictive strength and the model's structural limitations. Two notable studies that modeled this specific kind of system are Jepsen [11] and Wickramasuriya [19], each of which captures different aspects of land-use dynamics. Jepsen [11] investigated the rationale for shifting cultivation mosaics in remote Vietnam by developing an agent-based model grounded in simple behavioral assumptions to handle multiple temporal scales. They tuned the model to a quasi-equilibrium, a state in which the evaluation index, the clustering index known as the Mean Nearest Neighbor distance, becomes stable. The model successfully replicated observed clustering under specific parameter settings. However, this index captures only local land-cell interactions, making it insensitive to large-scale patterns.

Similarly, Wickramasuriya [19] modeled the "Chena" shifting cultivation system in Sri Lanka using an extended CA framework that incorporates land-use history to handle multiple temporal scales. The authors evaluated model performance by comparing land-use compositions across different aggregation scales, partitioning the map into windows at each scale to assess whether areas had the correct land-use composition (proportions). The model outperformed a random null model, but this was only evident at a coarse scale of approximately 8 by 8 kilometers. Thus, it could predict landscape-scale patterns but could not accurately allocate them at a fine local level. While Jepsen [11] tuned their model to a single local clustering statistic and Wickramasuriya [19] to composition at coarse scale only, neither could constrain landscape structure across scales simultaneously, leaving a gap in our ability to model fine-scale shifting cultivation.

Here, we develop a framework to model fine-scale shifintg cultivation land use. Building on the approach of Jepsen et al., we simulate the system to a quasi-equilibrium state, reached once the land-use composition stabilizes. We then fit the model using lacunarity, a fractal property describing geometric patterns that persist across scales; tropical land-use change has been shown to exhibit such fractal properties [20]. Because our fitted parameter persists across scales, a single model can reproduce both the fine-grained mosaic and the broad-scale organization that earlier approaches captured only one at a time.

We apply this framework to address the challenge of predicting land-use change in a cyclic agroecosystem in villages around Marojejy National Park in northeastern Madagascar. Madagascar is a major global biodiversity hotspot, yet rapid deforestation continues to threaten its ecosys-

tems, driven in many regions by intensive shifting cultivation [21]. At the same time, market integration is expanding nationwide through the growth of cash crops such as vanilla and cloves. In several regions, the transition from subsistence farming to vanilla agroforestry reduces deforestation and even drives reforestation [13], although deforestation continues in most regions [14, 22].

Smallholdings of less than one hectare per household are widespread in Madagascar, and landholdings continue to shrink over generations, producing a mosaic of increasingly fragmented plots [15]. Environmental risks are also severe: Madagascar faces a high burden of zoonotic diseases, including malaria, dengue, and lymphatic filariasis [23], as well as recurrent outbreaks of *Yersinia pestis* (plague) [24]. To analyze how long-term market-integration dynamics interact with short-term shifting-cultivation dynamics, we developed a model to predict land-use change at multiple spatial resolutions, ranging from fine-scale agricultural plots ($10\,\mathrm{m} \times 10\,\mathrm{m}$) to the broader village scale ($5\,\mathrm{km} \times 5\,\mathrm{km}$), using high-resolution classification maps [25]. With this model, we generate maps under different future scenarios of market integration in the study area—scenarios that drive the transition from predominantly subsistence shifting cultivation to agroforestry cash crops (vanilla)—and examine how they affect land-use characteristics.

To demonstrate the relevance of our model for understanding the consequences of land-use change, we couple it with a tick population model to predict tick exposure. Tick exposure can serve as a proxy for tick-borne disease risk, a major public health concern worldwide. Ticks are responsible for 40% of vector-borne zoonotic diseases, many of which occur in Africa—a region that remains underrepresented in tick research [2]. Furthermore, the relationship between ticks and land use requires examination at multiple temporal and spatial scales [26].

1.1 Research objectives

The objectives of this study are as follows:

O1 – Model development: To construct a predictive land-use model capable of simulating the dynamics of cyclical agroecosystems at short- and long-term timescales, by incorporating relevant socio-ecological processes.

O2 – Scenario projection: To generate alternative future land-use maps based on varying economic and agricultural scenarios, reflecting different potential development pathways.

O3 – Multiscale evaluation: To evaluate and compare these scenarios using a suite of land-use evaluation indices designed to capture distinct aspects of land-use configuration.

O4 – Use case: To demonstrate the usefulness of the approach for downstream scenario testing by integrating a tick population model into the land-use projections. In this use case, we will predict how different future scenarios influence human exposure to tick-borne pathogens.

Chapter 2

Materials and Methods

We provide detailed mathematical formulations, algorithms, and the full implementation code in the Supplementary Material. Here, we describe the main components of the modeling framework (Fig. 2.1).

2.1 Modeling Framework Overview

This research models land-use change in dynamic agroecosystems—a task that is challenging on two fronts:

1. **Data limitation.** Classifying dynamic land use, such as shifting cultivation, is inherently difficult, and as a result, complete land-use classification maps of such systems are rare.
2. **Modeling challenge.** The land use of individual cells changes annually, making it difficult to predict how these systems will evolve.

To address these challenges, we developed a model that, rather than tracking the land use of individual cells, captures higher-order spatial patterns evident across the landscape. Two components characterize these patterns:

1. **Land-use composition** — the proportion of each land-use type present on the map.
2. **Land-use configuration** — the spatial arrangement of those land-use types, quantified through a set of configuration indices.

To do so, we use a **CA–Markov framework**. First, we calibrate the **transition demand**, which represents the proportion of land cells of each type that are expected to transition to each other land-use type. Next, we calibrate the **transition potential**, a formula that estimates the suitability or fitness of each cell for each possible transition. Finally, using the transition potential, we stochastically allocate the required transitions by randomly selecting cells.

After calibrating the model to reproduce observed landscape patterns, we use it to explore how different development scenarios affect those patterns. Each scenario represents a plausible agroeconomic trajectory, operationalized by adjusting the corresponding model parameters. Using these scenarios, we generate different future land-use maps (Fig. 2.1b). We then examine how

the scenarios influence land-use indices, providing insights into the landscape-level consequences of specific agroeconomic decisions (Fig. 2.1c,d).

Finally, we demonstrate the broader applicability of this framework by coupling it with a tick population model (Fig. 2.1e). By running the tick model on the projected landscape maps for each scenario, we show that the framework can be directly integrated into epidemiological modeling. We further analyze how changes in land-use indices affect tick exposure, linking landscape dynamics to epidemiological outcomes (Fig. 2.1f).

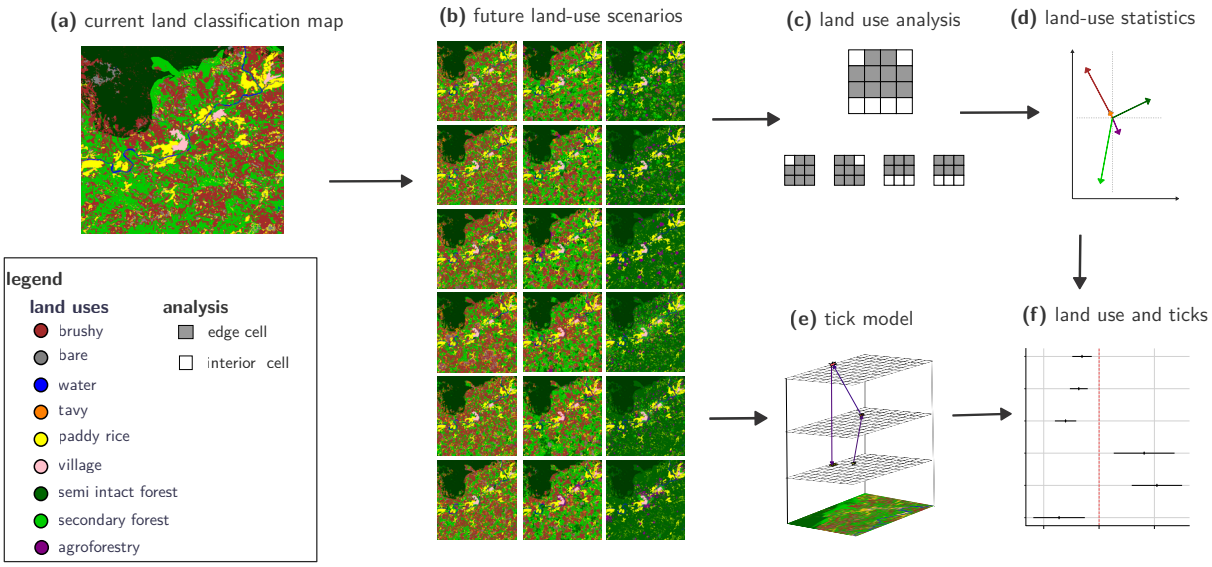


Figure 2.1: Model overview. A simplified overview of the modeling framework used throughout this paper. **(a)** The starting point is the current land-use classification map (Fig. 2.2). **(b)** Projected land-use maps are generated under multiple agroeconomic scenarios (Fig. 3.2). **(c,d)** Each projected map is analyzed using a suite of landscape composition and configuration indices (Fig. 2.5). **(e)** To demonstrate the biological relevance of the land-use projections, a tick population dynamics model is applied to each map to quantify how landscape patterns translate into human tick exposure (Fig. 2.6). **(f)** Finally, we apply linear regression to assess how landscape composition and configuration indices predict tick exposure (Fig. 3.5).

2.2 Study Area

We built this model for the agroecological system of the villages surrounding Marojejy National Park in Madagascar (Fig. 2.2a). This area is both an agroecosystem worthy of study in its own right and a compelling example of the unique challenges that characterize dynamic land-use modeling in developing regions.

Understanding this system is a vital step toward protecting it. The park is an important biodiversity hotspot, providing refuge for numerous endangered and endemic species across a steep elevational gradient. Local agricultural practices strongly shape both the park and its surrounding buffer zone. In particular, land degradation within shifting cultivation systems frequently pushes farmers to clear new forest at the park's margins, and occasionally within the protected boundaries (Fig. 2.2b).

Beyond its conservation value, the system is also vital to the communities that depend on it. The close coupling between agriculture and the surrounding ecosystems exposes local populations to

various diseases, while land degradation threatens their primary source of income.

The area also exemplifies the unique challenges of modeling shifting cultivation systems at an early stage of market integration. It is located in the remote **SAVA region** of northeastern Madagascar, a global center of Bourbon vanilla production—a labor-intensive cash crop grown in agroforestry systems (Fig. 2.2c). Although the shift toward vanilla agroforestry has opened pathways for market integration in otherwise isolated villages, most households still depend heavily on subsistence shifting cultivation. The region remains poor and geographically isolated, with only basic infrastructure.

Three main production systems dominate local agriculture: flooded rice cultivation, a traditional shifting cultivation practice known as *tavy*, and vanilla agroforestry for commercial sale. *Tavy* is a primary driver of deforestation and continues even within the Marojejy protected area due to weak law enforcement. Local communities also cultivate a variety of other crops [27].

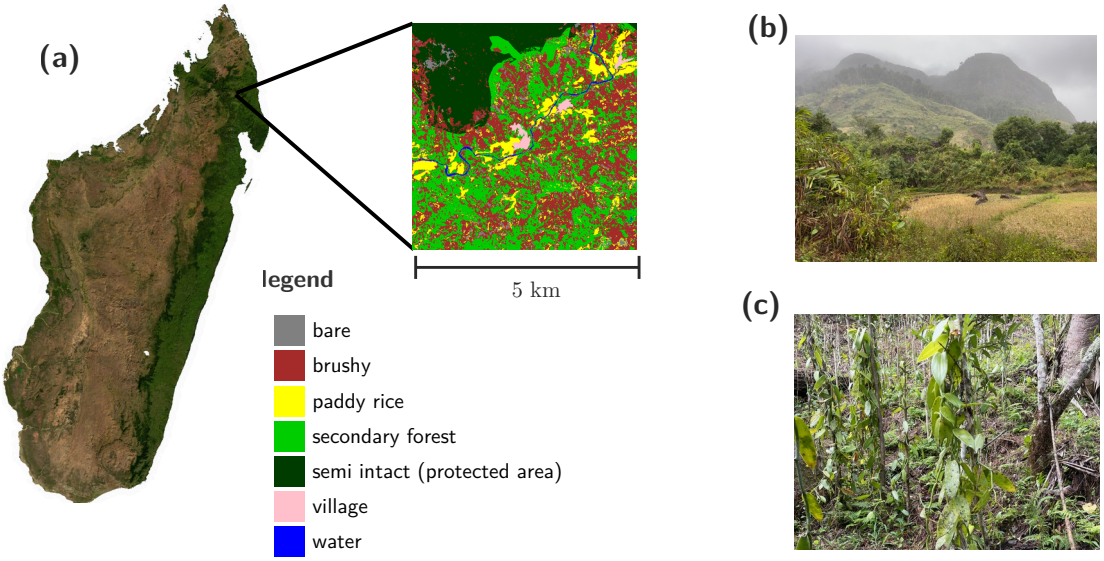


Figure 2.2: Study area description. (a) The study area is located in the SAVA region of northeastern Madagascar, a remote and economically underdeveloped area where many residents rely on traditional subsistence shifting cultivation (*tavy*). **(b)** The *tavy* cycle creates a complex mosaic of active agriculture, land in various stages of natural succession, and recovering forest. **(c)** Vanilla is the dominant cash crop in the area; although it represents a more sustainable practice than *tavy*, it nonetheless drives various forms of agroforestry-associated degradation. Images by Shai Pilosof.

2.3 Available Data and Their Limitations

We draw on two main data sources. The first is a land-cover map from Kauffman [25], who classified land cover using satellite imagery at 10 m resolution into the following categories: semi-intact forest, secondary forest, brushy regrowth, flooded rice, bare ground, and village. This dataset does not explicitly distinguish dynamic agricultural practices — such as *tavy* (the local form of shifting cultivation) or agroforestry — as separate land-cover classes; instead, these are often captured under the secondary forest category. This limitation stems from the fact that accurately mapping smallholder shifting cultivation remains a major challenge, requiring complex models trained on high-resolution spatial and temporal remote sensing time series [18]. Such data are costly to acquire and difficult to apply universally. We adopted an approach

that circumvents the need for explicit classification of shifting cultivation, thereby improving the transferability and robustness of our model.

The second source is documentation of traditional knowledge of shifting cultivation cycles in a nearby region of Madagascar, collected by Styger [12]. We used these data to estimate transition rates between different stages of fallow land, applying a simplified average of rates that, in reality, depend on the burning history of individual land cells. Incorporating burning history would improve accuracy [19] but is beyond the scope of the present study.

Together, these two sources provide a partial picture of the landscape: we know the proportions of different land-cover types across the *tavy* fallow cycle and the recovery rates between them, although we lack information on the proportion or spatial distribution of active *tavy* plots, or the rate at which new plots are created. Similarly, Styger [12] provides rates for different stages of fallow recovery, but not for *tavy* cultivation. By combining these two partial datasets, we can fill in the missing values using a Markov chain approach.

2.4 Markov Chain Model

We model land-use change as a cyclic, ordered system in which each cycle represents one year and the different stages represent seasons. In each season, a specific proportion of one land-use type transitions to another. To this end, we distinguish six dynamic land-use types:

$$\begin{aligned}
\text{ta} &= \textit{tavy} \text{ (active shifting cultivation)},\\
\text{ba} &= \text{bare ground},\\
\text{br} &= \text{brushy regrowth},\\
\text{sf} &= \text{secondary forest},\\
\text{rf} &= \text{fully recovered forest},\\
\text{af} &= \text{agroforestry}.
\end{aligned}$$

In addition to these dynamic land-use types, the study area contains static land-use types that rarely transition: *water*, *irrigated rice*, *village*, and *semi-intact forest* within the protected reserve. Although these categories can, in principle, change, such transitions are sufficiently rare to be treated as negligible. These categories are based on observations reported by Hänke [27].

Following the CA–Markov framework [11], we refer to the annual proportion of land-use type u transitioning into land-use type v as the *transition demand*, denoted by $\delta(u \to v)$ (Fig. 2.3d).

2.4.1 Estimating *Tavy* Transition Demand

The Markov framework above requires a transition-demand value for every transition. Since the agroecosystems of the remote villages on which we base our model are at an early stage of market integration, vanilla agroforestry remains relatively rare; we therefore model the current scenario as one in which *tavy* is the only dynamic land use. Consequently, the transition demands we need to estimate are those associated with *tavy*. The recovery transitions of abandoned *tavy* fields and the equilibrium proportions of different land-cover types across the recovery gradient

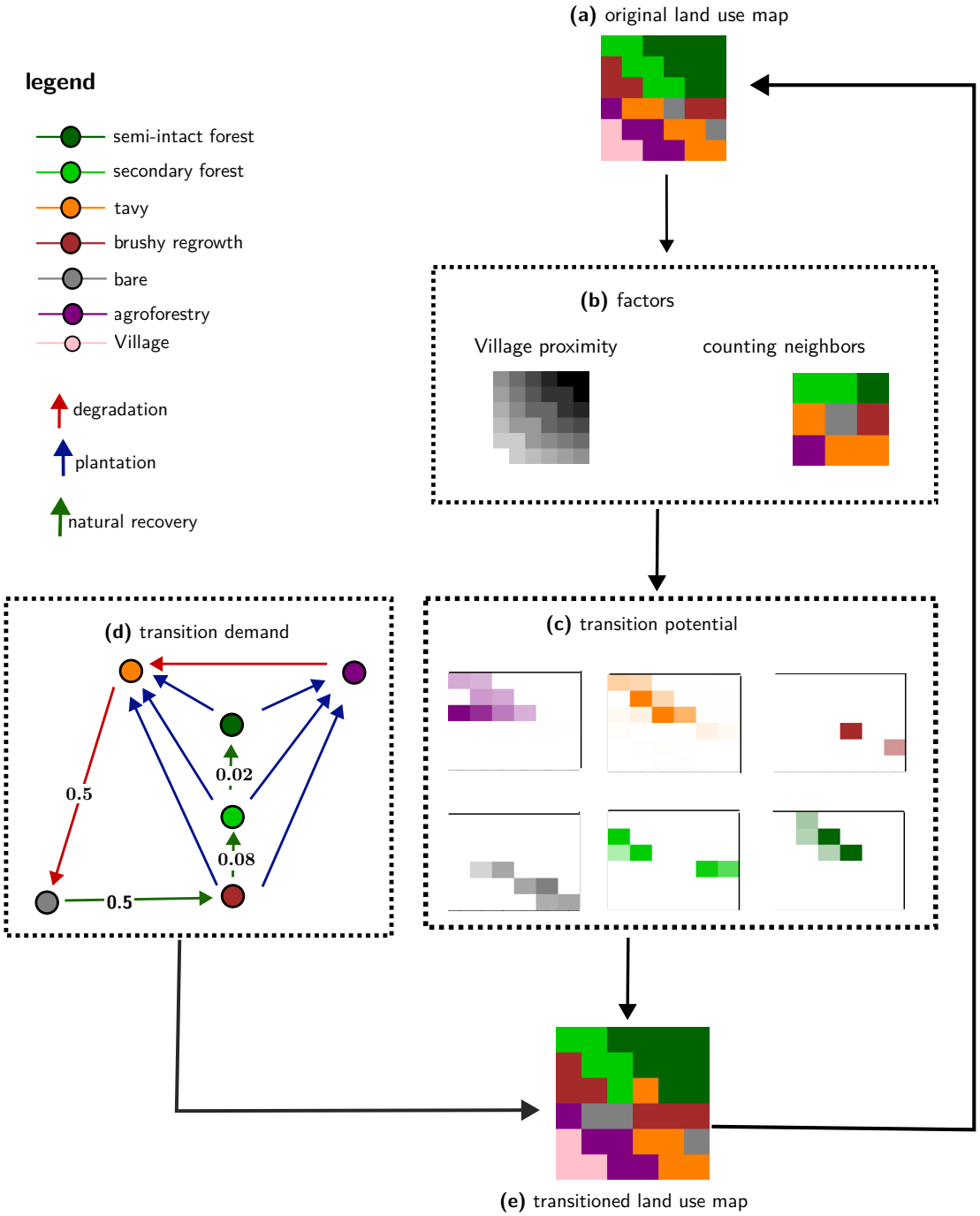


Figure 2.3: Stochastic Cellular Automata transition process. To generate a scenario-based land-use map, we use a CA–Markov model that operates through six main stages. **(a)** The process begins with an initial land-use map. **(b)** Spatial factors are then used to calculate **(c)** the transition potential, which represents the suitability of each land unit to undergo a specific land-use transition. This potential is combined with **(d)** the transition demand, which defines the amount of land expected to change for each transition type in each simulation round. Cells are then selected stochastically to satisfy the required transitions. After the transitions are applied, the model produces **(e)** an updated land-use map, which becomes the input for the next iteration.

are already known; the remaining demands are derived under an equilibrium assumption by solving the equilibrium equations.

Let $\mathbf{C}_t$ denote the vector of land-use proportions involved in the *tavy* cycle in year t:

$$\mathbf{C}_t = \big(C(\text{ta}),\ C(\text{ba}),\ C(\text{br}),\ C(\text{sf}),\ C(\text{rf})\big)^\top. \tag{2.1}$$

Let T_{all} denote the transition rate matrix governing land-use change, so that

$$\mathbf{C}_{t+1} = T_{\text{all}}\,\mathbf{C}_t. \tag{2.2}$$

Under the assumption of equilibrium, Markov chain theory implies that the observed land-use proportions correspond to the stationary distribution of T_{all}, satisfying

$$\mathbf{C}_{\text{current}} = T_{\text{all}}\,\mathbf{C}_{\text{current}}. \tag{2.3}$$

To obtain T_{all}, each transition is encoded as an elementary matrix. For example, the transition from brushy regrowth to *tavy* is represented as:

$$T_{\text{br}\to\text{ta}} = \begin{pmatrix} 1 & 0 & \delta(\text{br},\text{ta}) & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1-\delta(\text{br},\text{ta}) & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.4}$$

The full transition matrix is then given by:

$$T_{\text{all}} = \prod T. \tag{2.5}$$

Solving this system of equations yields all the transition rates needed to complete the composition side of the model. A full description is provided in Sections A.2 and A.3 of the Supplementary Material.

2.4.2 Finding stable equilibrium

To ensure comparability among all simulations, each was conducted for an equal number of iterations. The required number of iterations was determined as the point at which all simulations reached a stable equilibrium. Because the CA–Markov simulation does not converge exactly to the Markov-chain stationary distribution, a "stable equilibrium" is defined as a fluctuating steady state in which the Total Variation Distance (TVD) between land-use compositions, sampled at 20-year intervals, remains consistently below 0.01 across all scenarios. The TVD was computed as follows:

$$\text{TVD} = \frac{1}{2}\sum_{u\in\mathcal{L}} |C_t(u) - C_{t+20}(u)| \tag{2.6}$$

When this criterion is satisfied across all scenarios, we adopt the corresponding iteration count as the simulation horizon used throughout the study. In our system the criterion was met after 160 years; we therefore ran every simulation for 200 years.

2.5 Cellular Automata Model

To model land-use configuration, we assume the landscape is shaped mainly by local interactions. Each interaction raises the propensity of a cell to transform, a quantity we refer to as its transition potential (Fig. 2.3b,c). We define three types of interactions that can increase a land cell's potential.

1. Recovery attracts recovery: recovery occurs more rapidly in areas where a relevant seed bank exists in adjacent cells [12];
2. Cultivation attracts cultivation: cultivation occurs in cells close to already cultivated cells, making fencing and logistics easier [11];
3. Degradation attracts degradation: a degraded cell renders its neighboring cells more vulnerable to water damage.

Based on these interactions, we define an attracting set for each transition as the set of land use types present in the area that increase the transition potential. Because these interactions operate at the level of three broad processes, we group the seven land-use transitions into three *transition categories*—cultivation, degradation, and recovery—each defined by a shared attracting set (Table 3.1). This grouping specifies the attracting set $\mathrm{AT}(u \to v)$ for every transition and, during calibration, lets us tie the neighbor weight within each category.

For each transition, we calculate the transition potential from the origin land use to the target land use for each cell, and then use weighted random sampling—the Gumbel-top-k trick [28]—to determine which cells transition. To prevent a large transition demand from being allocated in a single draw—which would add stochasticity beyond that intended by the neighbor weight—the demand is allocated in successive batches: each batch converts the cells whose potential is within a fixed fraction of the current maximum, and the potentials are recomputed between batches, so the number of batches follows from the landscape. By construction, land-cell interactions affect the configuration of change but not its overall magnitude.

2.6 Transition Potential

The transition potential of cell i from type u to type v is given by

$$\Phi_{i,\,u\to v} = 1 + \alpha(u \to v) \cdot \frac{NW_i}{\overline{NW}}, \qquad NW_i = \sum_{j \in \mathcal{N}_i} \mathbb{I}\Big(s_j^t \in \mathrm{AT}(u \to v)\Big) \tag{2.7}$$

where $\alpha(u \to v)$ is the neighbor weight for this specific transition, $\mathcal{N}_i$ is the neighborhood (set of adjacent cells) of land cell i, s_j^t is the land use of cell j at step t, and $\mathbb{I}(s_j^t \in \mathrm{AT}(u \to v))$ is an indicator function equal to 1 if neighbor j belongs to the attracting type for the transition $u \to v$,

and 0 otherwise, so that NW_i is the number of cell i's neighbors belonging to the attracting set. This count is normalized by $\overline{NW}$, its mean over all cells currently eligible for the transition (i.e., all cells of type u), so that $\alpha(u \to v)$ and the village-proximity weight introduced below act on the same, comparable scale.

2.6.1 Approximating the Neighbor Weight

The equation above can be decomposed into two components. The constant term 1 represents a stochastic baseline: every cell of type u carries a baseline potential to transition to type v regardless of its neighborhood context. The second term, $\alpha(u \to v) \cdot NW_i/\overline{NW}$, adds a spatially contingent potential, amplifying the transition probability for cells whose neighbors already belong to the attracting type more than the eligible-cell average.

The parameter $\alpha(u \to v)$ thus controls the degree to which transitions are neighborhood-driven versus stochastic, but it is unbounded: $\alpha = 0$ yields a purely stochastic system in which neighbors have no effect, whereas a fully neighborhood-driven system requires $\alpha \to \infty$. To calibrate over a bounded domain, we instead use the *stochasticity ratio* α', defined as the ratio between the sampling weight of a cell with no attracting neighbors ($NW = 0$) and one with the mean number of attracting neighbors across eligible cells ($NW = \overline{NW}$):

$$\alpha' = \frac{w(NW = 0)}{w(NW = \overline{NW})} = \frac{1}{1 + \alpha} \in (0, 1]. \tag{2.8}$$

$\alpha' \to 1$ means neighbors are irrelevant and transitions are purely stochastic, whereas $\alpha' \to 0$ means transitions are strongly neighborhood-driven—large fitted α values, equivalently small α', therefore indicate a near-deterministic, neighbor-dependent regime. Calibrating α' is therefore essential to ensure that the simulated scenario reproduces the land-use configuration observed in the reference classification map.

To calibrate $\alpha'(u \to v)$ to the current system configuration, we use Bayesian optimization to minimize differences in the spatial summary statistics characterizing the map's structure between the original classification map and the generated base map. We employ two such measures: edge density and cross-scale lacunarity.

2.6.2 Edge Density

We derive an *edge map* from each land-use map. A cell is classified as an edge cell if at least one of its four von Neumann neighbors (up, down, left, or right; excluding diagonals) belongs to a different land-use category. The overall edge density ρ is then defined as the proportion of edge cells relative to the total number of cells. We use ρ as our quantitative index of landscape *fragmentation* during calibration, where composition is held at its observed value.

2.6.3 Lacunarity

Edge density captures local boundary structure but is insensitive to spatial organization at larger scales. To quantify cross-scale heterogeneity, we employ *lacunarity*, a scale-dependent measure of spatial variability. For a given window size r, we slide an $r \times r$ window across the edge

map (Fig. 2.5c) and record the local edge density at each window position. Let ρ_r denote the collection of local edge density values at scale r. Lacunarity at scale r is then defined in the standard Allain–Cloitre gliding-box form [29]:

$$\Lambda(r) = 1 + \frac{\mathrm{Var}(\rho_r)}{\mu(\rho_r)^2}, \tag{2.9}$$

where $\mathrm{Var}(\rho_r)$ and $\mu(\rho_r)$ are the variance and mean of the local edge densities at window size r, respectively, so that $\Lambda \geq 1$ with equality for a spatially uniform edge map.

Because Λ depends on the overall edge density of the map, we work throughout with the normalized form

$$\hat{\Lambda}(r) = \frac{\ln \Lambda(r)}{\ln \Lambda(1)}, \tag{2.10}$$

which equals 1 at $r = 1$ by construction and decays toward 0 as heterogeneity is smoothed away, allowing maps with different overall fragmentation to be compared [30, 31].

2.6.4 Calibration

We optimize the neighbor weights for each land-use transition to ensure that the simulated baseline map M_c matches the land-classification map M_b in spatial structure. Because the classification map does not resolve *tavy* or fully recovered forest as separate classes, the simulated map is first collapsed onto the observed class vocabulary—both *tavy* and fully recovered forest are relabeled as secondary forest—so that the two maps are compared over identical categories. The calibration therefore constrains the spatial structure of the mosaic, not the placement of classes the data cannot distinguish. The discrepancy between the two maps is quantified by:

$$\mathcal{D} = \sum_{r \in \mathcal{R}} \left| \hat{\Lambda}(r;\, M_c) - \hat{\Lambda}(r;\, M_b) \right| \;+\; w_\rho \, \frac{\left| E(M_c) - E(M_b) \right|}{\max\big(E(M_c),\, E(M_b)\big)}, \tag{2.11}$$

where $\hat{\Lambda}$ is the normalized lacunarity defined above, $E(\cdot)$ is the number of edge cells in a map, and $\mathcal{R} = \{2^0, 2^1, \ldots, 2^8\}$ is an exponentially spaced sequence of scales doubling up to half the smaller map dimension (nine scales, $r = 1$ to 256 cells). The edge weight $w_\rho = 0.25$ was calibrated so that the edge term contributes on average like a single scale of the lacunarity sum, about 10% of $\mathcal{D}$ at typical parameter values; cross-scale lacunarity therefore dominates the fit. We then apply Bayesian optimization to identify the parameter set that minimizes $\mathcal{D}$.

Treating all seven $\alpha'(u \to v)$ as free, independent parameters renders the optimization ill-posed: the discrepancy $\mathcal{D}$ admits many near-equivalent minima, so the fit does not converge to a unique, reproducible solution. We therefore first tied the stochasticity ratio within each of the three transition categories introduced above (Table 3.1), reducing seven free weights to three with no appreciable loss in final fitness; the calibration then showed that even three categories were more than the data could resolve, so we reduced them further to two (Results).

2.7 Adding vanilla

After modeling the current system under the *tavy*-only assumption, we examine how the system changes as market integration intensifies and the local economy becomes more reliant on

vanilla agroforestry. To generate a future scenario that includes vanilla agroforestry, we need to parameterize the different market-integration scenarios using the same parameters we defined earlier—transition demand and transition potential. For transition demand, we consider degradation and cultivation separately. For transition potential, we calibrate both the neighbor weight and a new parameter: the proximity weight. Baseline parameter values used in the calibration are summarized in Table 3.1.

2.8 Vanilla cultivation demand

We modeled the transition demand assuming that the total number of agricultural plots remains constant. For this we define the **vanilla fraction** ($B \in [0, 1]$), where $B = 0$ represents a scenario of exclusively *tavy* plots and $B = 1$ represents exclusively vanilla agroforestry, and any value in between represents some mix. The transition demand to vanilla is calculated to account for both the vanilla fraction and the vanilla degradation transition demand, ensuring the system reaches equilibrium when the combined number of *tavy* and vanilla plots is held constant, regardless of B.

To determine the number of sample points required along the vanilla-fraction gradient, we employ an adaptive sampling approach. We initialize with two points at the extremes of the gradient ($B = 0$ and $B = 1$), generate a landscape map for each, and compute the Index of Translational Homogeneity (ITH; see Section 2.10.2). We then use linear interpolation to predict the ITH at the midpoint $B = 0.5$, generate the corresponding map, and evaluate the error between the predicted and observed ITH values.

If the error exceeds a threshold of 0.01, we continue by generating maps at $B = 0.25$ and $B = 0.75$, iteratively inserting new sample points at the midpoint of each existing interval. After each insertion, we assess how well linear interpolation of the accumulated points approximates the full set, halting when the maximum approximation error falls below 0.01. This adaptive scheme ensures sufficient sampling density to capture the nonlinear behavior of the system [32], resulting in a final set of nine sample points.

2.8.0.0.1 Vanilla degradation In contrast to *tavy*, in which land degradation is intrinsic to the cycle, vanilla can, in principle, be cultivated on the same plot indefinitely. However, due to the sensitive nature of the vanilla plant and local farmers' limited access to knowledge and resources, vanilla plots tend to degrade through the spread of fungi and changes in soil composition. When vanilla plots degrade, farmers burn them to prepare the plots for *tavy*, as reported by World Bank [33]. We defined three levels of sustainability based on farmer experience and management techniques: (1) low-sustainability practices (∼15 years), (2) intermediate management (∼50 years), and (3) well-established, fully sustainable agroforestry practices (∼200 years).

2.8.0.0.2 Vanilla neighbor weight We assume that vanilla cultivation is sited like *tavy*: farmers have the same interest in clustering vanilla plots as they do *tavy* plots, so vanilla cultivation takes the fitted cultivation stochasticity ratio, $\alpha' = 0.0037$ (Table 3.1). For vanilla degradation we set a much stronger neighborhood dependence, $\alpha' = 0.001$, since vanilla degrades due to fungi (which spread from neighboring cells) and is highly sensitive to environmental conditions.

2.8.0.0.3 Village proximity (Fig. 2.3b) Vanilla is labor-intensive, and farmers must also guard against theft [33], biasing plot allocation toward locations that are easier to monitor — typically those close to the village. To test the impact of farmers' preference for cultivating vanilla near the village, we add a new factor to the transition potential equation: a village-proximity factor (see below), with proximity weight ω_{vil}. Rather than fitting ω_{vil} independently, we tie it to the fitted vanilla neighbor weight α for the same rule (Vanilla neighbor weight, above): ω_{vil} is either 0 (no proximity preference) or $\alpha/10$ (proximity preference at one-tenth the strength of the neighborhood term). Because α and ω_{vil} enter the transition potential on the same normalized scale (Combined transition potential, below), this ratio is a meaningful, comparable quantity rather than two independent constants, and the system is self-reinforcing: a plot nudged toward the village by even a weak proximity term becomes a neighbor, and the (order-of-magnitude larger) neighborhood term then recruits further plots around it, so proximity only needs to be strong enough to seed the first few plots. The 1:10 ratio was chosen by visual inspection of the resulting clustering in generated maps, not by fitting it against the lacunarity/edge-density objective that α was calibrated against.

2.8.0.0.4 Full experimental design Combining the three sources of variation introduced above—vanilla fraction, vanilla degradation rate, and village-proximity weight—yields the full experimental design. With 9, 3, and 2 values, respectively, we obtain $9 \times 3 \times 2 = 54$ scenarios (Table 2.1).

2.9 Village proximity

To represent farmers' preference for cultivating vanilla near villages, we introduce a village-proximity factor based on the distance to the nearest village. Following Decraene [34], this preference is modeled using a power-law distance-decay function:

$$\Psi_i = \omega_{\text{vil}} \cdot \frac{\psi_i}{\bar{\psi}}, \qquad \psi_i = \frac{1}{\left(d(x_i, x_{\text{village}(i)}) + 1\right)^{\gamma}}, \tag{2.12}$$

where Ψ_i is the proximity contribution to the transition potential of cell i, ω_{vil} is the village-proximity weight, and $d(x_i, x_{\text{village}(i)})$ is the Euclidean distance between cell i and the nearest village cell. The $+1$ keeps ψ_i finite at $d = 0$, the village cell itself, and $\bar{\psi}$ is the mean of ψ_i over all cells currently eligible for the transition, the same normalization applied to the neighbor term above so that ω_{vil} and $\alpha(u \to v)$ act on the same, comparable scale. Distance-decay functions of this form are commonly used in urban land-use models; however, their application to agricultural systems requires additional consideration regarding the choice of the decay exponent.

2.9.1 The distance exponent

The formulation of Decraene [34] implicitly assumes the existence of clearly defined urban boundaries, such as municipal limits. In remote agricultural systems, however, such boundaries are typically diffuse rather than explicit. Travel costs and accessibility naturally constrain farmers, but there is no sharply defined maximum cultivation distance.

To formalize this notion, we define an implicit spatial boundary through a convergence condition on the cumulative proximity contribution over an unbounded two-dimensional lattice.

Because the number of cells located at a distance d grows proportionally to $2\pi d$, the cumulative contribution of the distance-decay term scales as

$$N_{\text{cult}} \propto \int_1^\infty \frac{2\pi d}{d^\gamma}\,\mathrm{d}d = 2\pi \int_1^\infty d^{1-\gamma}\,\mathrm{d}d. \tag{2.13}$$

This integral converges if and only if $\gamma > 2$. To ensure that this convergence occurs at a realistic spatial scale, we set $\gamma = 3$. This choice guarantees that the cumulative proximity contribution remains finite.

2.9.2 Combined transition potential

By incorporating this new factor, we obtain the final transition potential as

$$\Phi_{i,\,u\to v} = 1 + \alpha(u \to v) \cdot \frac{NW_i}{\overline{NW}} + \omega_{\text{vil}} \cdot \frac{\psi_i}{\bar{\psi}} \tag{2.14}$$

where u is the current state (s_i^t) and v is the potential future state.

Table 2.1: Summary of simulation parameters used to define the 54 scenarios.

Parameter	Range	Description	Visualization
Vanilla fraction (B)	0–1	Gradient from *tavy* (subsistence) to agroforestry (cash crops).	position on the x-axis
Village proximity (ω_{vil})	$\{0, \alpha/10\}$	Preference for placing vanilla fields near the village; α is the fitted vanilla neighbor weight for the same rule.	▲ ●
Degradation Rate	$\frac{1}{15}, \frac{1}{50}, \frac{1}{200}$	Rate of land-productivity loss (1/years; characteristic timescales of 15, 50, and 200 years).	● ● ●

2.10 Analyzing Land-Use Maps

After generating future land-use maps for each scenario, we performed a comparative spatial analysis across three complementary dimensions of landscape structure. Of these, only lacunarity was also used during model calibration (Methods); composition and edge interspersion are introduced here to characterize how landscape structure responds to changes in the scenario parameters.

- **Land-use composition**, capturing the relative abundance of each land-use category;
- **Edge interspersion**, quantifying local fragmentation and land-use adjacency independently of composition;

- **Lacunarity**, characterizing cross-scale land-use configuration.

All configuration indices are computed on the dynamic mosaic only. The static classes—water, village, paddy rice, and the protected semi-intact forest—occupy an identical footprint in every scenario, so any boundary touching them contributes a constant that dilutes the scenario signal. Cells of these classes are therefore excluded from both the numerator and the denominator of every index, and boundaries between a dynamic and a static cell are not counted.

To compare land-use composition across multiple maps with diverse land-use types, we applied Principal Component Analysis (PCA) to the composition vectors. PCA identifies the dominant linear combinations of land-use types that capture the largest share of variance across maps (Fig. 2.4b). Beyond composition, we characterize spatial configuration with two complementary indices, taken up in turn below: edge interspersion, which is directly comparable across maps because it is normalized by composition; and lacunarity, which is characterized by a scale-dependent curve rather than a single value. To obtain a comparable scalar that captures the cross-scale structure described by lacunarity, we adopt the Index of Translational Homogeneity (ITH) of Malhi and Román-Cuesta [31], developed for tree-crown texture in forest-canopy imagery, and adapt it to the land-use edge map.

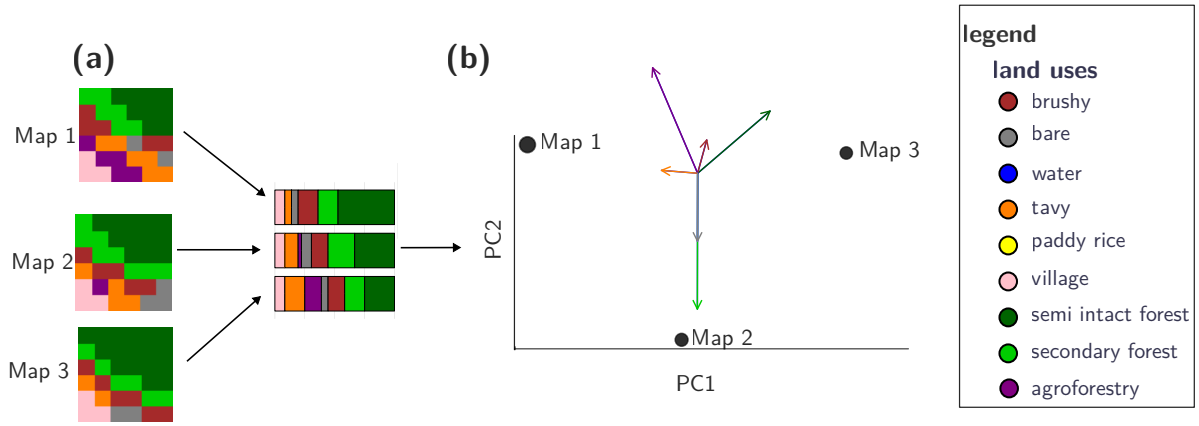


Figure 2.4: Land-use composition analysis. (a) To analyze the generated land use, we begin with land-use composition, defined as the proportion of each land-use category in a map. **(b)** To compare the composition across different maps, we conduct Principal Component Analysis (PCA) to reduce dimensionality, representing each map using only two features (PC1 and PC2).

2.10.1 Edge Interspersion

Edge density (Section 2.6.2) is not comparable across maps that differ in composition: a map split evenly among many land-use types has more opportunities for boundaries than a map dominated by one type, so ρ conflates a change in fragmentation with a change in composition. Because our scenarios deliberately move composition along the *tavy*-to-vanilla gradient, we compare configuration across scenarios using *edge interspersion* instead. Let $\mathcal{P}$ be the set of adjacent cell pairs in which both cells are dynamic, and let q_i be the proportion of dynamic cells belonging to class i. Interspersion is the observed share of unlike pairs divided by the share expected under

a random reallocation of the same class proportions:

$$\rho^* = \frac{\left|\{(a,b) \in \mathcal{P} : s_a \neq s_b\}\right| / |\mathcal{P}|}{\frac{N}{N-1}\left(1 - \sum_i q_i^2\right)}, \tag{2.15}$$

where N is the number of dynamic cells. Values near 1 indicate a spatially random arrangement, whereas $\rho^* < 1$ indicates that classes are more aggregated than expected by chance. Because the denominator absorbs the effect of composition, ρ^* isolates spatial arrangement and is comparable across maps whose composition differs. We report ρ^* for the scenario comparisons below and the raw edge density ρ for calibration (Methods), where composition is fixed at its observed value.

2.10.2 Index of Translational Homogeneity

The ITH introduced above collapses the multiscale lacunarity curve from Section 2.6.3 into a single characteristic scale: the scale at which the map attains *translational invariance*, meaning that the statistics of the pattern no longer depend on where the window is placed [31, 35]. Because $\hat{\Lambda}(r) = 0$ corresponds to $\Lambda(r) = 1$ and hence to zero variance in local edge density among windows, the scale at which the lacunarity decay reaches zero is the scale to which the map must be aggregated before it becomes homogeneous. We first describe the calculation and then illustrate its interpretation with an example.

Using the normalized lacunarity $\hat{\Lambda}(r)$ defined in Section 2.6.3, we plot $\hat{\Lambda}(r)$ against $\ln r$ over $r = 1$ to 7 cells and fit a line to the linear portion of the curve, $r = 2$ to 6 cells:

$$\hat{\Lambda}(r) \approx a \ln r + b. \tag{2.16}$$

Malhi and Román-Cuesta [31] select this window separately for each image, as the range of box sizes over which the decay is most nearly linear; we hold it fixed so that ITH values remain comparable across scenarios.

The ITH is then defined as the scale at which this line reaches zero, back-transformed from the logarithmic scale:

$$\mathrm{ITH} = \exp\left(-\frac{b}{a}\right). \tag{2.17}$$

Because ITH extrapolates the fitted line rather than tracking lacunarity out to the scale at which it vanishes, ITH can exceed the largest box size evaluated. It should therefore be read as a comparative measure of how far heterogeneity persists across scales rather than as a directly measured length. Lacunarity curves are convex, so the linear extrapolation tends to reach zero early and the resulting scales are conservative [31].

Box 1 — Reading land use indices

To understand how each of the indices captures different scales of information, we use a toy example. The maps in Figures 2.5a and 2.5b have the same composition (0.25 village, 0.25 brushy, 0.25 *tavy*, 0.25 semi-intact / fully recovered forest) but differ in edge density: Map 2 is more fragmented than Map 1,

$$\rho(\text{Map 1}) = \tfrac{48}{64} = 0.75,$$
$$\rho(\text{Map 2}) = \tfrac{56}{64} = 0.875.$$

Edge density alone, however, provides only a partial picture: the spatial distribution of fragmentation also differs between the two maps, and it is this aspect that lacunarity captures.

Fig. 2.5c shows windows of varying size applied to the edge map of each map, contrasting the window with the highest local edge density against the one with the lowest. For both maps, at window size $r = 2$, the local edge density ρ_2 spans the full range from 0 to 1. As the window size increases, this range narrows in both cases. In Map 1, which is composed of equally sized plots, the range collapses to zero at $r = 4$: every window of that size contains the same edge density, $\rho_4 = \frac{3}{4}$. This collapse is not coincidental — $r = 4$ corresponds to the size of the agricultural plots in Map 1, which are the largest geometric objects in the landscape. Once the window size exceeds the characteristic scale of the objects generating spatial heterogeneity, the averaging process smooths out all variability, effectively eliminating it.

In Map 2, by contrast, heterogeneity persists beyond $r = 4$. This occurs because the largest geometric object is not an individual plot, but rather the spatial separation between regions of large plots and regions of small plots—a higher-order structure that only becomes apparent at larger scales.

From the lacunarity equation,

$$\Lambda(r) = 1 + \frac{\mathrm{Var}(\rho_r)}{\mu(\rho_r)^2}, \tag{2.18}$$

Lacunarity increases with variance and therefore declines more slowly with increasing window size r in Map 2 than in Map 1 (Fig. 2.5d). In landscapes that are self-similar across scales, this behavior produces a characteristic three-phase pattern on log–log axes: high lacunarity at scales below the smallest self-similar object; an approximately linear decline over an intermediate range of scales; and low lacunarity at scales larger than the point where the map becomes translationally homogeneous. Fitting this intermediate, linear phase lets us estimate that scale using the ITH — the scale at which the linear interpolation of the lacunarity meets the x-axis — as illustrated in Fig. 2.5e. In Map 1, where the landscape is tiled by plots of a single size, this scale coincides with the plot size ($r = 4$); in general the ITH is larger than any single object's size, because the variation *among* objects must also be averaged away before the map reads as homogeneous.

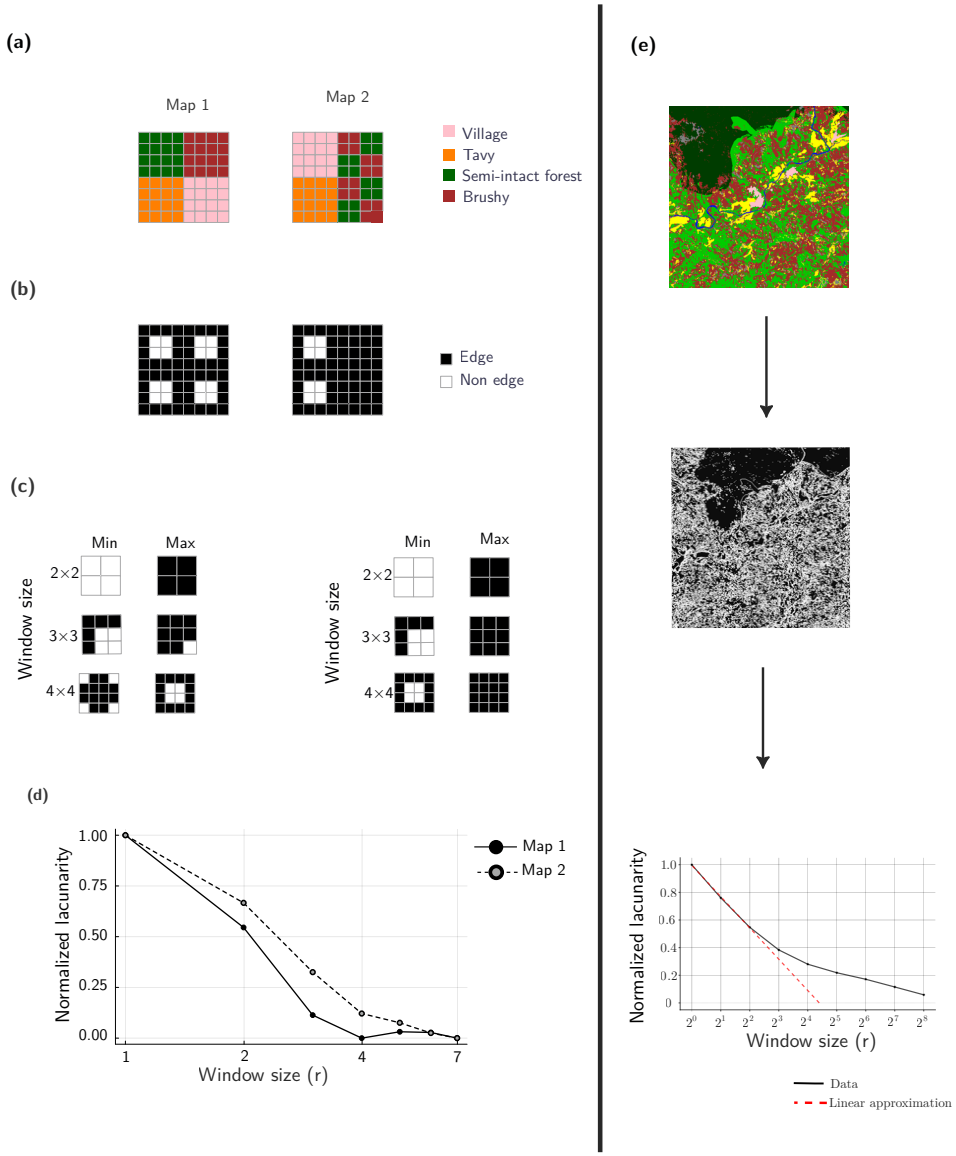


Figure 2.5: Land-use configuration analysis. Beyond land-use composition, we also examine land-use configuration. **(a)** To illustrate our configuration indices, we compare two land-use maps with identical compositions but distinct spatial arrangements: Map 1 consists of uniform 4×4 plots, whereas Map 2 contains a mixture of 4×4 and 2×2 plots. **(b)** Edge maps derived from the land-use maps, where an edge (black cell) is any pixel with at least one neighbor of a *different* land-use type. Edge density is lower in Map 1 ($\frac{3}{4}$) than in Map 2 ($\frac{7}{8}$), reflecting the negative correlation between edge density and mean plot size. **(c)** To capture higher-order structure, each edge map is subdivided into moving windows of size $r \times r$. We measure the variance in edge density across windows, which decreases as r increases and aggregation smooths spatial heterogeneity. In Map 1, heterogeneity collapses once the window reaches the plot size ($r = 4$). In Map 2, heterogeneity persists because of the larger-scale clustering of plots of different sizes. **(d)** Normalized lacunarity ($\hat{\Lambda}$) quantifies the proportion of spatial variance that persists at a given scale. Map 1 decays faster because its structure is dominated by a single characteristic scale and therefore loses variance rapidly once that scale is smoothed away; the persistence of lacunarity in Map 2 instead reflects multiscale spatial heterogeneity. **(e)** On a real land-use map, the curve typically begins with a linear section corresponding to a fractal structure; extrapolating this section to zero lacunarity yields the characteristic scale of that structure.

2.11 Tick Population Model

Configuration is not merely a descriptive endpoint: the spatial arrangement of land-use types governs the distribution of tick hosts and microclimate, and hence the opportunities for human–tick contact. Having shown that scenarios diverge in configuration even where they converge in composition, we asked whether those configurational differences propagate to a concrete, health-relevant outcome. We therefore coupled the projected landscapes to a tick population model, parameterized with data collected as part of the same broader study and in the same area from which we derived the land-use classification maps.

The model is an agent-based, four-layer lattice system in which each tick is treated as an individual agent (Fig. 2.6). The first layer represents land-use types and serves as input; the remaining three layers represent tick populations at successive life stages (larvae, nymphs, and adults). At each step, a tick can undergo one of two mutually exclusive events. In a **bite**, occurring with probability $Q(N_{i,\sigma}, K_i)$, the blood meal allows the tick to transition to the next life stage (or if it allready adult to lay eggs), and attachment to the host physically transports the tick in space; here $N_{i,\sigma}$ is the tick population at stage σ in cell i and K_i is the carrying capacity of that cell. In a **death**, occurring with the complementary probability $1 - Q(N_{i,\sigma}, K_i)$, the tick is removed from the simulation. This formulation yields a constant event rate per tick, so each iteration reduces to two steps: selecting the population and the time at which an event occurs, and then determining whether the event is a bite or a death. If the event is a bite, the new location is drawn from the movement kernel G.

2.11.1 Governing Equations

From these two events, the net change in the tick population in cell i at stage σ is the difference between the incoming and outgoing populations: the incoming population consists of ticks from neighboring cells that underwent a bite event with destination i, while the outgoing population consists of ticks in i that underwent any event. For nymphs and adults ($\sigma > 1$), the dynamics are

$$\frac{dN_{i,\sigma}}{dt} = \sum_{j \in \mathcal{N}_i} \lambda \, N_{j,\sigma-1} \, Q(N_{j,\sigma-1}, K_j) \; G(x_j, x_i) - \lambda \, N_{i,\sigma}, \tag{2.19}$$

where λ is the baseline transition rate. The first term is the influx of ticks from neighboring cells progressing from stage $\sigma - 1$; the second is the transition of ticks out of stage σ. Larvae ($\sigma = 1$) instead arise from reproduction: each adult that successfully feeds produces F eggs that develop into larvae, giving

$$\frac{dN_{i,1}}{dt} = \sum_{j \in \mathcal{N}_i} \lambda \, N_{j,3} \, Q(N_{j,3}, K_j) \; G(x_j, x_i) \, F - \lambda \, N_{i,1}, \tag{2.20}$$

where F is the average number of eggs per adult tick. Detailed definitions of Q and G, and the parameterization of F, are provided in Section A.6 of the Supplementary Materials.

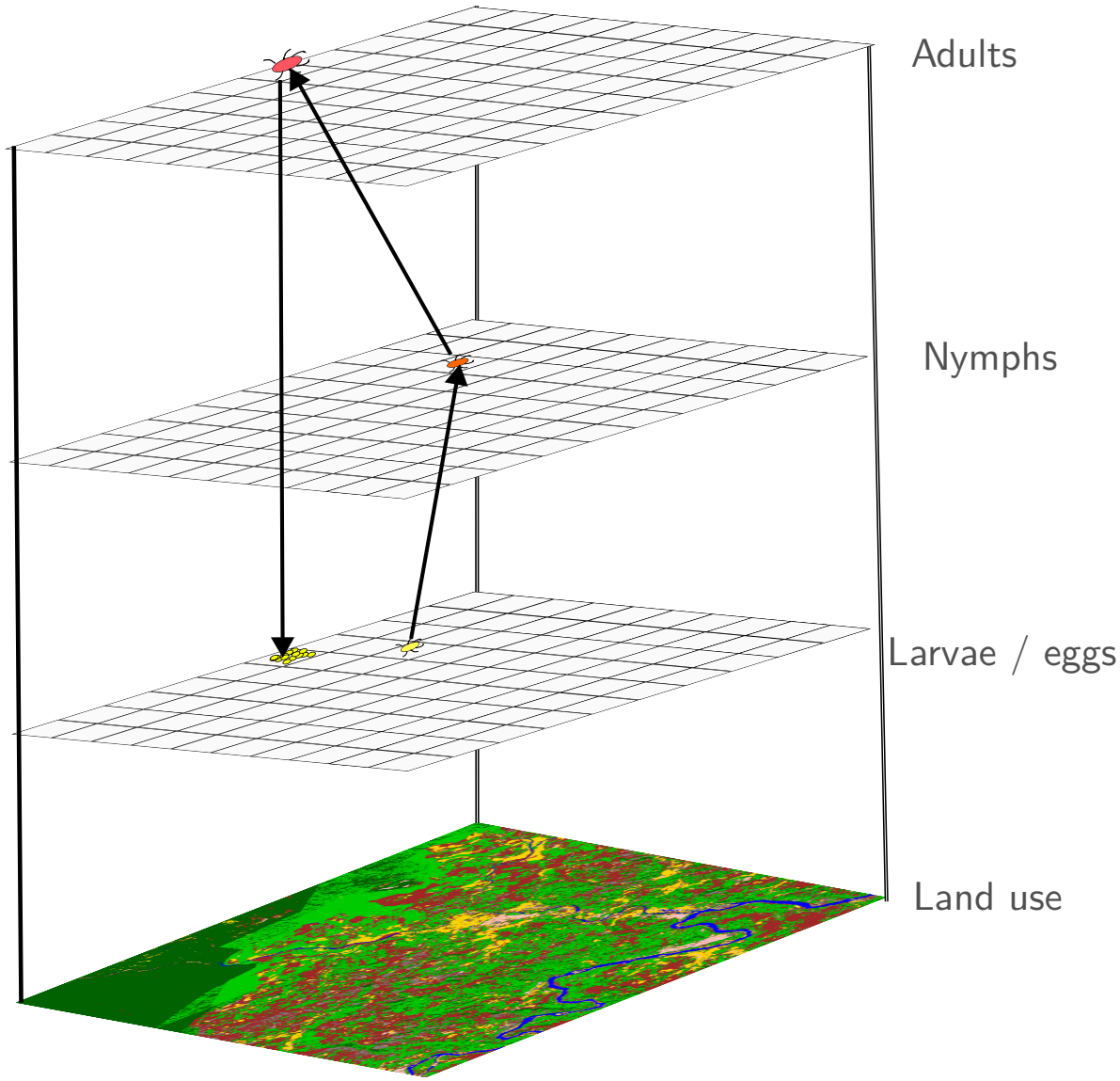


Figure 2.6: Tick population model. To demonstrate the biological relevance of the projected land-use scenarios, we applied a tick population model to estimate human exposure to ticks. The model is structured as a four-layer system: one layer represents land-use types (serving as input), and the remaining three layers represent tick populations at different life stages (larvae, nymphs, and adults). Ticks transition between life stages and move across space by biting hosts, which provides both a blood meal and physical transport.

2.11.2 Implementation and Parameterization

We implemented the simulation using the Gillespie algorithm [36], a continuous-time stochastic method, optimized as described by Lester et al. [37]. To ensure ecological relevance, we sampled small mammals and ticks across the main land-use types and pursued **qualitative fitting** rather than quantitative calibration, so that the model reflects generally observed patterns. We then verified model behavior and its sensitivity to land-use **composition and configuration** through a multi-step sensitivity analysis (Sections A.7 and A.8).

2.12 Integration of the Tick Model with Future Land-Use Scenarios

For each projected land-use map, we ran eight independent simulations of the tick population model to estimate tick **Risk**, defined as the mean tick abundance within anthropogenic cells (agricultural and village areas; see Section A.9). Preliminary analysis indicated low inter-run variability, so eight simulations sufficed to estimate Risk with a relative uncertainty below 5%.

To evaluate how landscape structure influences human exposure, we fitted linear regression models relating Risk to the four land-use indices (PC1, PC2, edge interspersion, and ITH) and their pairwise interaction terms. For each scenario map, mean Risk values were aggregated at the map level and used as the response variable; all variables were standardized. All four main effects were retained, and forward selection over the six pairwise interactions was performed using the Akaike Information Criterion (AIC) [38]. An interaction was admitted only if it lowered AIC by more than 2 [39].

2.13 Computational Implementation

All code was implemented using Julia version 1.10.4 and executed on Ben-Gurion University's high-performance computing service. The code is available at the following GitHub repository: `https://github.com/MadagascarEEID/LUC_Madagascar.git`, archived on Zenodo: `https://doi.org/10.5281/zenodo.21888470`.

Table 2.2: Summary of mathematical notation and associated units.

Symbol	Description	Unit
i, j	Indices of cells in the spatial grid	–
t	Time step	years
r	Window size in lacunarity analysis	cells
$\mathcal{L}$	Set of dynamic land-use states	–
u, v	Current and target land-use states	–
s_i^t	Land-use state of cell i at time t	–
ta, ba, br, sf, rf, af	Land-use type labels (*tavy*, bare ground, brushy regrowth, secondary forest, recovered forest, agroforestry)	–
$\mathcal{N}_i$	Neighborhood of cell i	cells
$\mathbb{I}(\cdot)$	Indicator function	–

continued on next page

Table 2.2 – continued from previous page

Symbol	Description	Unit
$\delta(u \to v)$	Transition demand (rate) from u to v	year^{-1}
T_{all}	Full transition rate matrix (product of elementary transitions)	dimensionless
$\mathbf{C}_t$	Vector of land-use proportions at time t	dimensionless
$C_t(u)$	Proportion of land-use u at time t (component of $\mathbf{C}_t$)	dimensionless
$\mathbf{C}_{\text{current}}$	Observed (current) land-use proportion vector	dimensionless
$\alpha(u \to v)$	Neighbor weight	dimensionless
$\text{AT}(u \to v)$	Attracting set for transition $u \to v$	–
$\alpha'(u \to v)$	Stochasticity ratio, $= 1/(1+\alpha) \in (0, 1]$	dimensionless
NW_i	Number of cell i's neighbors in the attracting set	cells
$\overline{NW}$	Mean NW_i over cells eligible for the transition	cells
ω_{vil}	Village-proximity weight	dimensionless
Ψ_i	Proximity contribution to the transition potential	dimensionless
$\psi_i,\ \bar{\psi}$	Normalized distance-decay term and its mean over eligible cells	dimensionless
$\Phi_{i,\,u\to v}$	Total transition potential	dimensionless
B	Vanilla fraction ($B = 0$: only *tavy*; $B = 1$: only vanilla agroforestry)	dimensionless
x_i	Spatial location of cell i	meters
$x_{\text{village}(i)}$	Location of nearest village cell	meters
$d(x_i, x_j)$	Euclidean distance between cells i and j	meters
γ	Distance-decay exponent	dimensionless
N_{cult}	Expected number of cultivated cells	cells
M_b	Land-classification map (ground-truth observed map)	–
M_c	Simulated baseline map (fitted to M_b)	–
$\mathcal{D}$	Discrepancy between M_c and M_b	dimensionless
$\mathcal{R}$	Exponential sequence of window scales $\{2^0, 2^1, \ldots, 2^k\}$	–
TVD	Total Variation Distance	dimensionless
ρ	Global edge density	dimensionless
ρ_{rand}	Edge density expected under random reallocation of the same composition	dimensionless
ρ^*	Edge interspersion, ρ/ρ_{rand}	dimensionless
ρ_r	Local edge density (window size r)	dimensionless
$\mu(\rho_r)$	Mean local edge density	dimensionless
$\text{Var}(\rho_r)$	Variance of local edge density	dimensionless
$\Lambda(r)$	Lacunarity at scale r	dimensionless
$\hat{\Lambda}(r)$	Normalized lacunarity at scale r	dimensionless
a, b	Slope and intercept of lacunarity fit	dimensionless
ITH	Index of Translational Homogeneity	meters
PC1, PC2	Principal component scores from PCA of land-use composition	dimensionless

continued on next page

Table 2.2 – continued from previous page

Symbol	**Description**	**Unit**
σ	Tick life stage (1 = larva, 2 = nymph, 3 = adult)	–
$N_{i,\sigma}$	Tick population at cell i and stage σ	ticks
K_i	Carrying capacity of cell i	ticks
$Q(N_{i,\sigma}, K_i)$	Bite probability function	dimensionless
$G(x_j, x_i)$	Tick movement kernel	dimensionless
λ	Event rate (Gillespie algorithm)	year^{-1}
F	Eggs per adult tick	ticks
Risk	Mean tick abundance in anthropogenic cells	ticks per cell

Chapter 3

Results

3.1 Calibration distinguishes only two transition categories: cultivation and natural (O1)

To ensure the model reaches equilibrium in all scenarios, we simulated 20-year intervals and measured the Total Variation Distance (TVD) of land-use composition between consecutive map pairs, recording the maximum TVD across scenarios; all scenarios reached equilibrium after 160 years (Methods). With this temporal horizon established, we fitted the simulated baseline map (M_c) to match the cross-scale lacunarity and edge density of the land-classification map (M_b), and asked how many of the model's process-level parameters this calibration can actually resolve.

For the cellular-automaton component, the transition potential of each cell is defined by its neighbors—specifically, how many of them belong to the set of land-use types that promote the transition (its attracting set, defined in the Methods). The strength of this dependency is summarized by the stochasticity ratio α', a bounded reparameterization of the neighbor weight (Methods): $\alpha' \to 1$ means that neighbors are irrelevant and transitions are purely stochastic, whereas $\alpha' \to 0$ means that transitions are strongly neighborhood-driven.

To test this, we first fitted a separate stochasticity ratio for each of the three processes—cultivation, degradation, and recovery. Bayesian optimization identified the cultivation ratio α'(cultivation) but not the degradation and recovery ratios: well-fitting solutions instead form a ridge along

$$\alpha'_{\text{rec}} \approx 0.094 - 0.15\,\alpha'_{\text{deg}}, \tag{3.1}$$

along which the differences in fitness between parameter sets are so small that, once the stochasticity of the simulation is accounted for, the exact optimum is non-identifiable (Fig. 3.1a). This indicates that the two configuration indices can distinguish only two transition categories—cultivation and natural (degradation and recovery combined).

Under this two-category assumption the solution becomes fully identifiable (Fig. 3.1b), with best-fit values α'(cultivation) = 0.0037 and α'(natural) = 0.0813. Parameter sets within $0 < \alpha'$(cultivation) < 0.0159 and $0.0727 < \alpha'$(natural) < 0.096 differ from the best solution by less than one standard deviation in fitness.

The optimization reduced the discrepancy between the indices of the two maps from $\mathcal{D} =$

1.34 ± 0.38, for parameter sets drawn at random from the prior, to $\mathcal{D} = 0.047 \pm 0.019$ at the optimum. By construction, matching these indices is expected, since the model was calibrated to them; it is therefore not evidence of predictive accuracy, but only confirms that the model behaves as intended. A more important question is how well a model fitted on edge density and lacunarity captures the "texture"—the family of patterns recognizable by eye. Placing the generated map and the classification map side by side (Fig. 3.1c) reveals both how much of this texture the two indices capture and where their limitations lie. Using only two indices, we already reproduce many aspects of the texture; future research could incorporate additional indices to improve the fit.

Overall, the configuration pattern captured by lacunarity and edge density is controlled almost entirely by two model parameters: the cultivation stochasticity, which is very low, and the natural-transition stochasticity, which is roughly twenty-fold higher ($\alpha' = 0.0813$ versus 0.0037). It is important to note that this parameter governs only the spatial distribution of transitions—determining *which* cells change—while the total number of transitions is captured entirely by the transition demand.

Table 3.1: Baseline parameters. The seven land-use transitions are grouped into three process-level categories, each defined by a shared attracting set. Demands were derived from the stationary distribution of the Markov chain and attracting sets were defined from ecological processes; the stochasticity ratios α' were approximated by Bayesian optimization to match the current land configuration, with a single α' fitted per category. Degradation and recovery are not separately identifiable and are therefore fitted jointly as one *natural* category (Fig. 3.1).

Category	Transition	Attracting set	Demand (δ)	Stochasticity ratio (α')
Cultivation	secondary forest → *tavy*	*tavy*	0.11	0.0037
	fully recovered forest → *tavy*	*tavy*	0.11	
	brushy → *tavy*	*tavy*	0.02	
Degradation	*tavy* → bare	bare	0.5	0.0813 (natural)
Recovery	bare → brushy	secondary, semi-intact, fully recovered forest	0.5	
	brushy → secondary forest	semi-intact, fully recovered forest	0.0833	
	secondary forest → fully recovered forest	semi-intact	0.02	

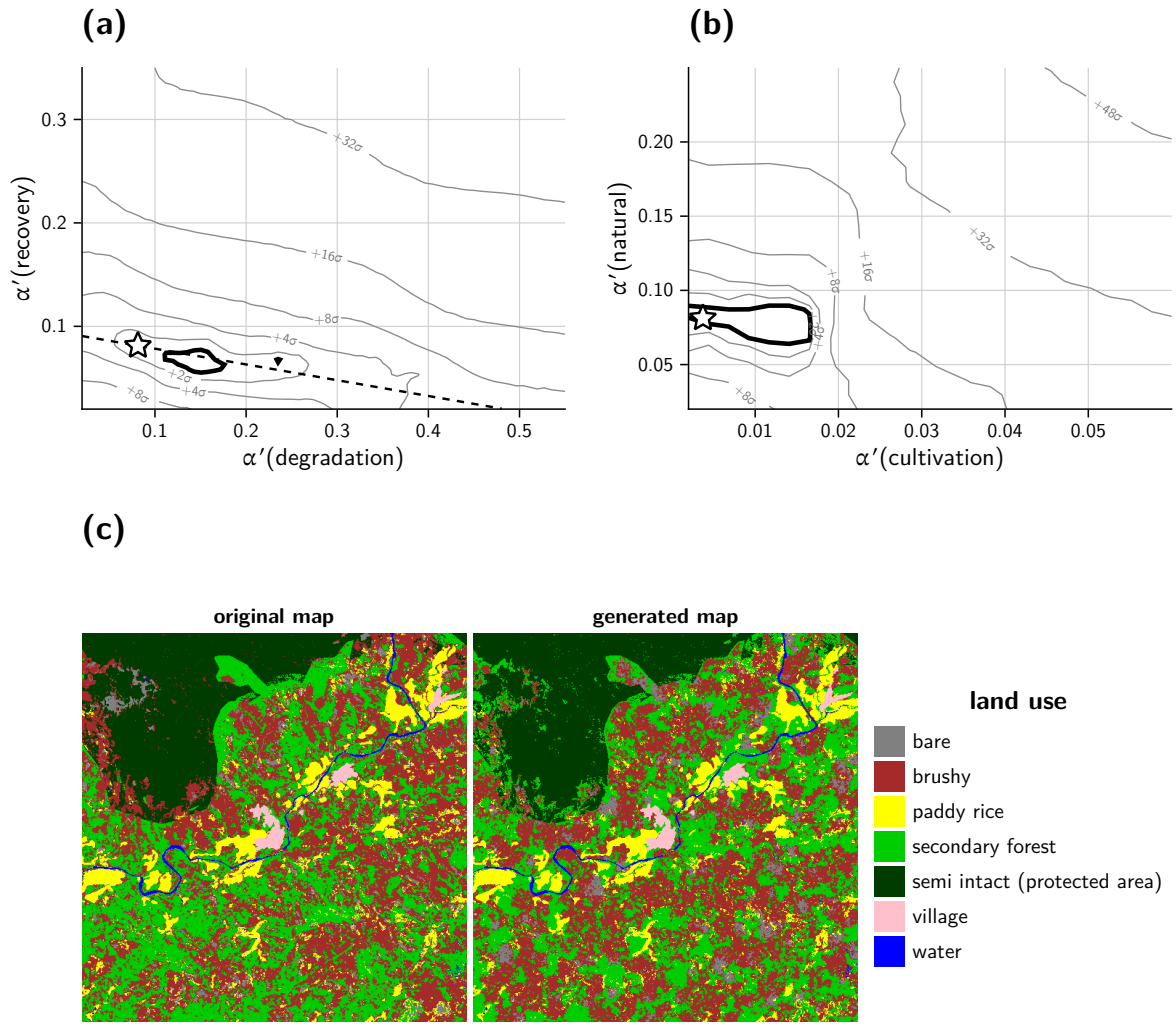


Figure 3.1: Simulation baseline calibration. **(a)** Fitness landscape over the degradation and recovery stochasticity ratios, α'(degradation) and α'(recovery), with the cultivation ratio held fixed; contour lines show fitness in standard deviations (σ) above the optimum. The gradient is far shallower along the dashed ridge (equation in text) than across it, so the optimum along the ridge is non-identifiable; the star marks the adopted representative point. **(b)** Fitness landscape under the two-category assumption, over the natural and cultivation stochasticity ratios α'(natural) and α'(cultivation); here the optimum (star) is well constrained. **(c)** Side-by-side visual comparison of the simulated baseline map and the land-classification map. Their similarity indicates that the model reproduces the original map given the fitted parameters.

3.2 Fast degradation overrides farmers' siting preferences for vanilla plots (O2)

From the simulated baseline map, we generate future land-use maps for market-integration scenarios (Fig. 3.2). They vary along three dimensions: the vanilla fraction B; the rate of vanilla plot degradation, set to fast (15-year), intermediate (50-year), or slow (200-year) characteristic timescales; and village proximity, representing farmers' bias toward planting vanilla near the village. Combining 9 vanilla-fraction levels, 3 degradation rates, and 2 village-proximity settings gives $9 \times 3 \times 2 = 54$ scenarios (Table 2.1), spanning a broad range of landscape outcomes under varying market integration, degradation, and village influence.

We compare these scenarios quantitatively in the following sections, but the main pattern is already visible in the maps. At $B = 0$ the scenarios are identical by construction, since they differ only in their assumptions about vanilla; the visible differences among these maps therefore measure simulation stochasticity rather than any scenario effect. As B increases, the progressive addition of vanilla agroforestry reshapes the landscape in ways that depend on both the degradation rate and village proximity.

At $B = 1$ the two parameters interact. Under fast degradation, vanilla plots are dispersed, because the continual search for new arable land conflicts with the tendency to aggregate plots. Under village proximity, vanilla stays concentrated near the village only when degradation is slow; when it is fast, the village becomes surrounded by degraded land, so although farmers site their first vanilla fields close to home, the plots ultimately end up spread across the landscape.

The broader pattern is that fast degradation makes plot placement a search for arable land, whereas slow degradation lets farmers' siting preferences—clustering and village proximity—govern where vanilla ends up.

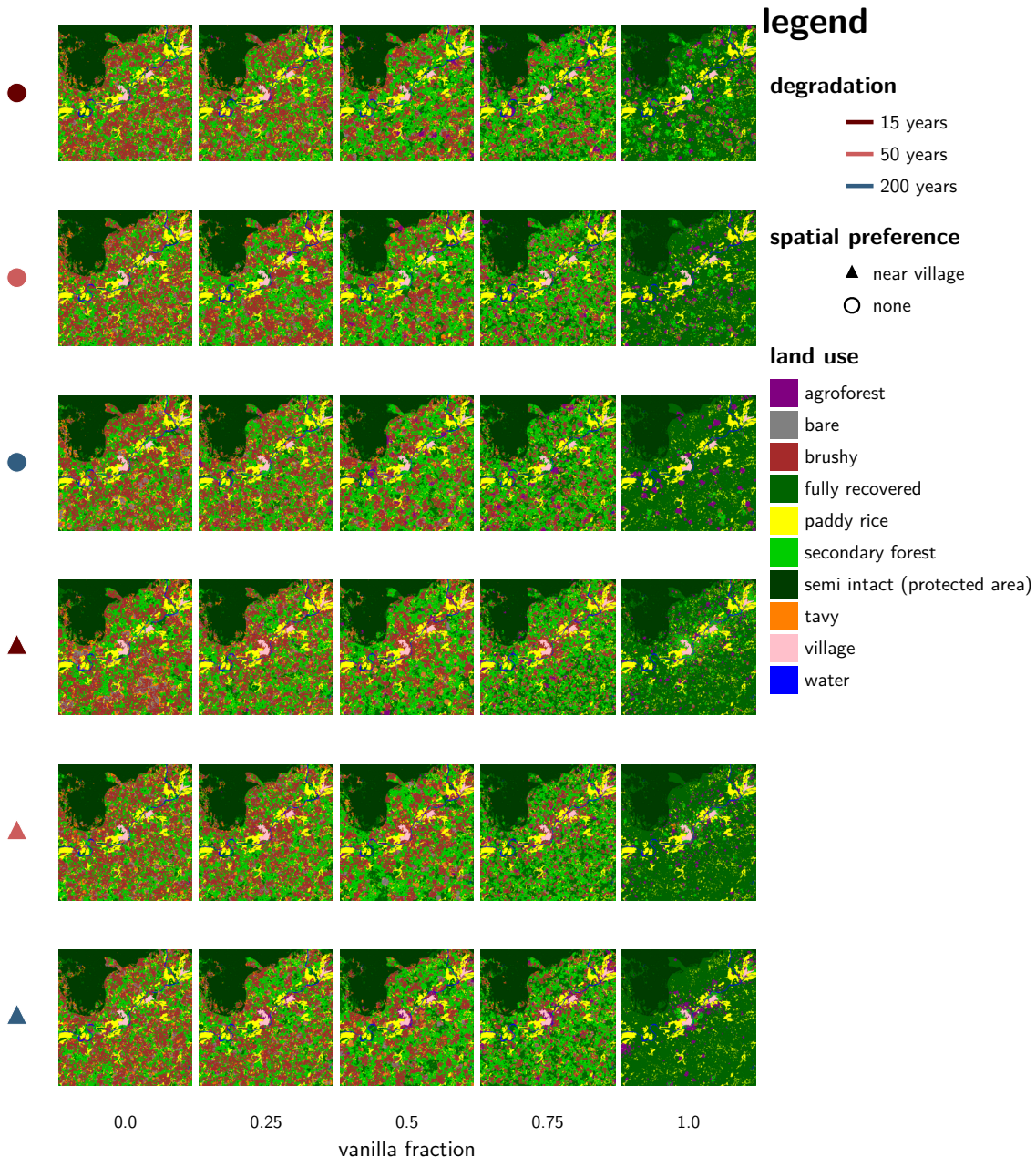


Figure 3.2: Grid of generated land-use maps across scenarios. Columns vary the vanilla fraction B from 0 (only *tavy*) to 1 (only vanilla agroforestry); rows vary the combination of vanilla degradation rate—fast (15-year), intermediate (50-year), or slow (200-year)—and village proximity. Shown are the 30 scenarios that combine all six degradation–proximity settings with five evenly spaced values of B $(0, 0.25, 0.5, 0.75, 1)$; the remaining four B values $(0.125,\ 0.375,\ 0.625,\ 0.875)$ are omitted for legibility.

3.3 Composition varies along two axes: the *tavy*-to-vanilla transformation and mid-successional cover (O3)

To compare land-use composition across maps, we performed PCA on the proportions of different land-use classes. We treated land-use types as features and maps as observations. The first two principal components jointly accounted for nearly all the variance: PC1 accounted for 93.7% and PC2 for 6.3%.

PC1 represents the overall transformation from *tavy* to vanilla agroforestry. It loads positively and most strongly on fully recovered forest, and negatively on brushy regrowth and secondary forest; agroforestry loads positively but weakly, so PC1 tracks the transformation mainly through the forest recovery that accompanies it (Fig. 3.3a). Consistent with this interpretation, PC1 increases monotonically with B, and most steeply when vanilla degradation is slowest (200-year timescale; Fig. 3.3b).

PC2 is positively associated with *tavy* and early degradation states (bare ground and brushy regrowth), as well as with fully recovered forest (Fig. 3.3a), and is therefore negatively associated with intermediate successional stages. As a result, PC2 reaches its most negative values in the upper-intermediate range ($B \approx 0.625$–0.875; Fig. 3.3c).

Village proximity, by contrast, does not affect land-use composition (overlapping triangles and circles in Fig. 3.3b,c). This result follows directly from the model's structure: composition is governed by the Markov chain and configuration by the cellular automaton, so configuration parameters cannot affect compositional outcomes. Village proximity therefore acts only on configuration, which we turn to next.

Together, the two axes divide the transformation into three phases. While the landscape is dominated by *tavy*, degraded land prevails (bare ground and brushy regrowth). In the intermediate phase, where *tavy* and agroforestry are mixed, secondary forest increases markedly. Finally, as agroforestry becomes the dominant crop, fully recovered forest accumulates.

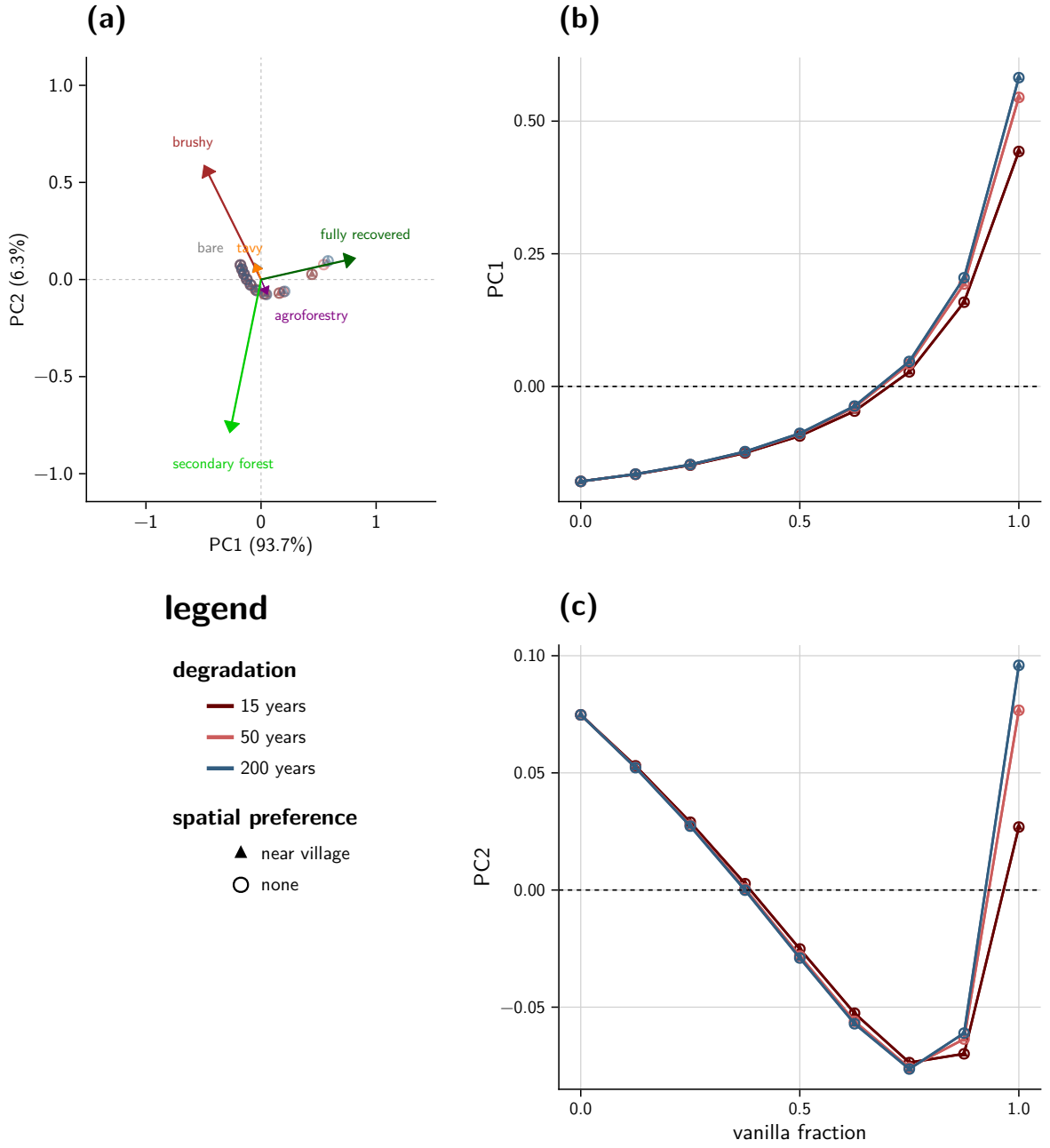


Figure 3.3: Principal component analysis (PCA) of land-use composition across scenarios. PCA was performed with land-use types as features and maps as observations, the values being the proportion of each land use in each map. **(a)** Biplot of PC1 and PC2 loadings for the land-use classes, with dots representing the different scenarios. PC1 represents the overall transformation from *tavy* to agroforestry: it loads strongly and positively on fully recovered forest, against brushy regrowth and secondary forest, with agroforestry loading positively but weakly. PC2 loads positively on brushy regrowth, bare ground, *tavy*, and fully recovered forest, and negatively on secondary forest, and therefore separates the intermediate successional stages from everything else. **(b)** PC1, which explains 93.7% of the variance, as a function of the vanilla fraction B, grouped by vanilla degradation rate and village proximity; each dot represents the PC1 value of one map. PC1 increases monotonically with B, most steeply when vanilla degradation is slowest (200-year timescale). **(c)** PC2, which explains 6.3% of the variance, plotted against B for the same scenario groupings. PC2 reaches its most negative values in the upper-intermediate range ($B \approx 0.625$–0.875), reflecting suppressed mid-successional stages. Triangles indicate scenarios with village proximity, and circles indicate scenarios without village proximity. Their overlap in **(b)** and **(c)** indicates that village proximity does not affect land-use composition.

3.4 Configuration indices respond to vanilla management only at high vanilla fractions (O3)

Composition tells us which land uses occupy the map but not how they are arranged. We therefore quantified configuration with two complementary indices: the Index of Translational Homogeneity (ITH), derived from lacunarity, which captures cross-scale land structure; and edge interspersion, the observed share of unlike adjacent cell pairs divided by the share expected under a random rearrangement of the same composition, which isolates fragmentation from composition (Methods). Both indices are computed from the edge map, so both are *class-agnostic*: they register that two adjacent cells differ, but not which land-use classes they are.

Edge interspersion ranged from 0.28 to 0.435, meaning the maps contain only about a third of the unlike-class boundaries expected under a random rearrangement of the same composition. At $B = 0$, where all scenarios are identical by construction, edge interspersion still varied between 0.375 and 0.435—a spread covering about 40% of the total range observed across all scenarios, which sets the noise floor against which scenario effects must be judged. Differences beyond that floor emerge only at $B = 1$, where *tavy* is absent entirely: there, both fast vanilla degradation and village proximity raise interspersion (Fig. 3.4a). Both parameters add competing pressures on plot siting—the continual search for arable land and the pull toward the village—which override the tendency to cluster plots and so leave more boundaries between unlike classes.

ITH remained almost perfectly constant at 12 cells (120 m) for B between 0 and 0.5. A small between-scenario variation appears between 0.5 and 0.75, and once vanilla dominates, ITH jumps sharply, spanning 47 to 209 cells (0.5–2.1 km) across scenarios—up to roughly 40% of the map extent (Fig. 3.4b). This reveals two regimes. In the first, a mosaic driven by *tavy* or by a *tavy*–vanilla mix, the relatively small plot size sets the pace at which aggregation erases heterogeneity, and a window barely over a hundred meters across already averages the landscape out. In the second, once vanilla greatly exceeds *tavy* in area, heterogeneity survives aggregation to kilometer scales, because vanilla plots are dispersed across the landscape and variation persists between whole management zones rather than only between neighboring plots.

In summary, landscape configuration differs appreciably with vanilla degradation rate and siting preferences only once vanilla clearly dominates the landscape. As long as *tavy* remains common, its highly fragmented, small-scale mosaic governs the configuration of the whole landscape.

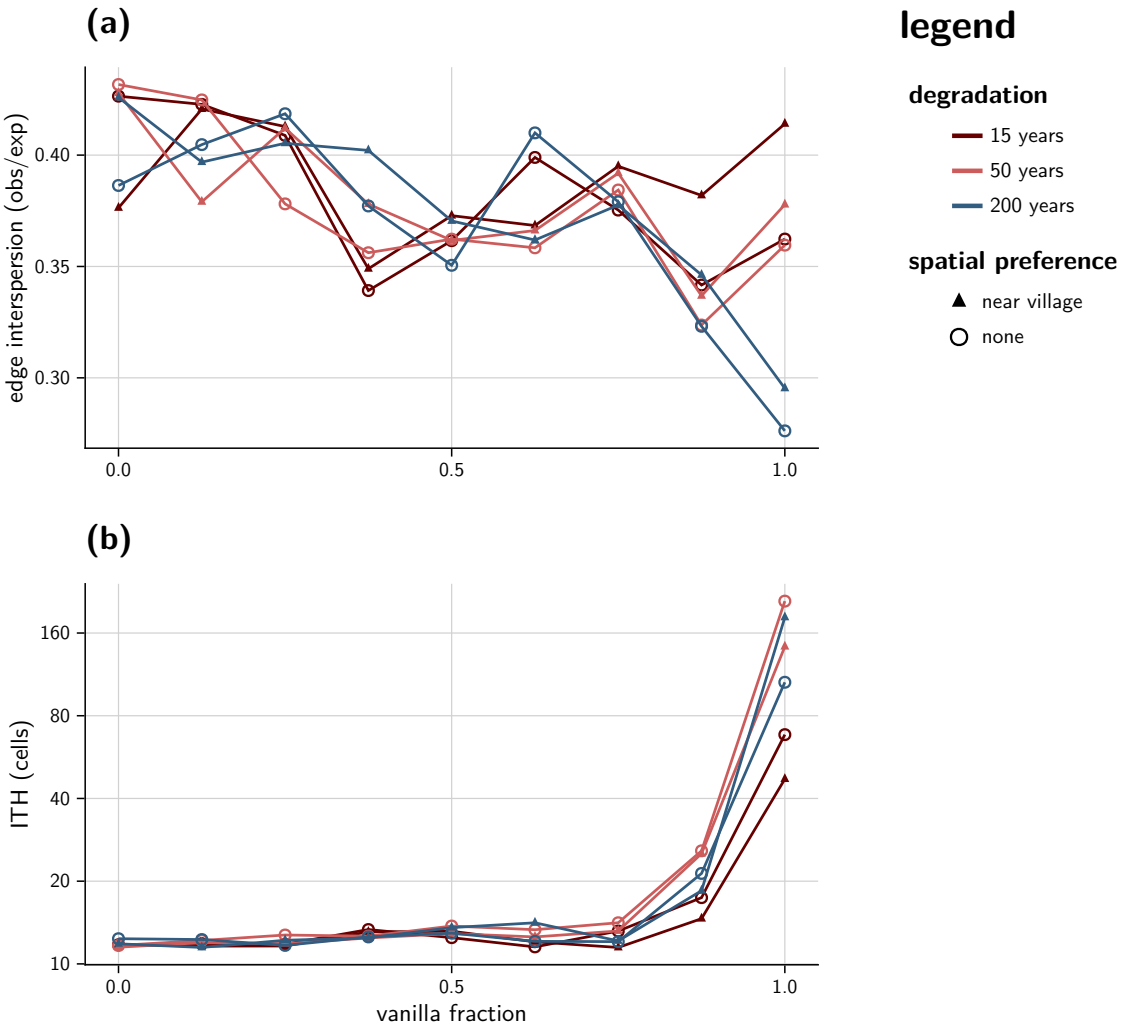


Figure 3.4: Landscape configuration across scenarios. (a) Edge interspersion as a function of the vanilla fraction B, grouped by vanilla degradation rate and village proximity. Scenarios separate only at $B = 1$, where fast degradation and village proximity both raise interspersion; the spread among the identical $B = 0$ scenarios indicates the noise floor. **(b)** Index of Translational Homogeneity (ITH) plotted analogously. ITH is flat at 12 cells (120 m) up to $B = 0.5$ and rises sharply once vanilla dominates, spanning 47–209 cells (0.5–2.1 km) at $B = 1$.

3.5 Tick Risk responds to land-use identity, which class-agnostic configuration indices cannot capture (O4)

To demonstrate a use case for the projected maps, we applied the tick population model to each of the 54 landscapes and estimated tick **Risk** (eight replicate runs per map; mean relative uncertainty below 5%, Methods).

Risk does not change monotonically along the transformation. It rises through the partial-transformation range and then falls once vanilla replaces *tavy* entirely, so the highest exposures occur in landscapes that are mixed rather than in either endpoint (Fig. 3.5a). At $B = 1$ the residual differences are set by management: Risk remains 2.5–5 times higher under fast (15-year) vanilla degradation than under intermediate or slow degradation at the same proximity setting, because degrading vanilla plots return to the *tavy* cycle and regenerate the mid-successional habitat that supports ticks.

Tick Risk resolved scenario differences that the landscape indices did not. Composition is insensitive to village proximity by construction, and the configuration indices separated the proximity scenarios only at $B = 1$; tick Risk, by contrast, separated them across almost the entire gradient. From $B = 0.125$ to $B = 0.875$, Risk in the village-proximity scenarios lay below its no-proximity counterpart in 20 of the 21 scenario pairs, whereas at $B = 1$ it lay above in all three (Fig. 3.5a). The reason is that our configuration indices are *class-agnostic*: computed from the edge map, they treat all land-use classes as interchangeable and count every boundary equally. They are also computed on the dynamic mosaic alone, so boundaries with the village—the class where human exposure concentrates—are excluded outright. The tick model is neither. It responds strongly to specific adjacencies—for example, a boundary between secondary forest (the most suitable tick habitat) and the village (where people spend most of their time)—and weakly to others, such as a boundary between brushy regrowth and bare ground, where neither ticks nor people concentrate.

Even so, the landscape indices retain substantial explanatory power: a linear model on PC1, PC2, edge interspersion, and ITH explained 81% of the variance in simulated tick Risk (adjusted $R^2 = 0.80$). Beyond the four main effects, which were retained throughout, AIC-based selection retained a single interaction, PC1 × edge interspersion (Fig. 3.5b). PC1, PC2, and edge interspersion were each individually significant ($p < 0.002$), as was the interaction ($p = 0.007$); ITH was not ($p = 0.62$), indicating that once composition (PC1) is accounted for, ITH adds no detectable explanatory power of its own. The interaction PC1 × edge interspersion is positive while both main effects are negative, and the two ends of the PC1 axis explain why. Where PC1 is low, the landscape holds many degradation stages at once, so most boundaries separate land types that differ only in tick suitability and not in how often people visit them; additional edges there mainly mix suitable with unsuitable habitat, depressing overall abundance. Where PC1 is high, the landscape collapses to two types—fully recovered forest and agroforestry—so most boundaries separate tick-rich, people-poor land from people-rich, tick-poor land, and additional edges increase spillover into occupied cells.

It is important to note, however, that this regression characterizes a relationship within our coupled modeling system rather than an empirically validated dose–response. Because all 54 landscapes derive from three scenario parameters, the configuration indices represent a struc-

tured, low-dimensional design space rather than independent samples—and so the analysis does not establish the predictive strength of these indices against field-measured tick burdens. It instead serves as a proof of concept for coupling the framework to epidemiological models. A further caveat is that we estimate Risk as the mean tick abundance in the cells people occupy, whereas the co-occurrence of people and ticks in the same place is only one of the factors that determine the probability of a bite.

Overall, the relationship between the *tavy*-to-vanilla transformation and tick Risk is nonlinear and stage-dependent, and it is only partly captured by the landscape indices: composition and configuration together account for most of the variance, but the residual reflects adjacencies whose importance depends on which land uses meet at the boundary—information that class-agnostic indices discard by design.

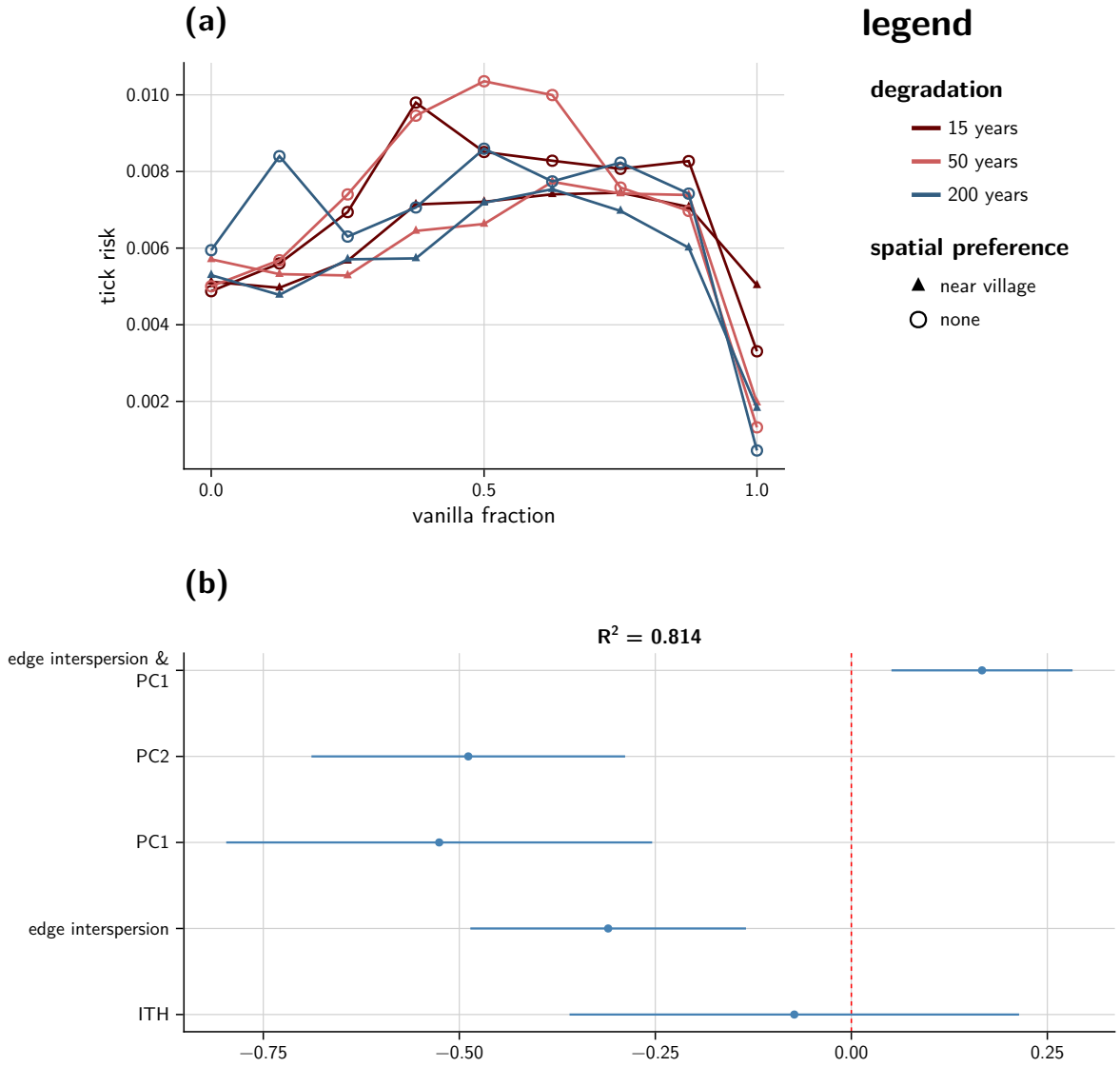


Figure 3.5: Tick Risk across scenarios. **(a)** Modeled tick Risk as a function of the vanilla fraction B, grouped by vanilla degradation rate and village proximity; each point is the mean of eight replicate simulations of one landscape. Risk rises during partial transformations and falls under full transformation ($B = 1$), where it nonetheless remains 2.5–5 times higher under fast (15-year) degradation than under intermediate (50-year) or slow (200-year) degradation at the same proximity setting. Village proximity lowers Risk across almost the whole partial-transformation range ($B = 0.125$–0.875) but raises it in all three degradation scenarios at $B = 1$—a separation of the proximity scenarios that the class-agnostic landscape indices do not resolve. **(b)** Coefficients of the linear regression of Risk on the land-use indices (PC1, PC2, edge interspersion, and ITH) together with the interaction term retained by AIC-based model selection, PC1 × edge interspersion ($R^2 = 0.81$, adjusted $R^2 = 0.80$); bars show 95% confidence intervals. PC1, PC2, and edge interspersion are each individually significant ($p < 0.002$), as is the interaction ($p = 0.007$), whereas the ITH interval overlaps zero ($p = 0.62$).

Chapter 4

Discussion

Land-use change models have been widely used to examine directional change (land-use systems where cells that change from one land use to another have a low probability of returning to their first state) such as urban expansion and modern agricultural intensification [5, 7, 8]. In contrast, dynamic traditional agricultural systems, including shifting cultivation and agroforestry, remain understudied, despite their importance for food security, livelihoods, and cultural continuity in many tropical rural communities [9].

Modeling dynamic, small-scale land-use systems requires rethinking what a model is expected to predict. In directional systems, models typically aim to predict the future land-use class of each cell. In shifting cultivation, however, the state of any given cell at a specific point in time is partly stochastic. Model evaluation therefore cannot rely solely on cell-by-cell agreement. Instead, it should assess whether simulated landscapes reproduce the broader spatial patterns that characterize the system[40].

We addressed this challenge by developing a Markov chain–cellular automata model designed for dynamic, shifting-cultivation land-use change: the model iterates repeatedly, transitioning each cell through multiple shifting-cultivation cycles until land-use composition reaches equilibrium, and it is calibrated against two complementary configuration indices spanning scales—edge density, capturing local interaction, and cross-scale lacunarity, capturing higher-order spatial structure. The calibration also bounds how much process detail these indices can support: the edge map separates only two sources of stochasticity, natural transition and cultivation, rather than the three we hypothesized, and resolving additional processes requires richer calibration targets, not more parameters.

By fitting the cellular automata ruleset to these indices, the model successfully reproduced key properties of the observed land-use map across scales, in contrast to earlier approaches that were able to do so only at a local scale [11] or a coarse scale [19]. The indices also varied across projected scenarios in ways consistent with their agroeconomic assumptions, indicating that the framework is sensitive to meaningful changes in land-use dynamics. Importantly, the indices also helped predict tick exposure, demonstrating their potential relevance to real-world ecological and public-health outcomes. The fact that tick exposure depended not only on the individual indices but also on an interaction between composition and fragmentation indicates that modeling only one aspect of land-use structure yields an incomplete picture.

The framework (model and its associated estimation indices) we developed is relevant for understanding the resilience of shifting cultivation systems. In many places, population pressure and economic intensification can shorten fallow periods and reduce the time available for natural succession [12, 41]. This may lead to soil degradation, declining productivity, and reduced availability of productive land, potentially creating a reinforcing cycle of further degradation and forest loss [12, 42, 43]. Linking our framework with models of land degradation, water runoff, or crop productivity could help identify thresholds at which these systems lose resilience.

Although the model was fitted to a specific region in northeastern Madagascar, the framework is transferable. It requires only a partial land-use classification map and qualitative prior knowledge of the system, as the remaining transition structure can be inferred through Markov chain calculations and Bayesian optimization. This flexibility is particularly valuable for modeling tropical and underdeveloped regions, where comprehensive land-use classification and detailed time-series data are often difficult to obtain. The ability to operate under data-scarce conditions allows the framework to establish a baseline understanding of land-use dynamics in regions where full data are unavailable. As more accurate maps and richer time-series datasets become available in the future, they can be incorporated to validate and refine model predictions — naturally improving accuracy as the available evidence grows.

A further contribution of this work is the link it draws between the fractal properties of land-use patterns and the dynamic processes that generate them; lacunarity and the ITH derived from it are the two such properties we use here. We showed that a cellular automaton can reproduce these properties from interpretable process-based rules, and that the properties respond to agricultural assumptions, consistent with earlier findings on the value of pattern-oriented modeling for capturing such complex, uncertain land-use systems [40]. A natural next step is to relax the class-agnosticism of these indices. Because they are computed from the edge map, ITH and edge interspersion transfer between systems without modification, but they also discard the identity of the land uses that meet at each boundary—precisely the information our coupled tick model responds to. A contrast-weighted index, which weights each boundary by the difference between the land uses it separates, would likely raise predictive power for a specific outcome such as disease exposure, at the cost of generality: it must be designed anew for each system and each outcome. This trade-off reflects a broader, well-recognized limitation of general configuration indices: their difficulty capturing the pattern information relevant to a specific biological process [44], which has led researchers to identify the linking of landscape metrics to ecological processes as a standing research priority [45]. Which side of this trade-off to prefer depends on whether the goal is transferable description or targeted prediction.

This research has several limitations that could be addressed in future work. First, the model was fitted to the current land-use map under an equilibrium assumption. In reality, land-use systems are rarely at equilibrium. Future models can use historical land-use maps to fit the model to observed dynamics rather than to a single state. Multiple temporal maps could also be used to validate projected outcomes [6]. Second, natural succession in *tavy* systems depends strongly on burning history [12], whereas the current model assumes constant recovery rates. Future models could incorporate burning history directly into the transition rules [19]. Third, degradation and recovery may not follow a simple Markov process, because transition probabilities can depend on the time since the last transition. Semi-Markov models could better

represent these duration-dependent processes [46]. Finally, the tick exposure model was built using limited empirical data, and some parameters were based on general ecological patterns rather than local measurements.

In conclusion, classical land-use change models are often designed to predict the specific land-use class of a specific cell at a specific time. Here, we demonstrate an alternative approach: fitting models to multiscale spatial indices rather than exact cell-level outcomes. This framework allows useful predictions in traditional, dynamic agricultural systems where cell-level land use is stochastic but broader spatial structure remains meaningful. Developing models that capture this structural complexity is an important step toward better policy decisions in landscapes shaped by tightly intertwined human and ecological processes, and toward more sustainable land-use transitions in regions of global importance for biodiversity, climate regulation, and disease dynamics.

Appendix A

Supplementary Materials

A.1 Code implementation

The code is available at `https://github.com/MadagascarEEID/LUC_Madagascar.git` (archived on Zenodo: `https://doi.org/10.5281/zenodo.21888470`); for guidance in reading or understanding the code, refer to the README file.

A.2 Fitting baseline transition demand in the cellular automata

In the manuscript, land-use demand is formally represented through matrix multiplication. While compact, this representation requires multiplying a sequence of sparse 6×6 matrices. Because most entries are zero, the same transition structure can be written transparently as a system of equations.

A.2.0.0.1 Notation We write $P(x, \tau)$ for the proportion of the landscape in land-use class x after step τ of a single iteration, and abbreviate the initial proportion as $P(x) \equiv P(x, 0)$. The within-iteration transitions are applied in the fixed order

$$br \to t, \quad sf \to t, \quad rf \to t \qquad \text{(cultivation, step 4)},$$

$$t \to ba, \quad ba \to br, \quad br \to sf, \quad sf \to rf \qquad \text{(degradation and recovery, steps 5–8)},$$

where t is *tavy*, ba bare land, br brushy regrowth, sf secondary forest, and rf recovered forest. Thus $P(t, 4)$, for example, is the amount of *tavy* present once the three cultivation transitions have been applied, before the degradation transition $t \to ba$ reduces it to $P(t, 5)$.

A.2.0.0.2 Cultivation step ($\tau = 4$)

$$P(t, 4) = P(t) + P(br)\delta(br, t) + P(sf)\delta(sf, t) + P(rf)\delta(rf, t), \tag{A.1}$$

$$P(ba, 4) = P(ba), \tag{A.2}$$

$$P(br, 4) = P(br)(1 - \delta(br, t)), \tag{A.3}$$

$$P(sf, 4) = P(sf)(1 - \delta(sf, t)), \tag{A.4}$$

$$P(rf, 4) = P(rf)(1 - \delta(rf, t)). \tag{A.5}$$

A.2.0.0.3 Degradation and recovery steps

$$P(t,5) = \tfrac{1}{2}P(t,4), \tag{A.6}$$
$$P(ba,5) = P(ba,4) + \tfrac{1}{2}P(t,4), \tag{A.7}$$
$$P(ba,6) = \tfrac{1}{2}P(ba,5), \tag{A.8}$$
$$P(br,6) = P(br,4) + \tfrac{1}{2}P(ba,5), \tag{A.9}$$
$$P(br,7) = \tfrac{11}{12}P(br,6), \tag{A.10}$$
$$P(sf,7) = P(sf,4) + \tfrac{1}{12}P(br,6), \tag{A.11}$$
$$P(sf,8) = \tfrac{49}{50}P(sf,7), \tag{A.12}$$
$$P(rf,8) = P(rf,4) + \tfrac{1}{50}P(sf,7). \tag{A.13}$$

The recovery transition $sf \to rf$ occurs with probability 1/50. We assume that secondary and recovered forest are equally susceptible to cultivation:

$$\delta(rf,t) = \delta(sf,t).$$

A.2.0.0.4 Stationary equilibrium We assume the system is in equilibrium; as a result, the value at the end of every iteration is the same as at the beginning, and we can write:

$$P(t) = P(t,5), \quad P(ba) = P(ba,6), \quad P(br) = P(br,7), \quad P(sf) = P(sf,8), \quad P(rf) = P(rf,8). \tag{A.14}$$

Equilibrium implications

Imposing stationarity on each class and eliminating the intermediate states gives, in turn, $P(t,4) = 2P(t)$ (from $P(t) = P(t,5)$) and $P(ba) = P(t)$ (from $P(ba) = P(ba,6)$ with $P(ba,4) = P(ba)$). Substituting these into the brush and secondary-forest balances ($P(br) = P(br,7)$ and $P(sf) = P(sf,8)$) and solving for the cultivation demands yields the closed forms

$$\boxed{\delta(br,t) = \frac{P(ba)}{P(br)} - \frac{1}{11}}, \qquad \boxed{\delta(sf,t) = \frac{P(br)}{11P(sf)} - \frac{1}{49}}.$$

Finally, the cultivation balance

$$P(t) = P(br)\delta(br,t) + P(sf)\delta(sf,t) + P(rf)\delta(rf,t),$$

together with $\delta(rf,t) = \delta(sf,t)$, reduces to

$$P(t) = P(br)\delta(br,t) + \big(P(sf) + P(rf)\big)\delta(sf,t).$$

A.3 Baseline calibration with recovery using an aggregated forest class

We now insert the observed proportions into the closed forms derived in Section A.2, retaining the recovery transition $sf \to rf$ (1/50) and the assumption $\delta(rf,t) = \delta(sf,t)$.

Observed proportions (aggregated class)

The land-cover data aggregate *tavy*, secondary forest, and recovered forest:

$$P_{ob}(ba) = 0.0567588857, \qquad P_{ob}(br) = 0.4957105006, \qquad P_{ob}(t + sf + rf) = 0.4475306137.$$

Stationarity gives $P(t) = P(ba) = 0.0567588857$, so the combined forest share (excluding bare land and brush) is

$$S := P(sf) + P(rf) = P_{ob}(t + sf + rf) - P(t) = 0.3907717280.$$

Solving for the transition demands

Substituting into the brush closed form gives $\delta(br, t) = P(ba)/P(br) - 1/11 = 0.0235909765$. The cultivation balance $P(t) = P(br)\delta(br, t) + S\,\delta(sf, t)$ then gives

$$\delta(sf, t) = \frac{P(t) - P(br)\delta(br, t)}{S} = 0.1153220352 = \delta(rf, t).$$

Finally, the secondary-forest closed form $\delta(sf, t) = P(br)/\big(11P(sf)\big) - 1/49$ identifies the split

$$P(sf) = \frac{P(br)}{11\left(\delta(sf, t) + \frac{1}{49}\right)} = 0.3320159513, \qquad P(rf) = S - P(sf) = 0.0587557766.$$

Numerical estimates (baseline)

$$\boxed{\delta(br, t) = 0.0235909765,} \qquad \boxed{\delta(sf, t) = \delta(rf, t) = 0.1153220352,}$$

$$\boxed{P(sf) = 0.3320159513,} \qquad \boxed{P(rf) = 0.0587557766.}$$

As a consistency check, the adding-up constraint holds: $P(t)+P(ba)+P(br)+P(sf)+P(rf) = 1$.

A.4 Calibrating neighbor weights with Bayesian optimization

The neighbor weights $\alpha(u \to v)$ that govern land-use configuration (main text, Section on the Transition Potential) are not known a priori. We calibrate them so that the simulated baseline map M_c reproduces the spatial structure of the observed land-classification map M_b, minimizing the discrepancy $\mathcal{D}(M_c, M_b)$ that combines edge density and cross-scale lacunarity (defined in the Calibration section of the main text). Because $\mathcal{D}$ has no closed form and each evaluation requires a stochastic cellular-automata simulation, we minimize it with Bayesian optimization (Algorithm A.1), implemented with the Optuna library [47] using a Tree-structured Parzen Estimator [48] as the surrogate model. We sample the weights in log space, because the system's qualitative behavior is sensitive to α across several orders of magnitude. Each trial runs the cellular automaton for 200 steps. We did not fix the number of trials in advance: we ran the optimization in batches of 100 trials and continued adding batches until the best discrepancy stopped improving.

Algorithm A.1: Bayesian optimization of the neighbor weights

Input: transitions to calibrate with weight ranges $[\alpha_{\min}, \alpha_{\max}]$; observed base map M_b; batch size $B = 100$ trials; cellular-automata steps per trial $S = 200$
Output: calibrated neighbor weights α^*
Precompute the cross-scale lacunarity curve and the edge density of M_b;
Initialize a Bayesian-optimization study (Tree-structured Parzen Estimator; objective direction: minimize);
repeat
  for $t = 1$ **to** B **do**
    Propose a candidate weight vector α, sampled in log space;
    Simulate the cellular automaton for S steps with weights α to obtain M_c;
    Evaluate the discrepancy $\mathcal{D}(M_c, M_b)$;
    Update the surrogate model with the pair $(\alpha, \mathcal{D})$;
  end
until *the best discrepancy did not improve over the last batch*;
return α^**, the weights achieving the lowest discrepancy observed*;

A.5 Adaptive sampling of the vanilla-fraction gradient

To decide how many vanilla-fraction values B to simulate along the *tavy*-to-vanilla gradient (main text, Section on Vanilla cultivation demand), we use an adaptive scheme that places new sample points only where the response is nonlinear (Algorithm A.2). We summarize each generated map by its Index of Translational Homogeneity (ITH) and insert sample points by interval bisection until linear interpolation of the accumulated points reproduces the full curve to within an error of 0.01. For our system, this procedure yielded nine sample points.

Algorithm A.2: Adaptive sampling of the vanilla-fraction gradient

Input: error threshold $\epsilon = 0.01$
Output: set of vanilla-fraction sample points $\{B\}$
Initialize samples at the extremes $B = 0$ and $B = 1$; generate a map and compute the ITH at each;
repeat
  foreach *interval* $[B_k, B_{k+1}]$ *between adjacent existing samples* **do**
    $B_m \leftarrow (B_k + B_{k+1})/2$;
    Predict $\widehat{\mathrm{ITH}}(B_m)$ by linear interpolation of the existing samples;
    Generate the map at B_m and compute the observed $\mathrm{ITH}(B_m)$;
    Record the error $\left|\widehat{\mathrm{ITH}}(B_m) - \mathrm{ITH}(B_m)\right|$;
  end
  Insert the midpoints B_m as new sample points;
until *the maximum interpolation error falls below* ϵ;

A.6 Fitting the tick-model parameters

A.6.1 Stochastic simulation algorithm

We advance the tick model with the Gillespie algorithm, a continuous-time stochastic method (Algorithm A.3), optimized as described by Lester et al. [37]. Because each tick undergoes events at a constant per-capita rate, every iteration reduces to two steps: selecting which tick acts and

the waiting time until it acts, and then deciding whether the event is a bite or a death. Per-cell stage populations are stored in an accumulating (Fenwick-style) weight tree, so that weighted selection and updates each cost $O(\log n)$.

Algorithm A.3: Event-based (Gillespie) tick-population simulation step

Input: lattice of per-cell stage populations $N_{i,\sigma}$; carrying capacities K_i; movement kernel G

Build an accumulating weight tree with one leaf per (cell, stage) pair, leaf weight = $N_{i,\sigma}$;

while *simulation time* $<$ *horizon* **do**

- Draw a (cell i, stage σ) by weighted random choice over the tree (probability $\propto N_{i,\sigma}$);
- Advance time by $\Delta t = -\ln U/W$, where $U \sim \mathrm{Unif}(0,1)$ and $W = \sum_{i,\sigma} N_{i,\sigma}$; `// exponential waiting time`
- Compute the hazard $\theta(N_{i,\sigma}, K_i) = \theta(0,K) + \left[\theta(K,K) - \theta(0,K)\right] N_{i,\sigma}/K_i$;
- **if** *the event is a death (probability* $\theta/(1+\theta)$*)* **then**
  - Remove one tick from (i,σ);
- **else**
  - The event is a bite (probability $Q = 1/(1+\theta)$); draw a destination cell j from kernel $G(x_i, \cdot)$;
  - **if** σ *is the adult stage* **then**
    - Add F new larvae to cell j (per-adult egg number; see below);
  - **else**
    - Move the tick to stage $\sigma + 1$ in cell j;
  - **end**
- **end**
- Update the affected leaves and propagate the partial sums up the tree in $O(\log n)$;

end

A.6.2 Survival Probability Q

The function $Q(N,K)$ represents the probability that a tick survives and successfully progresses to the next life stage when the local population size is N in a land-use cell with carrying capacity K.

To ensure that the carrying capacity has a clear biological interpretation, we model the *hazard ratio* between death and successful development rather than defining Q directly. Let

$$\theta(N,K) = \frac{1 - Q(N,K)}{Q(N,K)},$$

denote the density-dependent death-to-development hazard ratio.

We define $\theta(N,K)$ through two biologically interpretable boundary conditions:

$$\theta(0,K) \quad \text{(no competition)}$$

$$\theta(K,K) \quad \text{(at carrying capacity)}.$$

These two quantities represent the relative death pressure when the cell is empty and when it is saturated, respectively.

We construct $\theta(N, K)$ using linear interpolation between these two points:

$$\theta(N, K) = \theta(0, K) + [\theta(K, K) - \theta(0, K)] \frac{N}{K}.$$

The survival probability is then obtained as

$$Q(N, K) = \frac{1}{1 + \theta(N, K)}.$$

This formulation ensures that $Q(N, K)$ decreases smoothly with increasing density while remaining bounded between 0 and 1.

A.6.3 Number of Eggs per Adult Tick

To stabilize the population dynamics, we set the number of eggs produced per adult (F) such that, under high-density conditions ($N = K$), each tick replaces itself on average across the three life stages (larva, nymph, adult).

Since successful progression between stages occurs with probability $Q(K, K)$, the probability that a larva survives through all three stages is

$$Q(K, K)^3.$$

To maintain population stability at carrying capacity, this would set

$$F = \frac{1}{Q(K, K)^3} = (1 + \theta(K, K))^3.$$

In practice we add one further egg to this value:

$$F = (1 + \theta(K, K))^3 + 1.$$

The additional egg corrects for the derivation's assumption of equal competition at every life stage: real competition is stronger for larvae and weaker for adults, so the exact-derivation value above under-estimates the clutch a population needs to persist, and at low $\theta(K, K)$ a population parameterized with the uncorrected F goes deterministically extinct in simulation. The size of the correction was set by exploratory simulation of which values permit persistence rather than by the algebra above; its relative effect shrinks as $\theta(K, K)$ grows, matching where the equal-competition approximation is weakest (mortality lowest).

A.6.4 Carrying-capacity values

To decide on a carrying capacity for the different land uses, our collaborators captured small mammals in different land uses in the study area and counted how many ticks each one had (Table A.1).

Based on the capture data, we categorized habitats into four tick-abundance ranks based on

Table A.1: Proportion of animals with ticks and mean ticks captured per day by habitat type.

Habitat type	Proportion of animals with ticks	Mean ticks per capture day
Village	0.0	0.0
Agriculture	0.13	0.57
Agroforest	0.17	0.59
Brushy Regrowth	0.11	0.38
Flooded Rice	0.05	0.16
Secondary Forest	0.32	1.46
Semi-intact Forest	0.37	0.65

both the average number of ticks captured per day and the mean number of ticks per individual animal.

- **Village (blue):** No ticks were found, indicating unsuitable conditions for tick survival.
- **Agricultural areas and fallow land (yellow):** Moderately suitable habitats for ticks, with intermediate values for both ticks per capturing day and ticks per animal.
- **Semi-intact forest (dark green):** High tick load per animal but a moderate overall tick count, suggesting concentrated infestations in specific hosts.
- **Secondary forest (light green):** High values for both total tick abundance and mean ticks per animal, indicating optimal conditions for tick populations.

For each of these categories we set a carrying-capacity value. These values preserve the order of the data, but we do not have enough data to fit them accurately; we therefore did not fit them quantitatively, only preserved the order, obtaining the values shown in Table A.2.

Table A.2: Carrying-capacity values assigned to each habitat type. Values preserve the rank order observed in Table A.1 but are not quantitatively fitted.

Land use	Carrying capacity
Village	15
Agricultural areas and fallow land	30–35
Secondary	45
Semi-intact	60

A.7 Regional sensitivity analysis of the tick survival parameters

Ecological models often depend on parameters that are difficult to measure precisely, so before applying the tick model to real land-use scenarios we asked how much these uncertainties matter: which parameters most strongly influence the outcome, which combinations lead to persistence rather than extinction, and how robust the system is to changes in key assumptions. Rather than only quantifying sensitivity, we used this analysis to map the boundary between viable and non-viable ecological conditions—the tipping points at which tick populations either collapse or persist—and to select a parameter set that reproduces realistic spatial heterogeneity.

This step is a *regional* (and, where the full space is covered, *global*) sensitivity analysis over the parameters of the survival function. We did not use local, one-at-a-time analysis, because the interactions between parameters are too strong for it to be meaningful. The complementary *structural* sensitivity analysis, which varies landscape structure rather than parameters, is treated separately in Section A.8 owing to its additional complexity.

Mapping the viable parameter space with a support vector machine

We focused on the three parameters of the survival function $Q(N, K)$ introduced above: the carrying capacity K, the no-competition hazard ratio $\theta(0, K)$ (the death-to-development pressure when ticks are rare), and the carrying-capacity hazard ratio $\theta(K, K)$ (the death-to-development pressure when the cell is saturated). These correspond, respectively, to a carrying capacity and to low- and high-density death rates.

We ran many simulations with randomly sampled values of these parameters and labeled each outcome as

$$0\text{: the population went extinct,} \qquad 1\text{: the population persisted.}$$

To generalize these discrete outcomes into a continuous description of viability, we trained a support vector machine (SVM) classifier to learn the boundary between survival and extinction in the three-dimensional parameter space. After comparing several kernels, we adopted a polynomial kernel, which best captured the nonlinear shape of the viability boundary.

Active learning

To obtain an accurate boundary with as few simulations as possible, we used active learning: the current SVM identified parameter sets lying near its decision boundary—where classification was most uncertain—and we concentrated new simulations there. Iterating this targeted sampling refined the survival/extinction boundary far more efficiently than exhaustive sampling of the parameter space. We summarize the procedure in Algorithm A.4; it was implemented with the `LIBSVM.jl` library [49], exploring the ranges $K \in [1, 100]$, $\theta(K, K) \in [2, 20]$, and $\theta(0, K) \in [1, \theta(K, K))$.

Algorithm A.4: Active-learning support-vector classification of the viable region

Input: parameter ranges for K, $\theta(0, K)$, $\theta(K, K)$; total simulation budget n; candidate-pool size $m = 100$

Output: support-vector classifier of viability (persistence vs. extinction)

Sample an initial set of parameter points; run the tick model at each and label it 1 (population persists) or 0 (population goes extinct);

Train a support vector machine (polynomial kernel) on the labeled points;

while *the number of simulations* $< n$ **do**

- Draw m random candidate parameter sets;
- Select the candidate whose decision value is nearest the SVM boundary (most uncertain, closest to 0);
- Run the tick model at that candidate and label the outcome;
- Add the labeled point to the training set and retrain the SVM;

end

Behavior within the viable region

Having delineated the viable region, we examined how the model behaves inside it using three summary statistics: the proportion of occupied cells, the mean number of ticks per occupied cell, and the overall average tick density. Many parameter sets produced all-or-nothing outcomes, in which nearly all cells were either full or empty (Figure A.1). Real ecosystems are not binary, however, so we specifically sought *mixed* outcomes—some patches occupied and others empty—consistent with the field observation that some areas carry ticks while others do not, across a range of carrying capacities.

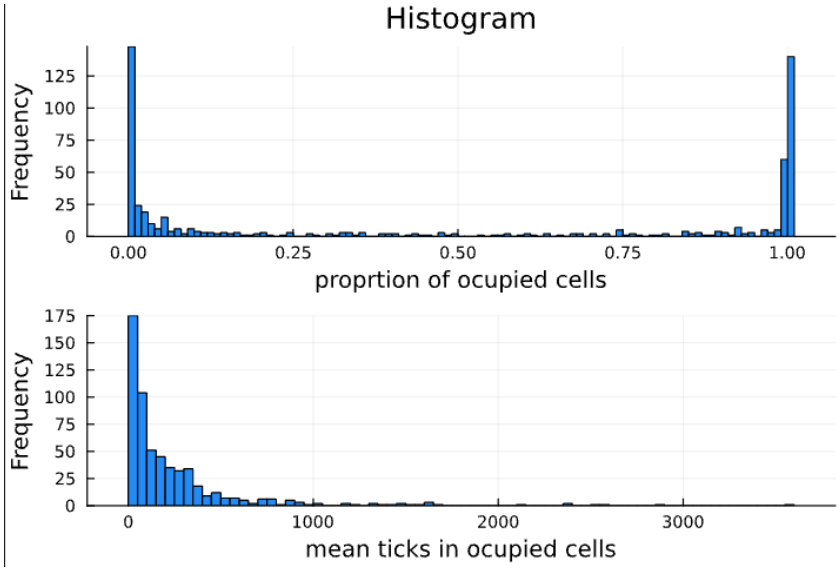


Figure A.1: Distribution of simulation outcomes across the viable parameter space. Many parameter combinations lead to near-total persistence or near-total extinction; intermediate outcomes with partial patch occupancy—the ecologically relevant regime—are less common.

Several consistent patterns emerged across the viable space (Figure A.2):

- Increasing $\theta(0, K)$ (higher death pressure when ticks are rare) reduces both the proportion of occupied cells and the mean number of ticks per occupied cell.
- Increasing K has the opposite effect, raising both patch occupancy and tick density.
- A larger gap between $\theta(0, K)$ and $\theta(K, K)$ increases the number of surviving patches but lowers the mean tick count within them, plausibly because cells saturate early.

Guided by these relationships, we selected a parameter combination that balances persistence against realism: tick populations persist in some areas while remaining ecologically plausible, and spatial heterogeneity is retained over a range of carrying capacities. This is the basis for the survival-function values fitted below ($\theta(0, K) = 4$, $\theta(K, K) = 12$). The remaining question—whether the model responds only to average carrying capacity or also to the structure of the landscape itself—is addressed by the structural sensitivity analysis below.

Fitting the survival function

To determine appropriate values for $\theta(0, K)$ and $\theta(K, K)$, we imposed two ecological constraints across a wide range of carrying capacities:

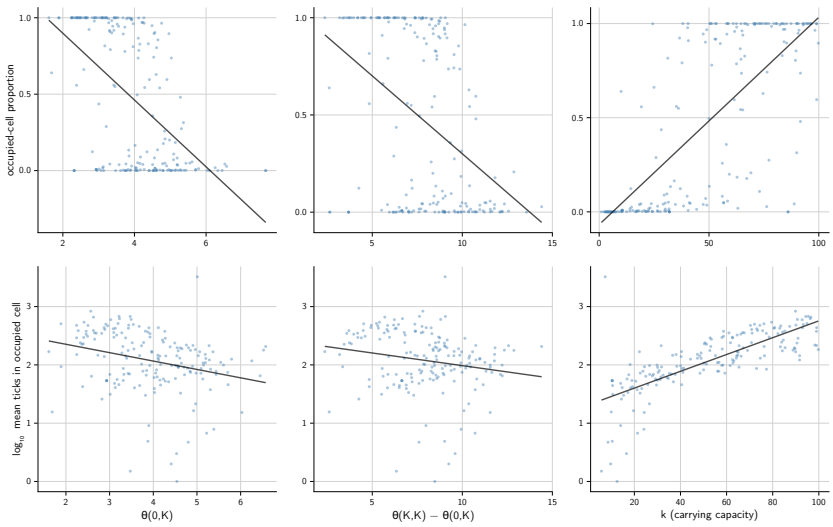


Figure A.2: Relationships between survival-function parameters and population outcomes. Adjusting $\theta(0, K)$ and $\theta(K, K)$ simultaneously controls mean tick density and the degree of spatial heterogeneity across simulated landscapes.

1. Partial occupancy: tick populations should occupy some, but not all, land cells.
2. Bounded abundance: the average tick abundance within occupied cells should not exceed 100 individuals.

We explored the parameter space using a support vector machine (SVM) classifier combined with manual calibration to identify regions satisfying both conditions.

This procedure yielded

$$\theta(0, K) = 4, \qquad \theta(K, K) = 12.$$

Substituting these values gives

$$\theta(N, K) = 4 + 8\frac{N}{K},$$

and therefore

$$Q(N, K) = \frac{1}{1 + \theta(N, K)} = \frac{1}{5 + 8\frac{N}{K}}.$$

Movement kernel G

We evaluated four alternative movement kernels defining the probability that a tick in cell i, after a successful bite event, relocates to cell j within a maximum radius R:

- **Random:** $G(i, j) = 1$
- **Linear distance decay:** $G(i, j) = R - d(i, j)$
- **Gaussian distance decay**

- **Land-use dependent:** $G(i, j) \propto K_j$

Sensitivity analysis showed that the kernel radius and the distinction between the random and distance-dependent kernels did have a statistically significant effect on overall tick burden, but with effect sizes small enough to be irrelevant relative to the influence of the other factors (carrying capacity and the survival-function parameters). Only the land-use-dependent kernel produced a substantial change in long-term abundance patterns.

Given the negligible effect size of the spatial decay structure, we adopted the random kernel for the main simulations to avoid introducing unnecessary complexity.

A.8 Structural sensitivity analysis

To ensure the ecological relevance and internal consistency of the tick population model, we tested whether it responds independently to variation in (i) land-use composition and (ii) land-use configuration. Because real landscapes often exhibit correlated changes in these two dimensions, we constructed synthetic maps that allow them to vary independently.

Synthetic Landscape Generation

We generated artificial landscapes using a two-dimensional midpoint displacement algorithm (diamond–square method) [50], which produces fractal surfaces characterized by a Hurst parameter H [51]. The Hurst parameter controls spatial aggregation:

- Low H values produce fragmented, spatially noisy landscapes.
- High H values generate smoother, more spatially aggregated patterns.

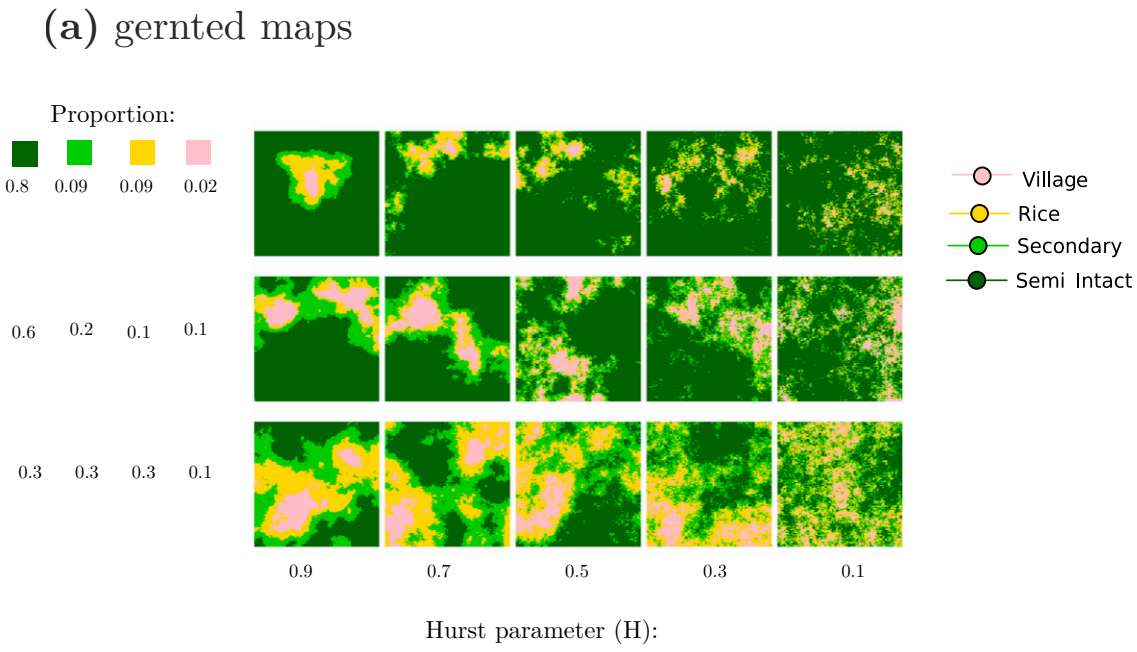
For each value of H, the continuous fractal surface was discretized into land-use categories by applying thresholds corresponding to predefined land-use proportions. By varying these thresholds, we independently controlled overall land-use composition while holding spatial configuration constant. Conversely, by varying H while maintaining fixed thresholds, we altered spatial configuration without changing composition.

This procedure produced a grid of synthetic landscapes spanning independent gradients of composition and aggregation.

Tick Model Evaluation

For each synthetic map, we ran the tick population model to equilibrium and recorded mean tick abundance within each land-use type. We then examined how tick abundance responded to:

1. Increasing aggregation (higher H) at fixed composition.
2. Changing land-use proportions at fixed spatial configuration.



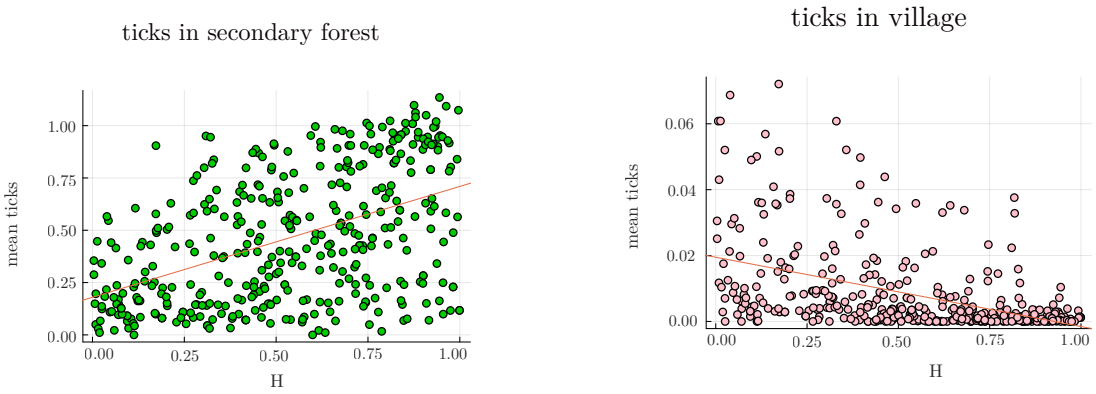


Figure A.3: Sensitivity analysis of the tick population model. (a) Synthetic landscapes were generated using the midpoint displacement (diamond–square) algorithm, allowing independent manipulation of land-use composition and spatial aggregation (Hurst parameter H). (b) Tick abundance responds systematically to both composition and configuration. Increasing aggregation (higher H) increases tick abundance in secondary forest (high-quality habitat) while decreasing abundance in village areas (low-quality habitat), consistent with expected edge effects.

Results and Interpretation

The model exhibited clear sensitivity to both composition and configuration (Figure A.3). Increasing the Hurst parameter (i.e., increasing spatial aggregation) led to:

- Higher tick abundance in secondary forest, the most suitable habitat.
- Lower tick abundance in village areas, which represent low-quality habitat.

This pattern is consistent with ecological expectations: greater aggregation reduces edge density, allowing high-quality habitats to form larger contiguous patches. As a result, tick populations concentrate more strongly within suitable habitats and are less frequently dispersed into marginal environments.

Importantly, composition effects remained detectable when configuration was held constant, and configuration effects remained detectable when composition was held constant. This confirms that the tick model is sensitive to both structural dimensions of the landscape and does not respond solely to overall land-use proportions.

Together, these results validate that the tick population model captures biologically meaningful spatial dynamics rather than being driven purely by global abundance constraints.

A.9 Risk calculation

For each resulting land-use map, we conducted eight independent tick population simulations. We calculated the Risk in each simulation as the average tick abundance in anthropogenic (agriculture or village) cells:

$$\text{Risk} = \frac{100}{3} \times (\text{average ticks in agricultural land}) + \frac{200}{3} \times (\text{average ticks in village}) \quad \text{(A.15)}$$

The calculation is based on the assumption that people spend 8 hours per day working in the fields and 16 hours in the Village. We multiply by 100 to place the result on a convenient scale.

Bibliography

[1] Hald-Mortensen, Christian. The main drivers of biodiversity loss: A brief overview. *Journal of Ecology & Natural Resources*, 2023.

[2] Swei, A. Patterns, drivers, and challenges of vector-borne disease emergence. *Vector Borne Zoonotic Dis.*, 20(3):159–170, 2020.

[3] Borrelli, P. An assessment of the global impact of 21st century land use change on soil erosion. *Nat. Commun.*, 8(1):2013, 2017.

[4] Smith, P. Agriculture, forestry and other land use (AFOLU). In *Climate Change 2014: Mitigation of Climate Change*, pages 811–922. Cambridge University Press, 2014.

[5] Hamad, R. Predicting land use/land cover changes using a CA-markov model under two different scenarios. *Sustainability*, 10(10):3421, 2018.

[6] Che, L. Land use evolution and prediction: A bibliometric review. *Int. J. Environ. Sci. Technol.*, 22(12):11077–11090, 2025.

[7] Zare, M. Modeling the effect of land use and climate change scenarios on future soil loss rate in kasilian watershed of northern iran. *Environ. Earth Sci.*, 76(8):286, 2017.

[8] Huang, Yun and Liao, Tie-Jun. An integrating approach of cellular automata and ecological network to predict the impact of land use change on connectivity. *Ecol. Indic.*, 98:149–157, March 2019.

[9] Nath, A J. The shifting cultivation juggernaut: An attribution problem. *Global Chall.*, 6 (8):2200051, 2022.

[10] Lowder, S K. The number, size, and distribution of farms, smallholder farms, and family farms worldwide. *World Dev.*, 87:16–29, 2016.

[11] Jepsen, M R. Agent-based modelling of shifting cultivation field patterns, vietnam. *Int. J. Geogr. Inf. Sci.*, 20(9):1067–1085, 2006.

[12] Styger, E. Influence of slash-and-burn farming practices on fallow succession and land degradation in madagascar. *Agric. Ecosyst. Environ.*, 119(3–4):257–269, 2007.

[13] Mariel, J. From shifting rice cultivation (tavy) to agroforestry systems: A century of changing land use on the east coast of madagascar. *Agrofor. Syst.*, 97(3):415–431, 2023.

[14] Llopis, J C. Effects of protected area establishment and cash crop price dynamics on land use transitions 1990–2017 in north-eastern madagascar. *J. Land Use Sci.*, 14(1):52–80, 2019.

[15] Fujiki, K. The evolution of a rural territory at plot scale: Between hyper-fragmentation and land grabbing. *Land Use Policy*, 43:170–185, 2015.

[16] Huang, D. Land use/land cover remote sensing classification in complex subtropical karst environments. *Applied Sciences*, 15(17):9641, 2025.

[17] Newman, E A. Scaling and complexity in landscape ecology. *Front. Ecol. Evol.*, 7:293, 2019.

[18] Rufin, Philippe. Large-area mapping of active cropland and short-term fallows in smallholder landscapes using PlanetScope data. *Int. J. Appl. Earth Obs. Geoinf.*, 112:102937, 2022.

[19] Wickramasuriya, R C. The dynamics of shifting cultivation captured in an extended constrained cellular automata land use model. *Ecol. Modell.*, 220(18):2302–2309, 2009.

[20] Taubert, F. Global patterns of tropical forest fragmentation. *Nature*, 554(7693):519–522, 2018.

[21] Ralimanana, H. Madagascar's extraordinary biodiversity: Threats and opportunities. *Science*, 378(6623):eadf1466, 2022.

[22] Zaehringer, J G. Beyond deforestation monitoring in conservation hotspots: Analysing landscape mosaic dynamics in north-eastern madagascar. *Applied Geography*, 68:9–19, 2016.

[23] Rajaofera, M J N. Public health surveillance of tropical diseases in madagascar: A scoping review. *BMC Public Health*, 25(1):2651, 2025.

[24] Rakotosamimanana, S. A decade of plague in madagascar: a description of two hotspot districts. *BMC Public Health*, 21(1):1112, 2021.

[25] Kauffman, Kayla. Comparing transmission potential networks based on social network surveys, close contacts and environmental overlap in rural madagascar. *J. R. Soc. Interface*, 19(186):20210690, 2022.

[26] Blache, N. A scoping review of the impacts of forest cover dynamics on acari-borne diseases: Beyond forest fragmentation. *Heliyon*, 11(2):e41893, 2025.

[27] Hänke, H. Socio-economic, land use and value chain perspectives on vanilla farming in the SAVA region (north-eastern madagascar). Technical report, Diskussionsbeitrag, 2018.

[28] Kool, Wouter, Van Hoof, Herke, and Welling, Max. Stochastic beams and where to find them: The Gumbel-Top-k trick for sampling sequences without replacement. In *Proceedings of the 36th International Conference on Machine Learning (ICML 2019)*, pages 3499–3508, 2019.

[29] Allain, C and Cloitre, M. Characterizing the lacunarity of random and deterministic fractal sets. *Phys. Rev. A*, 44(6):3552–3558, 1991. doi: 10.1103/PhysRevA.44.3552.

[30] Plotnick, Roy E, Gardner, Robert H, Hargrove, William W, Prestegaard, Karen, and Perlmutter, Marguerite. Lacunarity indices as measures of landscape texture. *Landsc. Ecol.*, 11(1):33–39, 1996. doi: 10.1007/BF02087112.

[31] Malhi, Yadvinder and Román-Cuesta, Rosa María. Analysis of lacunarity and scales of spatial homogeneity in IKONOS images of amazonian tropical forest canopies. *Remote Sens. Environ.*, 112(5):2074–2087, May 2008.

[32] Henrys, Peter A, Mondain-Monval, Thomas O, and Jarvis, Susan G. Adaptive sampling in ecology: Key challenges and future opportunities. *Methods Ecol. Evol.*, 15(9):1483–1496, September 2024.

[33] World Bank. Guide on sustainable vanilla cultivation and cost-benefit analysis: Case study of the SAVA region. Technical report, World Bank, October 2021.

[34] Decraene, James. The emergence of urban land use patterns driven by dispersion and aggregation mechanisms. *PLoS One*, 8:e80309, 2013.

[35] Gefen, Y, Meir, Y, Mandelbrot, B B, and Aharony, A. Geometric implementation of hypercubic lattices with noninteger dimensionality by use of low lacunarity fractal lattices. *Phys. Rev. Lett.*, 50(3):145–148, 1983. doi: 10.1103/PhysRevLett.50.145.

[36] Gillespie, Daniel T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977. doi: 10.1021/j100540a008.

[37] Lester, Christopher, Baker, Ruth E, and Yates, Christian A. Efficiently simulating discrete-state models with binary decision trees. *arXiv [q-bio.QM]*, January 2020.

[38] Akaike, Hirotugu. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.*, 19(6):716–723, 1974. doi: 10.1109/TAC.1974.1100705.

[39] Burnham, Kenneth P. and Anderson, David R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, 2 edition, 2002.

[40] Magliocca, Nicholas R and Ellis, Erle C. Using pattern-oriented modeling (POM) to cope with uncertainty in multi-scale agent-based models of land change. *Trans. GIS*, 17(6): 883–900, 2013. doi: 10.1111/tgis.12012.

[41] van Vliet, N. Trends, drivers and impacts of changes in swidden cultivation in tropical forest-agriculture frontiers: A global assessment. *Glob. Environ. Change*, 22(2):418–429, 2012. doi: 10.1016/j.gloenvcha.2011.10.009.

[42] Kim, Dong-Gill. Shifting cultivation in africa: a comprehensive review of its impacts on biodiversity, soil health, and livelihoods. *Agrofor. Syst.*, 99(6), August 2025.

[43] Magnuszewski, P. Resilience and alternative stable states of tropical forest landscapes under shifting cultivation regimes. *PLoS One*, 10(9):e0137497, 2015.

[44] Kupfer, John A. Landscape ecology and biogeography: Rethinking landscape metrics in a post-FRAGSTATS landscape. *Prog. Phys. Geogr.*, 36(3):400–420, 2012. doi: 10.1177/0309133312439594.

[45] Frazier, Amy E and Kedron, Peter. Landscape metrics: Past progress and future directions. *Curr. Landscape Ecol. Rep.*, 2(2):63–72, 2017. doi: 10.1007/s40823-017-0026-0.

[46] Raherinirina, Angelo, Hervé, Dominique, and Campillo, Fabien. Inferring land use dynamics by semi-markov model. In *12. African Conference on Research in Computer Science and Applied Mathematics*, pages 1–8, October 2014.

[47] Akiba, Takuya, Sano, Shotaro, Yanase, Toshihiko, Ohta, Takeru, and Koyama, Masanori. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2623–2631, 2019. doi: 10.1145/3292500.3330701.

[48] Bergstra, James, Bardenet, Rémi, Bengio, Yoshua, and Kégl, Balázs. Algorithms for hyper-parameter optimization. In *Advances in Neural Information Processing Systems 24 (NIPS 2011)*, pages 2546–2554, 2011.

[49] Chang, Chih-Chung and Lin, Chih-Jen. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):1–27, 2011. doi: 10.1145/1961189.1961199.

[50] Fournier, Alain, Fussell, Don, and Carpenter, Loren. Computer rendering of stochastic models. *Commun. ACM*, 25(6):371–384, 1982. doi: 10.1145/358523.358553.

[51] Peitgen, Heinz-Otto and Saupe, Dietmar, editors. *The Science of Fractal Images*. Springer-Verlag, 1988.

תוכן עניינים

יישום אוטומטים תאיים ושרשראות מרקוב לחיזוי שינויי נוף בקני־מידה מרובים בחקלאות מסורתית

יובל בלוך

עבודת גמר לתואר מוסמך למדעי הטבע

אוניברסיטת בן־גוריון בנגב

2026

תקציר

ינוי בשימושי קרקע הוא גורם מרכזי בהידרדרות סביבתית ובסיכונים לבריאות הציבור ברחבי העולם, התורם לאובדן המגוון הביולוגי, לדלדול קרקעות ולהופעת מחלות זואונוטיות (ZOONOTIC DISEASES) . אף על פי שקיימים מודלים רבים המסוגלים לחזות שינויים עתידיים בשימושי קרקע במגוון רחב של סביבות, רובם פותחו עבור נופים עירוניים או נופי חקלאות תעשייתית, המתאפיינים בתהליכי מעבר ארוכי טווח ורחבי היקף.לעומת זאת יכולתנו למדל מערכות של חקלאות נדדת (SHIFTING CULTIVATION) וחקלאות יער של משקים קטנים – הפועלות במגוון קני מידה מרחביים וזמניים – נותרה מוגבלת.

לפער מחקרי זה ישנה חשיבות רבה שכן חלקים גדולים מהאוכלוסיה באזורים טרופיים בעלי הכנסה נמוכה תלויים במערכות אגרו־אקולוגיות אלו למחייתם, לרוב בנופים החופפים לאזורי מפתח של מגוון ביולוגי (BIODIVERSITY HOTSPOTS). קהילות אלו פגיעות במיוחד לדלדול קרקע ולחשיפה למחלות, ולכן הבנה מעמיקה יותר של דינמיקת שימושי הקרקע שלהן חיונית הן לרווחת האדם והן לשימור אקולוגי.

כדי להתמודד עם אתגר זה, פיתחנו מערכת מודל (MODELING FRAMEWORK) המתאימה מדדים של הרכב (COMPOSITION) ומבנה מרחבי (CONFIGURATION) ברמת הנוף, במקום להסתמך על התאמה מרחבית נקודתית של תאי שטח. כדי ללכוד את המבנה של נופים אלו על פני קני מידה מרובים, שילבנו גיאומטריה פרקטלית בגישת המידול. כיילנו את המודל מול מפות שימושי קרקע שנצפו במחוז סאבא (SAVA) שבצפון־מזרח מדגסקר, ולאחר מכן חזינו את נופי שיווי המשקל (EQUILIBRIUM LANDSCAPES) הצפויים תחת מגוון תרחישים אגרו־כלכליים, המייצגים תלות גוברת בחקלאות יער מסחרית של וניל, והשווינו את ההרכב והמבנה המרחבי שלהם. כדי להוכיח את שימושיות המודל, צימדנו בין הנופים החזויים לבין מודל של אוכלוסיית קרציות.

במהלך כיול המודל מצאנו כי מדדי שימושי הקרקע שבהם השתמשנו איפשרו לכייל רק שני פרמטרים של מבנה מרחבי: הסטוכסטיות המרחבית של החקלאות והסטוכסטיות המרחבית של השינוי הטבעי, מה שמראה שהדרך למודלים מורכבים יותר מחייבת הרחבה של האינדקסים המשמשים את פונקציית המטרה. כמו כן מצאנו כי מדדי המבנה המרחבי היו רגישים להבדלים בין שיטות גידול רק כאשר הוניל היה הגידול הדומיננטי, בעוד

שמודל הקרציות היה רגיש להבדלים אלו כל עוד היו חלקות וניל במערכת -- ממצא המצביע על מגבלה של מדדי מבנה מרחבי בהשוואה למודל ביולוגי. במודל המשולב, החשיפה לקרציות השתנתה באופן לא־מונוטוני לאורך המסלול האגרו־כלכלי והשפעת ההחלטות המרחביות של החקלאים הייתה תלוית־שלב.

ממצאים אלו מדגימים את ערכה של מערכת המודל לקביעת מדיניות של ניהול קרקע ובריאות הציבור במערכות טרופיות של משקים קטנים. מעבר לחשיפה לקרציות, גישה זו ישימה גם לשאלות של דלדול קרקע, ניהול מים ושימור המגוון הביולוגי. במובן רחב יותר, מחקר זה שופך אור על הקשר שבין מאפייני נוף פרקטליים לבין מערכות חקלאיות דינמיות, ובכך מקדם את הבנתנו לגבי נופי חקלאות נדדת וחקלאות יער ברחבי העולם.

אוניברסיטת בן־גוריון בנגב
הפקולטה למדעי הטבע
המחלקה למדעי החיים

סיון תשפ"ו

יישום אוטומטים תאיים ושרשראות מרקוב לחיזוי שינויי נוף בקני־מידה מרובים בחקלאות מסורתית

חיבור זה מהווה חלק מהדרישות לקבלת התואר מוסמך למדעי הטבע (M.Sc)

יובל בלוך

בהנחיית פרופ', שי פילוסוף

חתימת המחבר _____ תאריך: 08/12/2026

אישור המנחה _____ תאריך: 12 August 2026

אישור יושב ראש ועדת מוסמכים מחלקתית _____ תאריך: _____

* אם יש מספר מנחים יש להקצות שורה לכל מנחה

אוניברסיטת בן־גוריון בנגב
הפקולטה למדעי הטבע
המחלקה למדעי החיים

יישום אוטומטים תאיים ושרשראות מרקוב לחיזוי שינויי נוף בקני־מידה מרובים בחקלאות מסורתית

חיבור לשם קבלת התואר מוסמך בפקולטה למדעי הטבע

יובל בלוך

בהנחיית פרופ', שי פילוסוף

סיון תשפ"ו June 2026