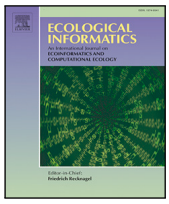
Agent-based ecosystem modeling with deep reinforcement learning

Ecological Informatics 96 (2026) 103819

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



Agent-based ecosystem modeling with deep reinforcement learning

Claes Strannegård [a,b,c,*], Michał Palak [d], Niklas Engsner [c], Alice Stocco [e], Alexandre Antonelli [b,f,g], Daniele Silvestro [b,f,h]

[a] Department of Applied Information Technology, University of Gothenburg, Sweden
[b] Gothenburg Global Biodiversity Centre, University of Gothenburg, Sweden
[c] Department of Molecular Medicine and Surgery, Karolinska Institutet, Sweden
[d] Department of Computer Science and Engineering, Chalmers, Sweden
[e] Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Italy
[f] Department of Biological and Environmental Sciences, University of Gothenburg, Sweden
[g] Royal Botanic Gardens, Kew, United Kingdom
[h] Department of Biosystems Science and Engineering, ETH Zurich, Switzerland

ARTICLE INFO

Dataset link: https://doi.org/10.5281/zenodo.19221981

Keywords:
Agent-based modeling
Deep reinforcement learning
Ecosystem modeling
Pattern-oriented modeling
Sustainable decision-making

ABSTRACT

Ecosystem models can support understanding, monitoring, and management of ecological dynamics across space and time. Yet many existing animal-behavior simulations depend on manually specified rules, which limits scalability, transferability, and realism. We present a flexible agent-based modeling framework that leverages deep reinforcement learning to generate adaptive animal behavior without hand-coded decision rules. As a case study, we construct a model of an Alpine ecosystem comprising wolves, chamois, and vegetation, and evaluate it using Pattern-Oriented Modeling. The resulting simulations reproduce key ecological patterns, including long-term coexistence across multiple landscapes, predator–prey dynamics, and behavior qualitatively consistent with that of the modeled species. We further show how the model can be used to explore ecosystem resilience under scenarios of habitat degradation, game hunting, and heat stress. Finally, we compare our machine-generated model to a rule-based, hand-crafted model and observe that it outperforms the latter. While ecosystem modeling with deep reinforcement learning remains nascent and experimental, our approach provides a scalable step toward more flexible computational ecosystem models for exploring ecosystem responses to disturbance.

1. Introduction

Ecosystems and their species provide us with goods and services such as food, clean water, and materials for buildings, clothing, and medicine (Daily and Matson, 2008). However, more than a million species are estimated to be threatened with extinction, which means that without radical policy changes, the contributions of nature to people may soon be disrupted, with profound consequences for biodiversity and human communities around the world (Antonelli, 2022). The urgency of addressing the biodiversity crisis has been recognized in landmark international agreements that call for major improvements in global protection of nature (Hughes and Grumbine, 2023). Part of our ability to address this crisis depends on understanding ecosystems and predicting biodiversity dynamics.

Recent research has highlighted the potential of ecosystem models to guide the development of science-based policies that protect biodiversity while supporting long-term human prosperity (Silvestro et al., 2022; Weiskopf et al., 2022). Some of these, referred to as "digital twins", combine data, models, and expert knowledge in continuous alignment with the real world (de Koning et al., 2023), enabling human interventions to trigger particular outcomes, such as avoided population collapses or local species extinctions. Modeling approaches have also been applied to study ecosystem dynamics (Geary et al., 2020; Vieira et al., 2022) and to explore the consequences of human activities, such as agriculture, forestry, urban development, and tourism, from the planning stage onward (Strannegård et al., 2024a), providing decision-makers with input to reduce negative environmental impacts or maximize the positive impact, e.g., in a restoration project.

As an example, ecosystem models have been used to predict biomass variation (Brigolin et al., 2009) and ecological interactions, including predator–prey dynamics and the effects of ecosystem changes driven by climate change and human activities (Geary et al., 2020).

One of the most promising approaches to ecosystem modeling is agent-based modeling (ABM) (Wilensky and Rand, 2015; DeAngelis

* Corresponding author.
E-mail address: claes.strannegard@gu.se (C. Strannegård).

https://doi.org/10.1016/j.ecoinf.2026.103819
Received 9 January 2026; Received in revised form 30 March 2026; Accepted 9 May 2026
Available online 15 May 2026

and Grimm, 2014; Crooks et al., 2019; Rollins et al., 2014), where animals are modeled individually. Although agent-based ecosystem models are highly flexible, their full potential remains unrealized (Siekmann and Osborne, 2023). A central limitation is the difficulty of realistically representing organism behavior across diverse environmental and physiological conditions. Animal behaviors—such as movement, feeding, and resting—must respond to complex external and internal inputs (DeAngelis and Grimm, 2014). These behaviors are often hand-coded using "if–then" rules grounded in behavioral ecology, but this approach is inherently limited (DeAngelis and Diaz, 2019). Organisms respond to multiple interacting sensory and physiological cues, making it nearly impossible to construct exhaustive and accurate rule sets (Kaul and Ventikos, 2015).

Ensuring ecological stability is another major challenge, as models must allow functional groups to coexist in a dynamic equilibrium under favorable conditions (Walters and Christensen, 2007). In addition, both hand-coded and supervised learning approaches require detailed empirical behavioral data, which are often scarce (LeCun et al., 2015). Finally, reliable population data for all functional groups are needed to define initial conditions and validate model outcomes.

The rapid rise of artificial intelligence has led to emerging powerful alternatives to rule-based approaches, proving highly successful in areas such as protein structure prediction (Jumper et al., 2021) and weather forecasting (Price et al., 2025). In this paper, we propose using reinforcement learning rather than hand-coding to construct behavioral models for agent-based ecosystem simulations. Our strategy for ecosystem modeling relies on three key ideas: (i) animals are modeled as deep reinforcement learning agents that make decisions based on their perception of the environment and their internal homeostatic variables; (ii) the agents are rewarded for maintaining homeostasis and thus, indirectly, for surviving by eating, drinking, avoiding predators and navigating efficiently in the landscape; and (iii) the agents are trained across multiple environments with varying ecological conditions, so that they become flexible enough to survive in models of diverse geographical areas.

We first present our strategy for agent-based ecosystem modeling. Then we use this strategy to construct an ecosystem model of an Alpine ecosystem featuring wolves, chamois, and three types of vegetation. We then use Pattern-Oriented Modeling (Grimm et al., 2005; Grimm and Railsback, 2012) to evaluate the ecosystem model against seven ecological patterns reported for Alpine ecosystems:

1. **Long-term persistence:** Stable wolf–chamois coexistence over decades; populations remain within empirical bounds.
2. **Predator–prey dynamics:** Coupled predator–prey fluctuations consistent with Lotka–Volterra-type dynamics.
3. **Spatial distribution:** Emergent habitat selection and space use are consistent with efficient use of resources.
4. **Life-history statistics:** Survival and fecundity of chamois and wolves are within reported ranges.
5. **Behavioral patterns:** Chamois track resources and avoid predation risk; wolves track prey and avoid barriers.
6. **Landscape generality:** Coexistence across multiple Alpine landscapes.
7. **Disturbance resilience:** Persistence and recovery under moderate environmental pressure in the form of hunting, heat stress, and habitat degradation.

Finally, we compare our model with a traditional rule-based ecosystem model.

2. Previous work

2.1. Analytic ecosystem models

Ecosystem dynamics have traditionally been described with analytic models such as the Lotka–Volterra (Lotka, 1925) and Arditi–Ginzburg equations (Arditi and Ginzburg, 1989). These models describe predator–prey dynamics using systems of ordinary differential equations valued for their ability to capture population interactions over time. Their strength lies in modeling species interactions and biomass fluctuations across multiple generations, thereby contributing to our understanding of energy and mass flow within ecosystems (Swartzman, 1979). However, analytically modeling organism behavior in complex and realistic scenarios remains challenging—for instance, when multiple functional groups (e.g., carnivores, herbivores, and primary producers) interact and when spatial features or environmental variables influence ecosystem dynamics (Geary et al., 2020).

2.2. Simulation-based ecosystem models

Another approach to modeling ecosystems is computer simulation (Soetaert and Herman, 2009; Grimm and Railsback, 2013), which helps address problems for which analytic solutions are impractical or unavailable—for example, ecosystems with multiple functional groups that go beyond simple predator–prey dynamics, or those that explicitly incorporate spatial features. Several simulation models have been developed to generate realistic representations of biological systems, including *population-based models*, which represent organisms at an aggregated level (Royama, 2012; Colléter et al., 2013; Silvestro et al., 2022; Hagen et al., 2021; Saraiva et al., 2014), and *agent-based models*, in which certain organisms are represented individually (Wilensky and Rand, 2015; DeAngelis and Grimm, 2014; Crooks et al., 2019; Rollins et al., 2014). In agent-based models, each agent typically has a decision-making mechanism that governs its actions, such as movement and feeding, within a spatially explicit environment. Thanks to these characteristics, agent-based models have proven useful in wildlife management (McLane et al., 2011), fishery ecology (Lindkvist et al., 2020), and even evolutionary contexts (Hagen et al., 2021).

2.3. Ecosystem model validation

Real ecosystems and their models can be compared along multiple qualitative and quantitative dimensions. Accordingly, any claim about model validity or biological plausibility should explicitly specify to which dimensions the claim refers (Rykiel, 1996). Even with these clarifications in place, model validity and biological plausibility are seldom crisp, binary properties; rather, they typically admit degrees and depend on context. A range of validation frameworks for ecosystem modeling has been developed to structure such comparisons, from broad integrative approaches such as the IPBES conceptual framework (Díaz et al., 2015) to applied toolkits such as InVEST for ecosystem-service assessment and decision support (Nelson et al., 2009).

A common limitation of purely quantitative validation approaches is data scarcity: biomass estimates for relevant functional groups at sufficiently fine spatial and temporal resolution are often unavailable, and collecting them can be prohibitively expensive. Moreover, no single dataset is typically sufficient to rigorously validate complex ecological models, particularly when models are used to project dynamics under novel environmental conditions for which no empirical data exist.

Pattern-Oriented Modeling (POM) (Grimm et al., 2005; Grimm and Railsback, 2012) addresses these limitations by integrating quantitative and qualitative validation through the use of multiple observed ecological patterns as simultaneous criteria for model development, calibration, and evaluation. By requiring models to reproduce several independent patterns rather than a single target dataset, POM constrains both model structure and parameterization, reduces the risk of achieving accurate predictions for incorrect reasons, and increases confidence that underlying processes are sufficiently realistic for explanation and scenario-based forecasting.

2.4. Deep reinforcement learning

Reinforcement learning (Sutton and Barto, 2018) is a paradigm in artificial intelligence that enables an *agent* to interact with an *environment* and learn behavior through trial and error. The agent engages with the environment by making *observations* and performing *actions*. It receives feedback in the form of a *reward* signal—a real number that, at each time step, quantifies the positive or negative effect of the interaction on the agent. Reinforcement learning algorithms optimize models that encode how agents behave in their environment to maximize their reward. These models, called *policies*, map observations to actions. A common way to represent policies is with *policy networks*: artificial neural networks that take observations as input and return actions as output. This is the basis of *deep reinforcement learning* (Mnih et al., 2015).

Deep reinforcement learning has been applied in domains such as robotics, autonomous driving, finance, natural language processing, and healthcare (Sutton and Barto, 2018). It has also been used to develop agents that perform at high levels in games such as Pac-Man (Badia et al., 2020), Go (Silver et al., 2018), and Minecraft (Hafner et al., 2025). In environmental sciences, reinforcement learning has been applied to select optimal areas for biological conservation (Silvestro et al., 2022, 2025) and to assess the impact of economic activities (Strannegård et al., 2024a), among other topics (Zhang et al., 2021).

While reinforcement learning has been applied to predator–prey simulations in maze-like environments (Sunehag et al., 2019; Yamada et al., 2020) and to environmental decision models (Andrew et al., 2024), its use for generating animal behavior within spatially explicit agent-based ecosystem models remains largely unexplored. Here, we investigate how deep reinforcement learning can be integrated with spatially explicit agent-based ecosystem modeling. Specifically, this work contributes: (1) a framework for learning behavioral policies in agent-based ecosystem models using deep reinforcement learning; (2) integration with Pattern-Oriented Modeling for ecological validation; and (3) a case study demonstrating ecosystem-level dynamics emerging from learned behaviors.

3. Materials and methods

3.1. Modeling strategy

Given a set of ecosystems and ecological phenomena to be modeled, we first define an ontology that contains the building blocks of our ecosystem models. An *ontology* consists of (i) a set of *functional groups*, divided into decision-making and non-decision-making groups, where decision-making groups (e.g., wolves and chamois) are modeled individually as agents, and non-decision-making groups (e.g., different kinds of vegetation) are modeled collectively; (ii) a set of *agent properties*, such as age, weight, maximum speed, maximum age, energy level, hydration level, position, observation space, and action space; and (iii) a set of *cell properties*, such as biological properties (e.g., abundance of each functional group), physical properties (e.g., altitude and temperature), and landscape properties (e.g., land-cover class, with values such as rock, field, sand, or water). Cell properties may also include smell intensity of different functional groups and of organisms associated with land-cover classes.

We use spatial models to represent ecosystems at a given time. A *spatial model* consists of a grid of *cells*, where each cell has its own properties and a set of *agents*, each with its own properties.

A *behavioral model* consists of the following components for each decision-making functional group: (i) an *observation space*; (ii) an *action space*; and (iii) a *policy network* computing a function from the observation space to the action space. We use one policy network per decision-making functional group to compute the actions of each agent, based on the properties of the surrounding cells and its own internal state (Fig. 1).

A *dynamic model* is a mechanism that takes as input a spatial model and the decisions (actions) of all its agents and outputs an updated spatial model. Dynamic models are typically defined by update rules specifying how primary producers grow, the consequences of locomotion and feeding actions, and how the physical properties of the cells develop over time.

Given an ontology Ω and a dynamic model D, we use reinforcement learning to construct a behavioral model B. Specifically, we fix an architecture for the policy network of each decision-making functional group and train their parameters (weights and biases) using deep reinforcement learning. The reward signal evaluates the impact of the agent's actions—for example, actions that improve energy levels yield positive rewards, while actions that result in the agent's death yield negative rewards. The goal of training is to optimize the parameters of the policy network to maximize cumulative reward over the long run.

The policy networks of all decision-making groups can be constructed simultaneously as follows: (i) Define a class of spatial models S suitable for training; (ii) Initialize the parameters of the behavioral model B using randomization; (iii) Repeat for a fixed number of iterations: For every spatial model $S \in \mathcal{S}$, use B and D to run a simulation starting from S for a fixed number of steps, then update B using reinforcement learning. Training B on a class of spatial models $\mathcal{S}$, rather than on a single model, promotes generalization (Cobbe et al., 2019). When defining $\mathcal{S}$, one may also consider synthetic data, which have been used successfully in both supervised learning and reinforcement learning when real data are scarce (Liu et al., 2022).

An ecosystem simulation proceeds by iteratively applying the behavioral model to compute agent actions and the dynamic model to update the environment. Thus, we can define an *ecosystem model* to consist of an ontology, a spatial model, a behavioral model, and a dynamic model.

3.2. Alpine model

We constructed a general model targeting Alpine ecosystems in northern Italy, more precisely the Rhaetian Alps. Accounts of the flora and fauna of the Rhaetian Alps are provided in Lenoir et al. (2012), Gentili et al. (2010), Gazzola et al. (2007). A more detailed description of the model in the ODD format (Grimm et al., 2020) can be found in Appendix A.

Ontology

Our ontology includes three functional groups: vegetation, chamois (*Rupicapra rupicapra*), and wolves (*Canis lupus*). Vegetation represents primary producers at different altitudes and is modeled collectively at the cell level, while chamois and wolves are represented as individual agents.

The landscape consists of land and water cells. Agents perceive environmental conditions within a local neighborhood, together with internal physiological variables such as energy and hydration. Based on these observations, they select movement actions through a neural-network policy, which therefore maps environmental observations to movement actions, specifying direction and speed.

For simplicity, all agents within the same functional group share the same behavioral model. In our implementation, a single policy network is used for both the chamois and the wolf agents, with functional group identity included as an input variable. The inputs to the policy network also included the agent's internal state and properties of the surrounding 3×3 cells, for a total of 144 inputs. The purpose of using a shared network instead of two separate networks was to simplify and potentially speed up the training process, while including the functional group identity as part of the input ensured that the networks could reach diverging outputs even under the same environmental observations. The output of the network is a triplet of real numbers

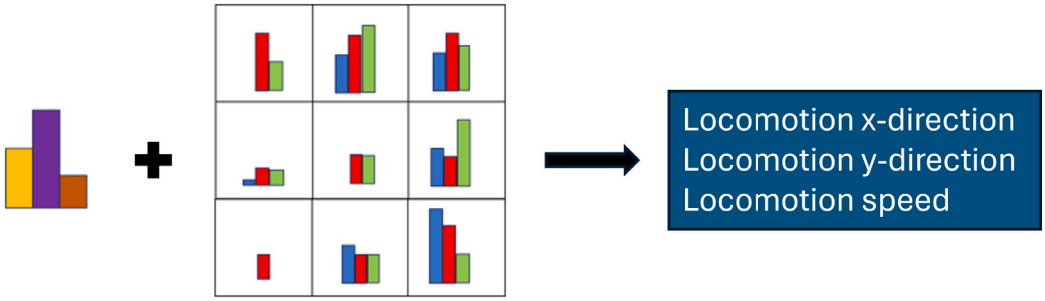


Fig. 1. Interface of a shared policy network for chamois and wolf agents. The agent observes its internal state (left), consisting of its functional group and current energy and hydration levels, as well as the properties of the surrounding 3 × 3 neighborhood (middle), including the abundance of each functional group, terrain type, and smell signals. From these observations it can infer nearby resources, obstacles, predators, and competitors. The policy network then outputs a movement action (right), represented by a speed and two values specifying direction.

$(\Delta x, \Delta y, v)$, specifying movement in the direction $(\Delta x, \Delta y)$ with speed v. The speed v is a number in $[0, 1]$ representing a fraction of the agent's maximum speed.

In our experiments, we used policy networks with a fully connected architecture and two hidden layers of 64 nodes each, using tanh as the activation function (Fig. 1).

Dynamic model

The dynamic model specifies how agent actions modify the environment and how agent states evolve over time. Key processes include vegetation consumption and regrowth, predator–prey interactions, hydration at water sources, reproduction conditional on energetic thresholds, and mortality from starvation, dehydration, predation, or age.

Agents move in continuous space, although the environment itself is represented as a grid of cells. Movement actions generated by the policy network determine the direction and speed of each agent, subject to species-specific limits on maximum movement speed.

Vegetation biomass regenerates after grazing, and wolves gain energy by killing chamois occupying the same cell. Reproduction occurs probabilistically when agents exceed species-specific energy thresholds. Environmental signals, including simplified long-range cues indicating the proximity of water or predators, are recalculated at each simulation step.

Behavioral model

To train the policy network described above, we used the reinforcement learning algorithm PPO (Schulman et al., 2017) from Stable-Baselines3 (Raffin et al., 2021). Moreover, we opted for a *homeostatic* reward signal, defined as a quadratic function of the agent's energy and hydration levels (Fig. 2). Thus, the policy network was optimized to maintain energy and hydration above a threshold of 80% of maximum values.

During training, agents interact with synthetic landscapes generated using Perlin noise (Perlin, 1985). Each training episode consists of a randomly generated 50 × 50 environment populated with chamois and wolves placed at random locations (Fig. 3). Episodes last 4000 time steps, and the full training process comprises multiple episodes totaling approximately four million simulation steps.

To simplify training and maintain computational stability, population sizes remain constant during training: when an agent dies it is replaced by a new individual of the same functional group. This was motivated by computational efficiency in the training process; however, in our subsequent simulations based on the trained models, population sizes were allowed to vary. At the beginning of each training episode, a new Perlin environment was generated. The purpose of training in multiple Perlin environments was to produce relatively versatile behavioral models, enabling the chamois and wolves to survive in a wider range of environments. After training, the learned policy networks were evaluated in simulations based on real Alpine landscapes. The training procedure is summarized in Algorithm 1.

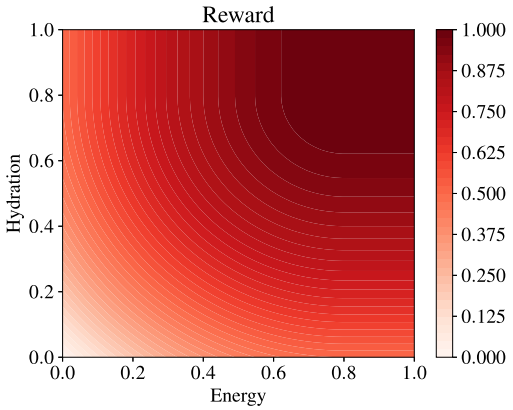


Fig. 2. Reward signal used for training the policy network, defined in terms of energy and hydration. Agents with energy and hydration levels above the saturation threshold of 0.8 do not receive additional reward from further increases.

Algorithm 1 Training process for policy networks.

```
Initialize the policy network π_θ with random parameters θ
for counter n ← 1 to 25 do
    Generate a new 50 × 50 Perlin environment
    Spawn 40 chamois and 10 wolves
    for time t ← 1 to 4000 do
        for agent i ← 1 to 50 do
            Sample action a_it ~ π_θ(a_it | s_it) for each agent
        end for
        Update environment based on actions performed by the agents:
        Update energy and hydration levels of all agents
        Kill agents based on predation, starvation, or age
        Spawn new agents to replace dead ones and maintain 40 chamois
        and 10 wolves
        Update the policy network π_θ using PPO
    end for
end for
```

The main design choices underlying the behavioral model are discussed in Appendix B. Unlike conventional agent-based ecological models, where behavioral rules are explicitly specified by the modeler, the present framework learns behavioral policies through reinforcement

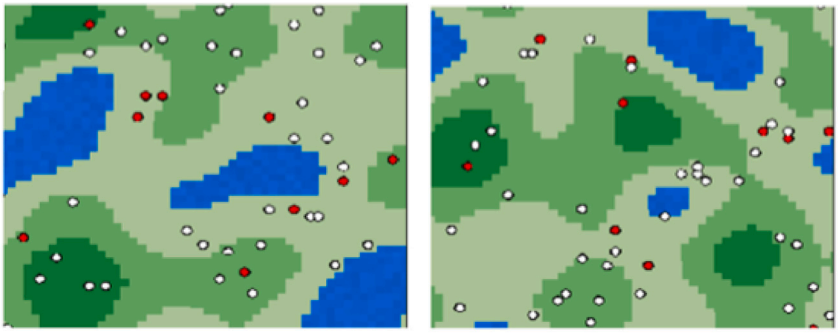


Fig. 3. Two 50 × 50 Perlin environments used for training chamois agents (white dots) and wolf agents (red dots).

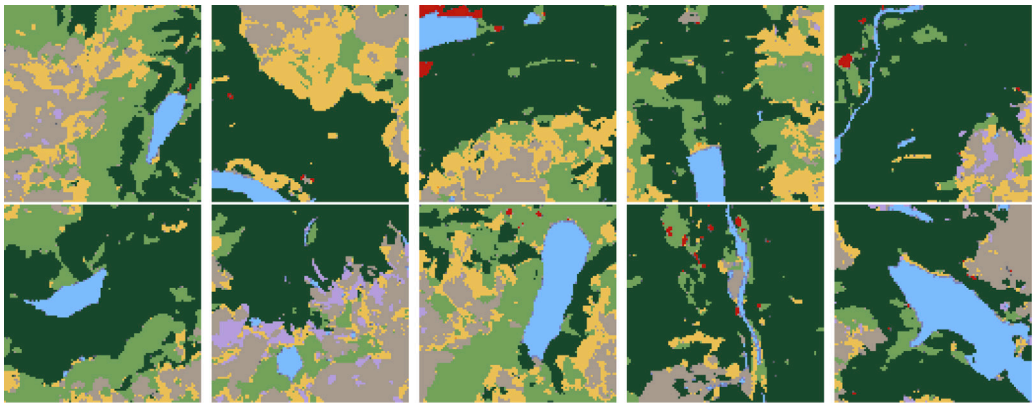


Fig. 4. Maps of ten areas in the Rhaetian Alps with 100 × 100 pixels. Light blue represents water; gray, rocks and bare soil; lilac, snow; yellow, sparsely vegetated areas; light green, pastures and grasslands; dark green, shrubs and tree-covered areas; and red, buildings and artificial settlements.

learning. Instead of defining detailed rule sets governing movement, foraging, or predator avoidance, agents develop behavioral strategies through interaction with simulated environments and a reward signal capturing basic physiological needs. As a result, the policy network replaces the rule-based decision mechanisms, allowing behavioral strategies to emerge from the training process.

3.3. Landscape model

We used land-cover maps of ten non-overlapping 4 × 4 km areas near Stelvio National Park in the Rhaetian Alps. The maps were derived from multiband satellite images from 2023, collected by the Copernicus Sentinel-2 fleet and discretized into 100 × 100 cells (Fig. 4). We used these ten maps and the Alpine model to construct several ecosystem models, including the *Stelvio model* (Fig. 5). The Stelvio model was populated with the three vegetation types, distributed according to land-cover class and initially set to 50% of their maximum biomass in each cell. Moreover, 100 chamois and 25 wolves were randomly spawned on the land cells of the model.

4. Results

In this section, we assess the extent to which the Stelvio model reproduces the seven ecological patterns introduced above. Because high-resolution time-series data for the study area were unavailable, validation is based on qualitative comparison of ecological patterns rather than statistical calibration. Thus, the model is evaluated using Pattern-Oriented Modeling rather than through direct quantitative comparison with empirical datasets.

4.1. Long-term persistence

We ran a relatively long simulation with the Stelvio model to evaluate whether chamois and wolf agents (hereafter, the chamois and wolves) could coexist for several decades in the absence of external disruptions. The simulation lasted 10,000 time steps, corresponding to 20 wolf lifetimes. The results show that chamois and wolves were able to coexist throughout the simulation (Fig. 6). The number of chamois alive at any given time during the simulation ranged from 64 to 267, whereas the number of wolves ranged from 7 to 35. Let us compare these population fluctuations to empirical data about wild chamois and wolves. Chamois and wolves have coexisted in the Rhaetian Alps for centuries, if not millennia (except for one period in the 20th century when the wolves were hunted to extinction and later repopulated the region) (Group, 2024; Gazzola et al., 2007).

Recent census data reported 82 wolf packs and 16 wolf pairs in the Italian Alps (Group, 2024), while other estimates quantified the chamois population in the Italian Alps at ~137,000 individuals, with an average density of 4.6 individuals per km^2 (Soglia et al., 2010). We did not have access to specific data for the Stelvio area, but these densities suggest that both chamois and wolf populations were larger on average in the simulations than in reality.

4.2. Predator–prey dynamics

Lotka–Volterra dynamics emerged during the simulation with the Stelvio model: chamois population surges were followed by wolf surges,

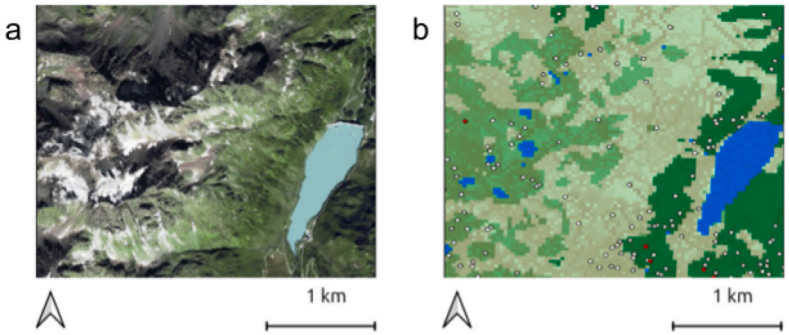


Fig. 5. (a) Aerial photograph (2020) of the Stelvio area in the Rhaetian Alps. (b) A spatial model of the same area with 100 × 100 cells. Blue represents water, and three shades of green represent vegetation types, with darker green indicating faster-growing vegetation. White and gray dots represent chamois of different ages, red dots represent wolves, and brown indicates vegetation recently grazed by chamois.

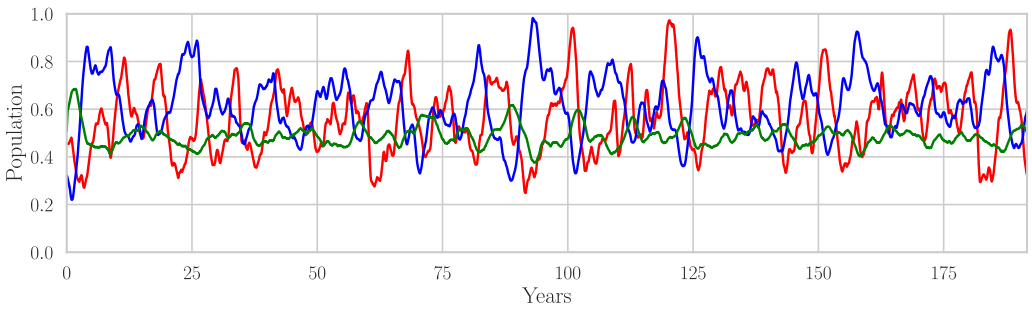


Fig. 6. Relative abundance of the functional groups during simulation with the Stelvio model. The curves show normalized amounts of vegetation (green), chamois (blue), and wolves (red) during a simulation. The value 1.0 on the *y*-axis represents 35 wolves, 267 chamois, and vegetation amounting to 100% of the carrying capacity. The value 0.5 on the *y*-axis represents half those quantities.

which triggered chamois declines and subsequent wolf declines before the cycle restarted (Fig. 6). Despite the noise, the curves suggest alternating peaks in chamois and wolves. Cross-spectral analysis using Welch's method (Welch, 1967) revealed a peak coherence of 0.834 at the dominant frequency, indicating strong coupling between predator and prey dynamics. Both species show a periodicity of ~500 time steps (about ten years) with a temporal offset of ~160 steps (about three years). Notably, these dynamics were not hand-coded but arose as emergent properties of the model.

4.3. Spatial distribution

During the simulation, the wolves were concentrated around the large lake, while the chamois were distributed more broadly across the landscape (Fig. 7).

4.4. Life history statistics

Life-history data from the Stelvio simulation showed that many chamois died young, with few reaching maximum age, here set to 10 years (Fig. 8(a)). Wolves experienced similar high juvenile mortality, but survivors often lived relatively long lives (Fig. 8(b)).

These simulated data are comparable with empirical observations, which show that, while wild wolves can live for up to 13 years and chamois up to 21 years, most individuals in both species die much younger (Mech and Boitani, 2003; Corlatti et al., 2015).

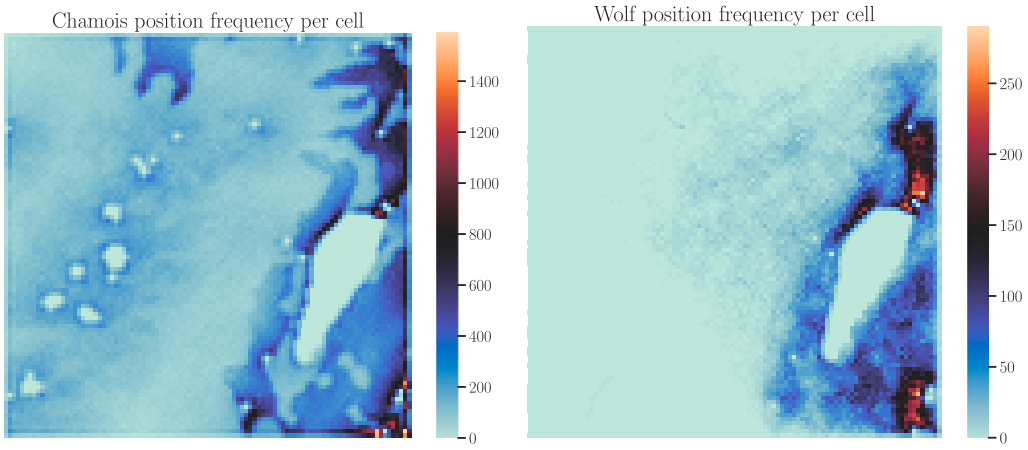
Recent studies report that the maximum lifetime reproductive success for wolves can reach up to 20 offspring per female (Stahler et al., 2024), while it is estimated at ten offspring per female for chamois (Corlatti et al., 2015). We note that sex was not explicitly represented in our model, allowing all chamois and wolves that met the reproduction criteria to reproduce. Yet, these data suggest that the life-history statistics found in our simulations are of the correct order of magnitude.

4.5. Behavioral patterns

In the Stelvio simulation, the agents displayed several behaviors consistent with those of their natural counterparts: chamois moved between vegetation and water sources, wolves hunted chamois, chamois fled from wolves, both species traveled along straight paths while avoiding lakes, and both showed a certain tendency to remain close to other individuals of the same species (Supplementary Video S1).

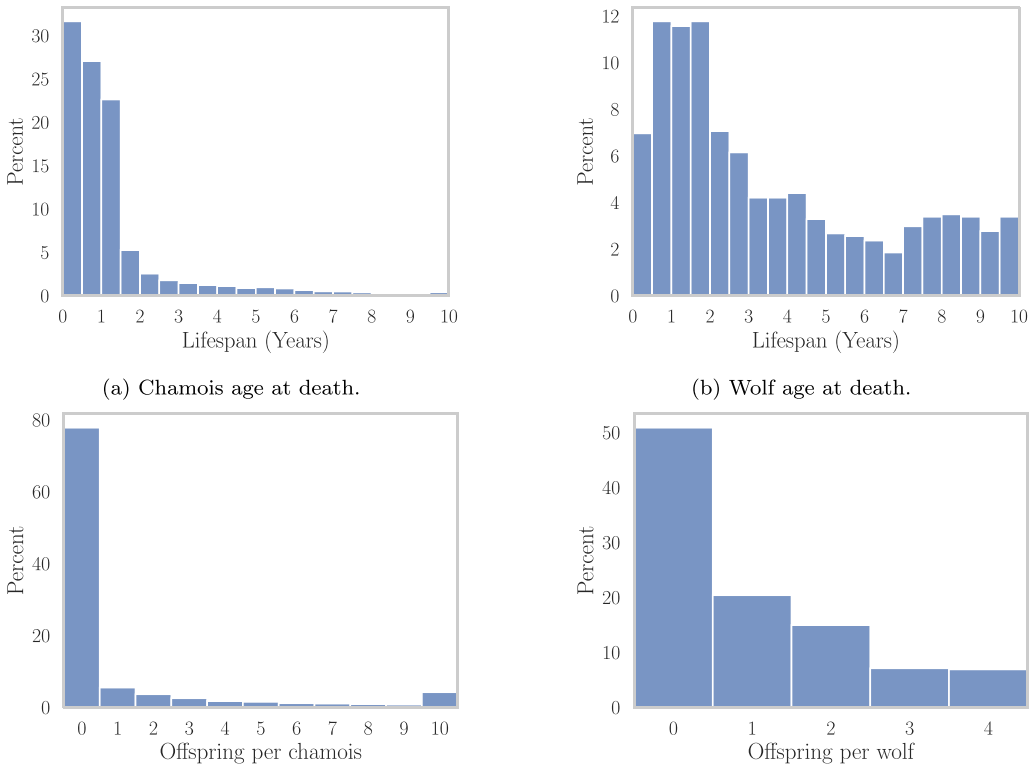
4.6. Landscape generality

To test the ability of the chamois and wolves to coexist in different landscapes, we ran several simulations on each map of ten areas of the Rhaetian Alps (Fig. 4). Recall that these spatial models were not used in the training process, which only featured artificial landscapes. Ten simulations were run on each map, starting from randomly spawned populations, giving a total of 100 simulations. Each simulation lasted up to 4000 steps—eight times the maximum wolf lifetime of 500 steps—or was stopped earlier if either species went extinct. In 83 of the 100 simulations, both species survived; in the remaining 17, wolves went extinct (Fig. 9).



(a) Heat map of the chamois. (b) Heat map of the wolves.

Fig. 7. Heat maps of the chamois (a) and wolves (b), showing their spatial distributions during the simulation with the Stelvio model.



(a) Chamois age at death. (b) Wolf age at death.

(c) Chamois offspring distribution. (d) Wolf offspring distribution.

Fig. 8. Life-history statistics from the simulation with the Stelvio model (which can be compared with those of wild wolves and chamois in the area). (a) Age at death for chamois. (b) Age at death for wolves. (c) Offspring distribution of chamois. The rightmost bar represents ten or more offspring. (d) Offspring distribution of wolves.

4.7. Disturbance resilience

We ran additional simulations to evaluate the resilience of the ten ecosystems in the Rhaetian Alps (Fig. 4) under habitat degradation, heat stress, and game hunting. Thus, we could identify tipping points where pressures would likely reduce biodiversity.

Habitat degradation—potentially caused by agriculture, forestry, mining, infrastructure projects, or urban development—was simulated by replacing areas of fast-growing vegetation with slow-growing vegetation. This mimics a shift from productive pastures to areas that recover biomass much more slowly after grazing. Degradation thus reduced resources for chamois. We tested degradation levels from 0% to

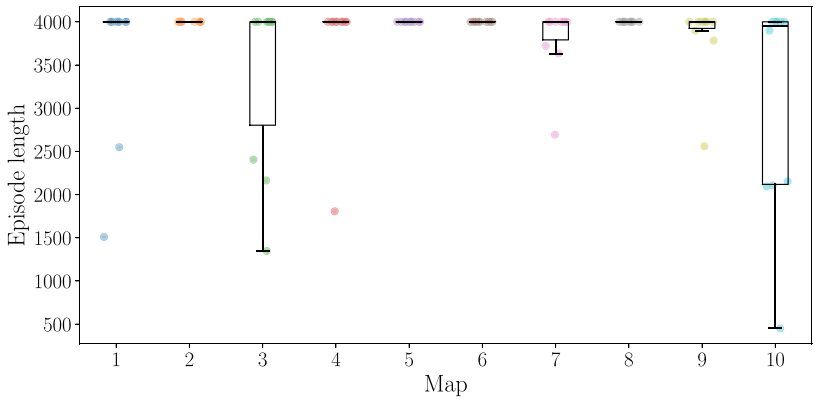


Fig. 9. Results of simulations based on the ten maps from the Rhaetian Alps. Ten runs were performed on each model, starting from randomized agent distributions. Each run lasted up to 4000 steps but was stopped earlier if wolves or chamois went extinct.

100% of the vegetated area. Each level was run on ten spatial models, with ten replicates per model, for 1100 simulations in total. The results indicate a tipping point at ~40% degradation (Fig. 10(a)).

We also explored prolonged heat stress, for example, from climate change. Increased temperature raises metabolic rates, increasing energy consumption. We tested eight levels of elevated consumption (0%–35%) for both species, with ten replicates each (80 simulations in total). The results show little effect up to ~20% increased consumption, after which extinction risk rose sharply (Fig. 10(b)). Again, wolves consistently died out first.

Game hunting was modeled by randomly removing animals at nine levels of hunting pressure (0%–90% killed every 100 steps). Wolf and chamois hunting were studied separately, with ten simulations per level, giving 90 data points for each species. The results show tipping points at ~50% for chamois (Fig. 10(c)) and ~40% for wolves (Fig. 10(d)), above which extinctions occurred. In all extinction cases, wolves died out first, even when only chamois were hunted, highlighting the greater vulnerability of predators.

4.8. Comparison with hand-coded models

We constructed one hand-coded behavioral model for the chamois and another for the wolf. The code of both models can be found in the code repository. In essence, the wolf agent follows the food gradient, avoids stepping into the water, and follows the water-gradient if it is thirsty, while the chamois agent escapes wolves if they are closer than a certain distance, follows the food gradient, avoids stepping into the water, and follows the water-gradient if it is thirsty. The hand-coded models include a handful of parameters, for example one that indicates how close the wolf should be before the chamois flee. The parameters of the hand-coded models were optimized using the TPESampler, a Bayesian optimization method available in Optuna (Akiba et al., 2019), a framework for automated parameter search.

Simulation runs were conducted using both the hand-coded model and the PPO model across the ten maps from the Rhaetian Alps (Fig. 4). In this experiment, the PPO model outperformed the hand-coded model (Fig. 11).

5. Discussion

5.1. Perception

Our animal agents received two types of input: internal signals (interoception) and external signals (exteroception). Internal signals inform agents about their physiological state, here limited to energy and hydration. External signals provide information about the quantity and location of food, water, competitors, predators, and obstacles in the environment.

External signals can be further divided into short- and long-range cues. Short-range signals provide relatively precise quantitative information about each cell in the agent's immediate neighborhood. Long-range signals, which we refer to as 'smells' for simplicity, offer information about cells that are further away. By comparing smell intensity across the nine surrounding cells, agents could approximate both the amount and the direction of the source. In reality, such remote perception may derive from multiple senses and possibly also memories of previous observations. Moreover, these smells can be critical to survival in certain environments (Strannegård et al., 2024b). Smell intensity can be defined in a number of alternative ways, for example, by using the inverse square law. To make perception more realistic, one may consider adding random noise to the signals and including models of wind and humidity.

5.2. Behavior

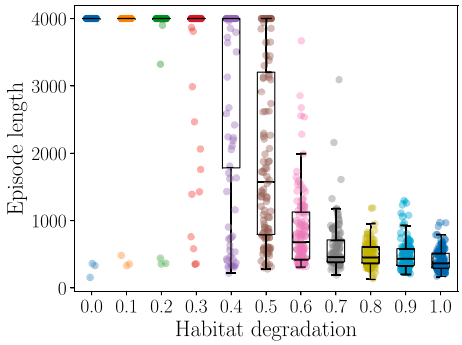
During training, agents were directly rewarded for maintaining homeostasis and indirectly for surviving, since survival enabled further opportunities to eat and drink, thereby increasing cumulative reward. Thus, the agents were trained to maintain homeostasis across a wide range of environments, as real animals must.

Our simulations showed that trained agents were able to survive in spatial models based on real geographical data—even though they had been trained only in synthetic environments. For the chamois, survival required securing food and water, navigating obstacles, and escaping predation. For wolves, it required locating, chasing, and hunting chamois. These behaviors, which are qualitatively consistent with those of their natural counterparts, were observed in our behavioral simulations (Supplementary Video S1).

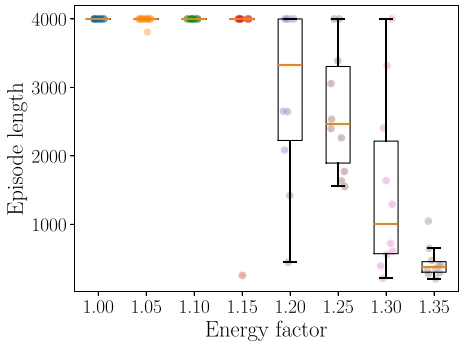
Since the animal models perceive each other's location and species, some degree of social behavior is possible in their behavioral models. Empirically, we observed in our experiments a tendency for two nearby animals of the same species to remain together (Supplementary Video S1). This observed flocking tendency may arise because co-located agents of the same species perceive similar environmental signals and, therefore, behave in similar ways. However, social or aversive behavior could emerge if there were for instance benefits in hunting success rates or competitive access to resources.

5.3. Generality

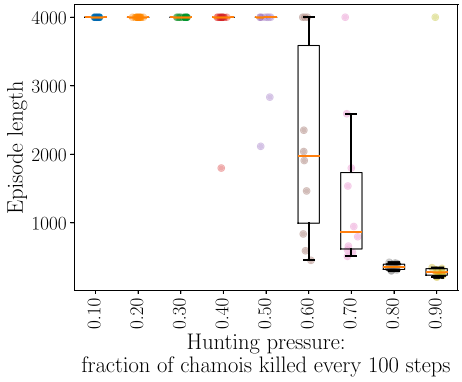
Our framework represents a first step toward using reinforcement learning for ecosystem modeling and for exploring the impact of interventions such as habitat degradation, selective hunting, or ecological restoration.



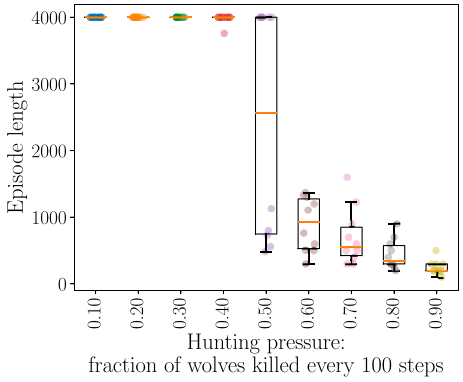
(a) Habitat degradation. There are 100 dots of each color.



(b) Heat stress. There are ten dots of each color.



(c) Chamois hunting. There are ten dots of each color.



(d) Wolf hunting. There are ten dots of each color.

Fig. 10. Results of resilience simulations. The colored dots represent individual simulation runs. The colors are not essential but help identify runs with the same level of environmental stress. The y-coordinate indicates the number of time steps the simulation lasted (episode length). The effects of (a) habitat degradation, (b) heat stress leading to increased energy consumption of the animals by a factor in the range 1.00–1.35, (c) chamois hunting, and (d) wolf hunting were tested.

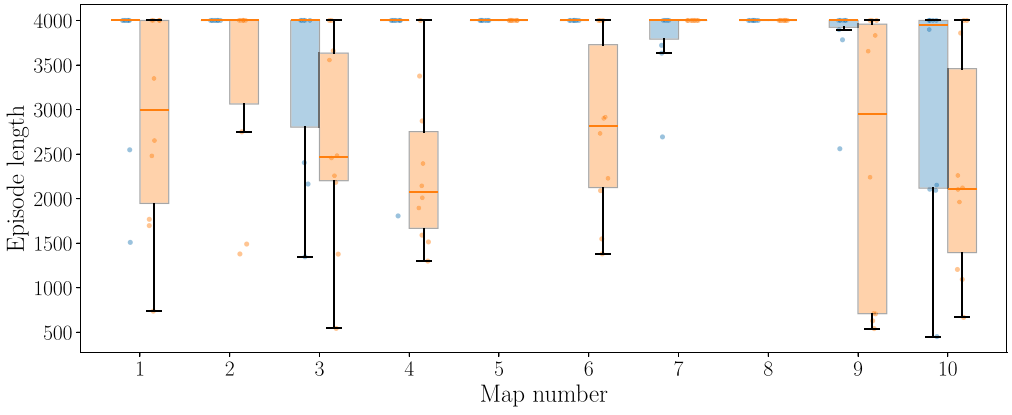


Fig. 11. Results of simulations with PPO (blue dots) and the hand-coded models (orange dots) based on the ten maps from the Rhaetian Alps. Ten runs were performed on each map, starting from randomized agent distributions. Each run lasted up to 4000 steps but was stopped earlier if wolves or chamois went extinct.

The behavioral and dynamic models in our simulations can be trained to represent different ecosystems in any geographical area, giving the framework considerable flexibility in biogeographic scope and the species included. The model can be extended with additional functional groups (e.g., scavengers or pollinators) to make food webs more complex and realistic, as well as with further land-cover classes (e.g., more vegetation types). Once the properties of new species and vegetation types are defined, the ecosystem models can be trained through deep reinforcement learning, following the same procedure used here.

Our agents were able to coexist for multiple generations across multiple environments, suggesting that the learned policies captured behavioral strategies that were sufficiently robust to support survival under varying conditions. The emergence of these strategies indicates that deep reinforcement learning is a promising approach for modeling complex social–ecological systems.

5.4. Scalability

There are limits to the number of agents that can be included, depending on available computational resources and training time. We used hundreds of agents in our simulations, while other reinforcement learning studies have scaled to millions (Yamada et al., 2020), leaving substantial room for expanding the scope of our models.

As the complexity of the ecosystem model grows, for example by adding more functional groups, environmental variables, and interaction patterns, the complexity of hand-coding behavioral models grows correspondingly. In such settings, the relative advantages of using machine-generated behavioral models—with or without optimizers that help setting the parameters—may become more pronounced.

5.5. Remaining challenges

Although our models and simulations demonstrate the potential of deep reinforcement learning for agent-based ecosystem modeling, several challenges remain in making this approach more realistic and scalable.

Real ecosystems may contain thousands of species, millions of individuals, and countless known and unknown mechanisms driving animal behavior and ecosystem dynamics. Ecosystem functioning and responses to change depend heavily on processes that are not fully understood and are often modeled as stochastic, such as reproduction, death, migration, genetic mutation, and climate or landscape change (Lande et al., 2003).

Because of this complexity, ecosystem models must rely on simplifying assumptions; for example, when defining food webs. The assumptions should be chosen with the model's intended use in mind, whether to explore ecological patterns, evaluate management interventions, or predict resilience and tipping points. In our case, assumptions included a food web limited to a few representative species, imposed spatial boundaries without migration, and a reproduction model that does not fully reflect biological complexity.

Although the current temporal resolution allows us to run simulations spanning many generations, it should ideally be finer when cells measure 40 × 40 meters since wolves and chamois can move much more than 40 m in a single week. This issue cannot be resolved simply by redefining the time scale, however. In fact, if a wolf must chase a chamois for hundreds or even thousands of time steps before being rewarded, the reward signal may be substantially delayed relative to the actions that produced it. In such situations, the credit assignment problem (Sutton and Barto, 2018) may arise, making it difficult for the learning algorithm to determine which earlier actions contributed to the eventual reward.

In future developments, calibration with real-world values will be crucial for increasing realism. This calibration must be performed individually for each geographical area to reflect its local conditions. In particular, empirical knowledge of the energy intake from feeding, movement costs, reproduction rates, and natural mortality rates plays an important role in calibration. For some systems, such data may be available, but in many cases, parameter calibration will likely rely on approximations. Other features that can be considered in future developments include probabilistic hunting success, noise in animal perception, and more realistic reproduction models.

6. Conclusions

We developed a flexible agent-based modeling strategy in which behavioral rules are not hand-coded but learned from data. We used this strategy to construct a model of an Alpine ecosystem and ran multiple simulations, indicating that seven ecological patterns of Alpine ecosystems were robustly reproduced in the model. We also found that behavioral models learned through deep reinforcement learning outperformed the hand-coded models in our simulation setup.

These findings suggest that machine-generated behavioral models can capture ecologically relevant dynamics in spatially explicit, multi-species systems. Such models provide a foundation for systematically exploring the potential effects of human interventions—including habitat degradation, selective hunting, and ecosystem restoration. Although further validation across ecosystems is needed, integrating machine learning with agent-based modeling represents a promising step toward more flexible and realistic ecosystem simulations for conservation and management.

CRediT authorship contribution statement

Claes Strannegård: Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Michał Palak:** Writing – review & editing, Software, Methodology, Formal analysis, Data curation. **Niklas Engsner:** Writing – review & editing, Visualization, Supervision, Software, Methodology, Investigation, Formal analysis. **Alice Stocco:** Writing – review & editing, Investigation, Data curation. **Alexandre Antonelli:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition. **Daniele Silvestro:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Claes Strannegard reports financial support was provided by Sten A Olsson Foundation for Research and Culture. Claes Strannegard reports financial support was provided by Foundation Erik and Lily Philipsons Memorial Fund. Daniele Silvestro reports financial support was provided by Swedish Research Council. Alexandre Antonelli reports financial support was provided by Swedish Research Council. Alexandre Antonelli reports financial support was provided by Swedish Foundation for Strategic Environmental Research. Daniele Silvestro reports financial support was provided by Swedish Foundation for Strategic Environmental Research. Claes Strannegard reports a relationship with Ecotwin Sweden AB that includes: board membership and equity or stocks. Daniele Silvestro reports a relationship with Ecotwin Sweden AB that includes: equity or stocks. Daniele Silvestro reports a relationship with Captain Ltd that includes: equity or stocks. Alexandre Antonelli reports a relationship with Captain Ltd that includes: equity or stocks. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Design details

This appendix describes the ecosystem model in greater detail using the ODD (Overview–Design concepts–Details) protocol (Grimm et al., 2020).

A.1. Overview

The model is a spatially explicit agent-based ecosystem simulation in which animal behavior emerges from deep reinforcement learning. Individual wolves and chamois interact with a landscape composed of grid cells containing vegetation and environmental features. Instead of using hand-coded behavioral rules, agents learn movement strategies through interaction with the environment and a reward signal which promotes physiological homeostasis.

Purpose

The purpose of the model is methodological: to investigate whether reinforcement learning can generate ecologically plausible behavioral strategies in a spatially explicit agent-based ecosystem model. The case study represents a simplified Alpine ecosystem with wolves (predators), chamois (herbivores), and vegetation resources.

The model is not intended as a fully calibrated representation of a specific site. Rather, it is used to assess whether learned behavior can reproduce ecological patterns characteristic of Alpine predator–prey systems. Model evaluation follows a Pattern-Oriented Modeling approach and focuses on long-term predator–prey coexistence, predator–prey cycles, realistic spatial habitat use, plausible life-history statistics, behavioral patterns, landscape generality, and resilience to disturbances such as habitat degradation, heat stress, and hunting.

Entities, state variables, and scales

Entities. Two types of entities are represented.

- *Cells*: landscape grid cells containing environmental information such as vegetation biomass, land-cover class, water presence, and simplified long-range cues (e.g., distance to predators or water).
- *Agents*: individual wolves and chamois that move, interact, and reproduce. Vegetation is represented collectively as biomass within cells rather than as individual plants.

Agent state variables. Animal agents include the following state variables:

- species identity
- spatial position (x, y)
- age
- energy level
- hydration level
- maximum movement speed (age-dependent)
- alive/dead status
- reproductive eligibility

Sex is not explicitly represented.

Cell variables. Each cell stores:

- biomass values for three vegetation functional groups
- land-cover class and water presence
- simplified long-range cues such as distance to the nearest wolf or water cell

Spatial and temporal scales. Training environments consisted of 50×50 synthetic landscapes generated with Perlin noise. Evaluation simulations used a real Alpine landscape, represented as 100×100 grids covering approximately 4×4 km (40 m spatial resolution). One time step in the simulation corresponds approximately to one week. Agents have a maximum lifespan of 500 steps (about 10 years). Training episodes last 4000 steps, while evaluation simulations may run up to 10,000 steps.

Process overview and scheduling

At each time step, agents perceive their internal state and local environment, compute an action using the policy network, and move accordingly. Environmental interactions and demographic processes are then applied. Actions are computed from the state at time t and then applied synchronously.

The simulation proceeds as follows:

1. Agents perceive internal and environmental signals.
2. The policy network generates a movement action.
3. Agents move according to the chosen action and maximum speed.
4. Automatic interactions occur: chamois graze vegetation, wolves kill chamois in the same cell, and agents drink when adjacent to water.
5. Energy and hydration are updated.
6. Mortality occurs (starvation, dehydration, drowning, predation, or old age).
7. Reproduction may occur if energetic thresholds are met.
8. Vegetation regenerates.
9. Long-range cues are recalculated.
10. During training simulations, reinforcement learning updates the behavioral policy.

During training, population sizes remain fixed (40 chamois and 10 wolves) by replacing dead individuals. Evaluation simulations allow population sizes to vary.

A.2. Design concepts

Basic principles. The model combines agent-based modeling with deep reinforcement learning. Behavioral strategies emerge from optimization of a reward signal promoting homeostasis rather than from manually specified rule sets.

Emergence. Population dynamics, predator–prey cycles, spatial habitat use, and ecosystem responses to disturbance arise from interactions between learned behavior, environmental structure, and resource distribution.

Adaptation. During training, agents adapt behavior through reinforcement learning. After training, the learned policy is fixed during evaluation simulations.

Objectives. The reward function encourages agents to maintain energy and hydration levels above approximately 80% of their maximum values.

Learning. A single shared neural-network policy is used for both species, with species identity included as an input variable. The network contains 144 input variables, two hidden layers with 64 neurons each, hyperbolic tangent activation functions, and three output variables representing movement actions. The policy is optimized using Proximal Policy Optimization (PPO).

Prediction. Agents do not maintain an explicit internal predictive model of future environmental states. Their actions are reactive outputs of the learned policy based on current observations.

Sensing. Agents perceive their internal state and environmental properties within a 3×3 neighborhood, along with simplified long-range cues representing distances from water or the distance between the agent and a predator.

Interaction. Interactions include grazing, predation, drinking, and indirect competition for space and resources. No explicit social rules are imposed, although clustering may emerge indirectly through shared environmental responses.

Stochasticity. Randomness arises from synthetic landscape generation, initial agent placement, reproduction probability, exploration during reinforcement learning, and disturbance events in scenario experiments.

Outputs. Model outputs recorded for analysis include species abundances over time, spatial distributions, life-history statistics, episode lengths, and system responses to environmental disturbances.

A.3. Details

Initialization. Training simulations begin with randomly generated Perlin landscapes. The set-up procedure creates 40 chamois and 10 wolves. The neural-network parameters are initialized randomly.

Evaluation simulations use Alpine land-cover maps, in which vegetation biomass initially equals 50% of maximum values. Populations start with 100 chamois and 25 wolves randomly placed in cells.

Input data. Two types of landscapes are used: synthetic Perlin-noise landscapes for training and real Alpine land-cover maps derived from Sentinel-2 satellite imagery (2023).

Submodels. Each agent receives a 144-dimensional observation vector consisting of internal variables and environmental properties from a 3×3 neighborhood.

The actions of the individual agents only include movement, whereas feeding (grazing or predation), drinking, reproduction, and death are automatic processes. Cells store simplified signals representing Euclidean distance to environmental features, such as water or other individual agents, and mediate the interaction between the individual agents and the conditions in the surrounding cells. With regard to vegetation, it regenerates over approximately 20 to 200 time steps, depending on the vegetation type, which is a cell property.

Movement. The neural network outputs $(\Delta x, \Delta y, v)$ describing movement direction and speed. Coordinates are updated at each time step according to the equation:

$$(x_{t+1}, y_{t+1}) = (x_t + \Delta x \cdot v,\ y_t + \Delta y \cdot v).$$

Maximum agents' speeds increase during early life, remain stable during adulthood, and decrease toward maximum lifespan.

Feeding and predation. Chamois occupying vegetation cells can feed on vegetation by consuming plant biomass, as stored in the cell properties, and gain up to 3% of their own maximum energy. Wolves can predate and kill chamois that occupy the same cell and gain up to 18% of their maximum energy as a consequence of predation. *Drinking.* When interacting with cells that host water, individual agents can drink and restore hydration to 100%. *Reproduction.* Agents reproduce with a probability of 0.1 if energy exceeds species-specific thresholds (95% of maximum for wolves and 50% for chamois). *Death.* Agents die from starvation, dehydration, drowning, predation (chamois), or reaching maximum age.

Reward and learning. A quadratic homeostatic reward encourages energy and hydration above 80% of maximum values. Behavioral policies are optimized using PPO.

Appendix B. Design choices

B.1. Machine learning algorithm

Several approaches could in principle be used to optimize policy-network parameters, including gradient-free methods such as Approximate Bayesian Computation (Beaumont, 2010) and Adaptive Random Search (Mania et al., 2018), as well as gradient-based reinforcement learning algorithms such as PPO, TRPO, and TQC (Raffin et al., 2021). Because gradient-based methods scale more favorably to high-dimensional policy networks, we adopted deep reinforcement learning as our main approach.

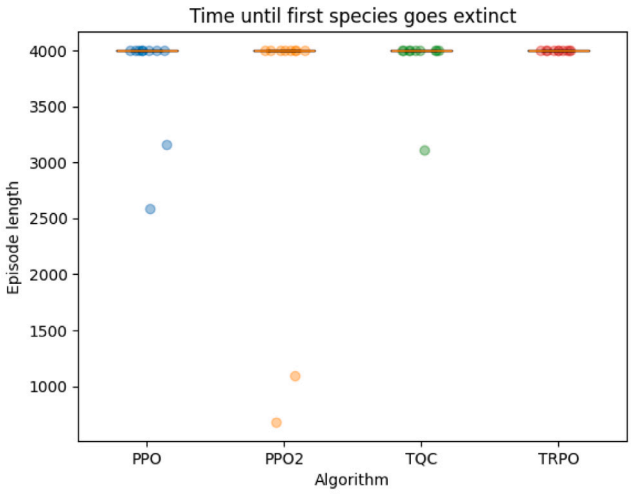


Fig. B.12. Episode lengths for deep reinforcement learning agents on ten previously unseen Perlin worlds. Each dot represents one simulation and there are ten dots of each color.

Next, to select a suitable reinforcement learning algorithm, we conducted a simple study. We trained three policy networks, each with two hidden layers, using the algorithms PPO, TRPO, and TQC. We also trained a fourth policy network without hidden layers using PPO, here called PPO2 (Wong et al., 2024). After training, the four networks were tested on ten spatial models created with Perlin-noise. For each network, a 4000-step simulation was run on each of the ten spatial models. Of the 40 simulations, all but five lasted the full 4000 steps (Fig. B.12).

We further compared the four networks in a long simulation where chamois and wolves with different (heritable) policy networks coexisted in the same spatial model. In this test, PPO2 agents died out first (Supplementary Video S2). Based on these results, we chose the PPO algorithm from Stable-Baselines3 (Raffin et al., 2021). PPO is a widely used reinforcement learning algorithm that performs above human level in many video games (Schulman et al., 2017). Among others, it excels at Pac-Man, where the task is to survive in a maze-like environment by collecting food and avoiding obstacles and enemies, somewhat reminiscent of animals navigating ecosystems. PPO enables effective decision-making strategies to be learned through repeated interaction with the environment. The algorithm improves the decision policy step by step based on feedback on how well previous actions performed, while deliberately limiting how much the policy can change at each update to maintain stable learning. This is done using a simple mechanism that discourages overly large updates, helping the system learn reliably even in noisy and non-stationary settings. PPO separates the tasks of choosing actions and evaluating their outcomes, which improves learning efficiency and consistency.

B.2. Reward signal

Many reward signals are possible in the context of agent-based ecosystem modeling with deep reinforcement learning (Strannegård and Engsner, 2025). One possibility is to use a version of fitness (Grimm and Railsback, 2013). Since fitness only provides relatively sparse feedback on individual behavior, the sparse reward problem (Sutton and Barto, 2018) is likely to arise, however.

To select a suitable reward signal, we compared two main alternatives: the *homeostatic* reward (defined above) and the *survival* reward,

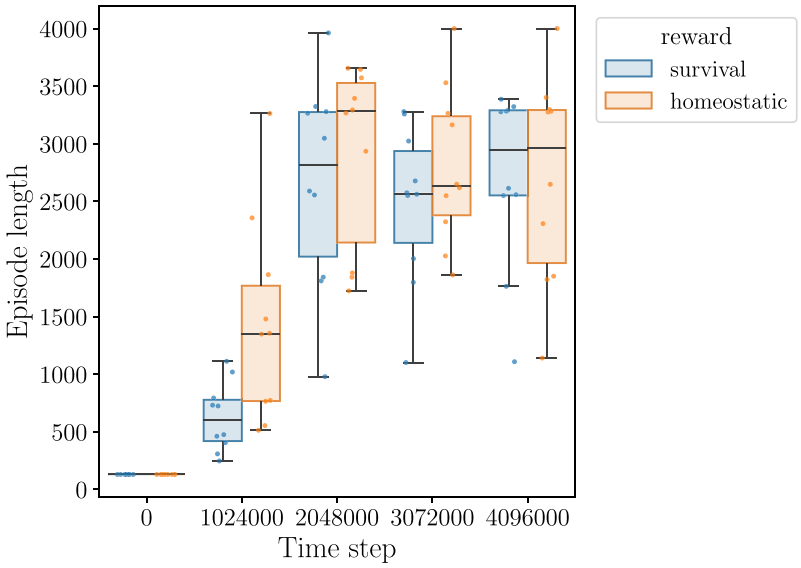


Fig. B.13. Episode length at five training stages for agents trained with homeostatic (blue) and survival (orange) rewards. Each dot denotes a test run lasting up to 4000 steps, with earlier termination in cases of extinction.

which grants +1 for each step as long as no functional group goes extinct (Supplementary Video S3). The homeostatic reward primarily encourages eating and drinking, while the survival reward encourages long-term coexistence. The survival reward is quite general, as it requires no reward engineering and can be applied across functional groups—for example, an anteater might learn to eat ants and a hummingbird to drink nectar by trial and error, without explicitly rewarding those actions.

Training with either reward improved the agents' ability to coexist in Perlin worlds over time (Fig. B.13). At the end of the training processes, the mean performance of the homeostatic and the survival agents was very similar. This similarity may reflect that homeostatic reward indirectly favors survival (to collect more reward in the future), while survival reward indirectly encourages feeding and drinking (along with predator avoidance and efficient navigation). In the end, we chose the homeostatic reward since it led to somewhat faster learning.

B.3. Hyperparameter settings

Table B.1 shows the hyperparameter settings that were used when training the policy network. In most cases, the default settings from the PPO implementation of Stable-Baselines3 were preserved. Additional hyperparameters in the dynamic model, such as maximum animal age, energy intake and consumption, and reproduction probability, also substantially influenced the population dynamics and were manually fine-tuned to maintain stable populations.

B.4. Training process

Randomness plays a key role in the training process (Algorithm 1), e.g., when generating training environments based on Perlin noise. Thus, the behavioral models are partially dependent on random noise. To test the robustness and repeatability of the training process, we trained seven different policy networks using Algorithm 1 and ran 190 simulations with each of the seven models. The simulations lasted up to 4000 steps but ended earlier if some species went extinct. For each model, we computed the average population size and the average lifespan of chamois and wolves. The results indicate moderate variations between the seven models, with the medians essentially falling within a 25% margin of the medians of the other models (Fig. B.14). We also tracked the locations of the chamois and wolves during the simulations with the seven models and calculated the corresponding heat maps. The heat maps vary between models, but the animal distributions are clearly correlated with the resources in all cases (Fig. B.15).

Table B.1
Hyperparameters used for training the policy network.

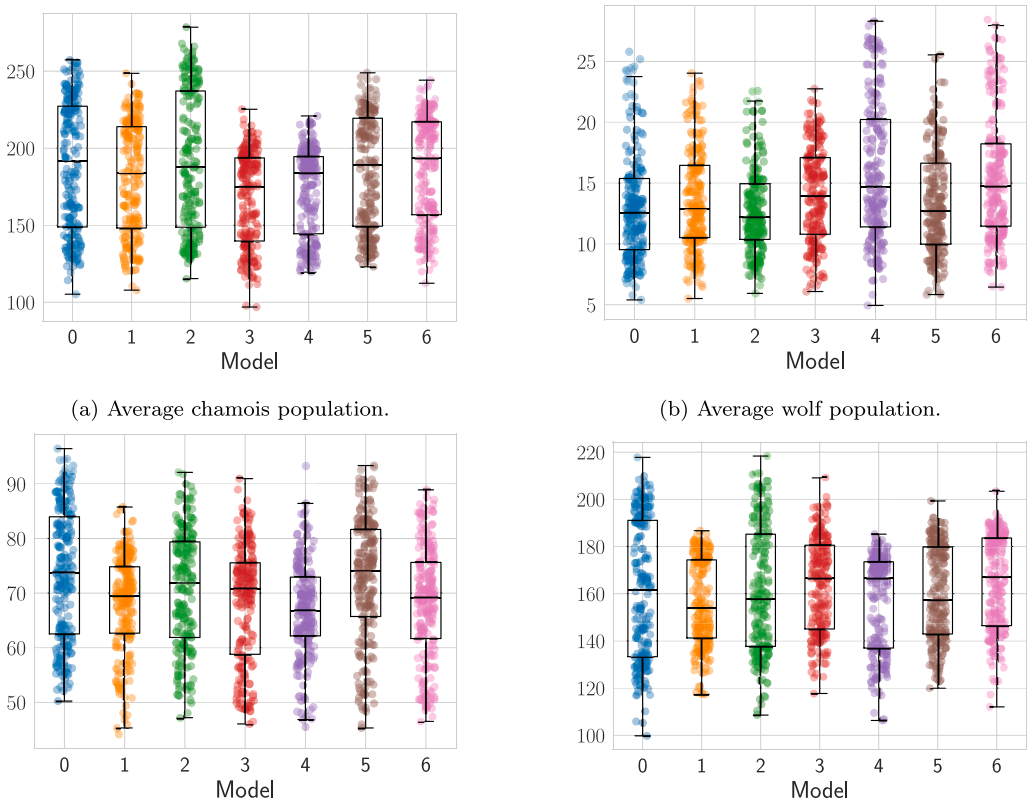
Hyperparameter	Value
`inputs`	144
`outputs`	3
`hidden_layers`	2
`hidden_layer_size`	64
`activation_function`	tanh
`learning_rate`	0.0003
`n_steps`	512
`batch_size`	64
`n_epochs`	10
`gamma`	0.99
`gae_lambda`	0.95
`clip_range`	0.2

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ecoinf.2026.103819.

Data availability

The data and source code supporting the results of this study are publicly available on Zenodo and can be accessed at https://doi.org/10.5281/zenodo.19221981. The repository includes the trained policy networks, as well as documentation and instructions required to reproduce the main experiments.



(a) Average chamois population.

(b) Average wolf population.

(c) Average chamois lifespan.

(d) Average wolf lifespan.

Fig. B.14. Results of robustness simulations with seven different behavioral models. Each dot represents one simulation. There are 190 dots of each color. (a) average chamois population (individuals); (b) average wolf population (individuals); (c) average chamois lifespan (time steps); and (d) average wolf lifespan (time steps).

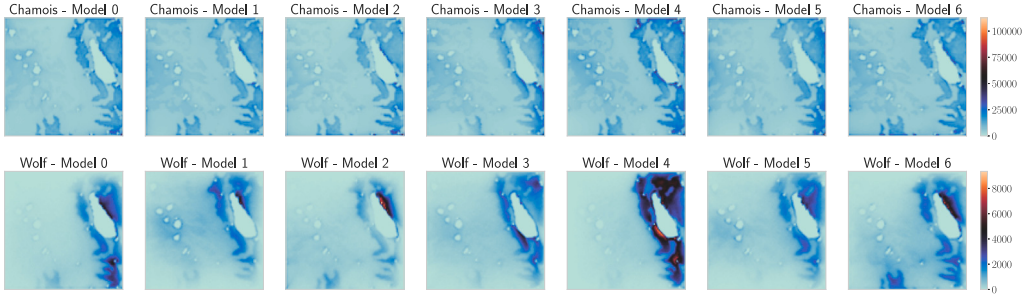


Fig. B.15. Heat maps showing the aggregated distribution over time of the chamois and wolves with the seven different behavioral models.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 2623–2631.

Andrew, K., Zia, A., Rizzo, D., 2024. Integrating deep reinforcement learning into agent-based models for predicting farmer adaptation under policy and environmental variability. In: Arai, K. (Ed.), Intelligent Systems and Applications. Springer Nature Switzerland, Cham, pp. 221–238.

Antonelli, A., 2022. The Hidden Universe: Adventures in Biodiversity. University of Chicago Press.

Arditi, R., Ginzburg, L.R., 1989. Coupling in predator-prey dynamics: ratio-dependence. J. Theoret. Biol. 139, 311–326.

Badia, A.P., Piot, B., Kapturowski, S., Sprechmann, P., Vitvitskyi, A., Guo, Z.D., Blundell, C., 2020. Agent57: Outperforming the atari human benchmark. In: International Conference on Machine Learning. PMLR, pp. 507–517.

Beaumont, M.A., 2010. Approximate bayesian computation in evolution and ecology. Annu. Rev. Ecol. Evol. Syst. 41, 379–406. http://dx.doi.org/10.1146/annurev-ecolsys-102209-144621.

Brigolin, D., Dal Maschio, G., Rampazzo, F., Giani, M., Pastres, R., 2009. An individual-based population dynamic model for estimating biomass yield and nutrient fluxes through an off-shore mussel (mytilus galloprovincialis) farm. Estuar. Coast. Shelf Sci. 82, 365–376.

Cobbe, K., Klimov, O., Hesse, C., Kim, T., Schulman, J., 2019. Quantifying generalization in reinforcement learning. In: International Conference on Machine Learning. PMLR, pp. 1282–1289.

Colléter, M., Valls, A., Guitton, J., Lyne, M., Arreguín-Sánchez, F., Christensen, V., Gascuel, D.D., Pauly, D., 2013. EcoBase: a Repository Solution to Gather and Communicate Information from Ewe Models (Ph.D. thesis). Fisheries Centre, University of British Columbia, Canada.

Corlatti, L., et al., 2015. Reproductive senescence in female chamois. Oecologia 178, 187–196. http://dx.doi.org/10.1007/s00442-015-3232-9.

Crooks, A., Malleson, N., Manley, E., Heppenstall, A., 2019. Agent-Based Modelling and Geographical Information Systems: a Practical Primer. SAGE Publications.

Daily, G.C., Matson, P.A., 2008. Ecosystem services: from theory to implementation. Proc. Natl. Acad. Sci. 105, 9455–9456.

DeAngelis, D.L., Diaz, S.G., 2019. Decision-making in agent-based modeling: A current review and future prospectus. Front. Ecol. Evol. 6, 237.

DeAngelis, D.L., Grimm, V., 2014. Individual-based models in ecology after four decades. F1000Prime Rep. 6.

Díaz, S., Demissew, S., Carabias, J., et al., 2015. The IPBES conceptual framework — connecting nature and people. Curr. Opin. Environ. Sustain. 14, 1–16. http://dx.doi.org/10.1016/j.cosust.2014.11.002.

Gazzola, A., Avanzinelli, E., Bertelli, I., Tolosano, A., Bertotto, P., Musso, R., Apollonio, M., 2007. The role of the wolf in shaping a multi-species ungulate community in the Italian western alps. Ital. J. Zoöl. 74, 297–307.

Geary, W.L., Bode, M., Doherty, T.S., Fulton, E.A., Nimmo, D.G., Tulloch, A.I., Tulloch, V.J., Ritchie, E.G., 2020. A guide to ecosystem models and their environmental applications. Nat. Ecol. Evol. 4, 1459–1471.

Gentili, R., Armiraglio, S., Rossi, G., Sgorbati, S., Baroni, C., 2010. Floristic patterns, ecological gradients and biodiversity in the composite channels (central alps, Italy). Flora-Morphology Distrib. Funct. Ecol. Plants 205, 388–398.

Grimm, V., Railsback, S.F., 2012. Pattern-oriented modelling: a 'multi-scope' for predictive systems ecology. Phil. Trans. R. Soc. B 367, 298–310. http://dx.doi.org/10.1098/rstb.2011.0180.

Grimm, V., Railsback, S.F., 2013. Individual-Based Modeling and Ecology. Princeton University Press.

Grimm, V., Railsback, S.F., Vincenot, C.E., Berger, U., Gallagher, C., DeAngelis, D.L., Edmonds, B., Ge, J., Giske, J., Groeneveld, J., et al., 2020. The odd protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism. J. Artif. Soc. Soc. Simul. 23.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W.M., Railsback, S.F., Thulke, H.H., Weiner, J., Wiegand, T., DeAngelis, D.L., 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310, 987–991. http://dx.doi.org/10.1126/science.1116681.

Group, W.A., 2024. The Wolf Alpine Population in 2020–2024 Over 7 Countries. Technical Report for Life wolfalps eu Project LIFE18 NAT/IT/000972, Action C4. Technical Report, Life Wolfalps EU project team, https://www.lifewolfalps.eu/en/archivi/technical-reports-en/.

Hafner, D., Pasukonis, J., Ba, J., Lillicrap, T., 2025. Mastering diverse control tasks through world models. Nature 1–7.

Hagen, O., Flück, B., Fopp, F., Cabral, J.S., Hartig, F., Pontarp, M., Rangel, T.F., Pellissier, L., 2021. gen3sis: A general engine for eco-evolutionary simulations of the processes that shape earth's biodiversity. PLoS Biol. 19, e3001340.

Hughes, A.C., Grumbine, E.R., 2023. The kunming-montreal global biodiversity framework: what it does and does not do, and how to improve it. Front. Environ. Sci. 11, 1281536.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589.

Kaul, H., Ventikos, Y., 2015. Investigating biocomplexity through the agent-based paradigm. Brief. Bioinform. 16, 137–152.

de Koning, K., Broekhuijsen, J., Kühn, I., Ovaskainen, O., Taubert, F., Endresen, D., Schigel, D., Grimm, V., 2023. Digital twins: dynamic model-data fusion for ecology. Trends Ecol. Evolut. 38, 916–926.

Lande, R., Engen, S., Saether, B.E., 2003. Stochastic Population Dynamics in Ecology and Conservation. Oxford University Press, USA.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436.

Lenoir, J., Svenning, J.C., Dullinger, S., Pauli, H., Willner, W., Guisan, A., Vittoz, P., Wohlgemuth, T., Zimmermann, N., Gégout, J.C., 2012. The alps vegetation database—a geo-referenced community-level archive of all terrestrial plants occurring in the Alps. Biodivers. Ecol 4, 331–332.

Lindkvist, E., Wijermans, N., Daw, T.M., Gonzalez-Mon, B., Giron-Nava, A., Johnson, A.F., van Putten, I., Basurto, X., Schlüter, M., 2020. Navigating complexities: agent-based modeling to support research, governance, and management in small-scale fisheries. Front. Mar. Sci. 6, 733.

Liu, C., Ventre, C., Polukarov, M., 2022. Synthetic data augmentation for deep reinforcement learning in financial trading. In: Proceedings of the Third ACM International Conference on AI in Finance. pp. 343–351.

Lotka, A.J., 1925. Elements of Physical Biology. Williams & Wilkins.

Mania, H., Guy, A., Recht, B., 2018. Simple random search of static linear policies is competitive for reinforcement learning. In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 1–22, URL: https://proceedings.neurips.cc/paper_files/paper/2018/file/7634ea65a4e6d9041cfd3f7de18e334a-Paper.pdf.

McLane, A.J., Semeniuk, C., McDermid, G.J., Marceau, D.J., 2011. The role of agent-based models in wildlife ecology and management. Ecol. Model. 222, 1544–1556.

Mech, L.D., Boitani, L. (Eds.), 2003. Wolves: Behavior, Ecology, and Conservation. University of Chicago Press, Chicago.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al., 2015. Human-level control through deep reinforcement learning. Nature 518, 529–533.

Nelson, E., Mendoza, G., Regetz, J., Polasky, S., Tallis, H., Cameron, D.R., Chan, K.M.A., Daily, G.C., Goldstein, J., Kareiva, P.M., Lonsdorf, E., Naidoo, R., Ricketts, T.H., Shaw, M.R., 2009. Modeling multiple ecosystem services, biodiversity conservation, commodity production, and tradeoffs at landscape scales. Front. Ecol. Environ. 7, 4–11. http://dx.doi.org/10.1890/080023.

Perlin, K., 1985. An image synthesizer. ACM Siggraph Comput. Graph. 19, 287–296.

Price, I., Sanchez-Gonzalez, A., Alet, F., Andersson, T.R., El-Kadi, A., Masters, D., Ewalds, T., Stott, J., Mohamed, S., Battaglia, P., et al., 2025. Probabilistic weather forecasting with machine learning. Nature 637, 84–90.

Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., Dormann, N., 2021. Stable-baselines3: Reliable reinforcement learning implementations. J. Mach. Learn. Res. 22, 1–8.

Rollins, N.D., Barton, C.M., Bergin, S., Janssen, M.A., Lee, A., 2014. A computational model library for publishing model documentation and code. Environ. Model. Softw. 61, 59–64.

Royama, T., 2012. Analytical Population Dynamics, vol. 10, Springer Science & Business Media.

Rykiel, Edward J.J., 1996. Testing ecological models: the meaning of validation. Ecol. Model. 90, 229–244. http://dx.doi.org/10.1016/0304-3800(95)00152-2.

Saraiva, S., van der Meer, J., Kooijman, S., Ruardij, P., 2014. Bivalves: From individual to population modelling. J. Sea Res. 94, http://dx.doi.org/10.1016/j.seares.2014.06.004.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., 2017. Proximal policy optimization algorithms. arXiv preprint: 1707.06347.

Siekmann, I., Osborne, J.M., 2023. Editorial: Do individuals matter? - Individual-based versus population-based models applied to biology and health. Front. Appl. Math. Stat. 9, http://dx.doi.org/10.3389/fams.2023.1272392, URL: https://www.frontiersin.org/articles/10.3389/fams.2023.1272392.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., et al., 2018. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. Science 362, 1140–1144.

Silvestro, D., Goria, S., Groom, B., Jacobsson, P., Sterner, T., Antonelli, A., 2025. Using artificial intelligence to optimize ecological restoration for climate and biodiversity. BioRxiv, 2025-01.

Silvestro, D., Goria, S., Sterner, T., Antonelli, A., 2022. Improving biodiversity protection through artificial intelligence. Nat. Sustain. 5, 415–424.

Soetaert, K., Herman, P.M., 2009. A Practical Guide to Ecological Modelling: Using R as a Simulation Platform. Springer.

Soglia, D., Rossi, L., Cauvin, E., Citterio, C., Ferroglio, E., Maione, S., Meneguz, P.G., Spalenza, V., Rasero, R., Sacchi, P., 2010. Population genetic structure of alpine chamois (rupicapra r. rupicapra) in the italian alps. Eur. J. Wildl. Res. 56, 845–854.

Stahler, D.R., et al., 2024. Lifetime reproductive characteristics of gray wolves. J. Mammal. 105, 1–13. http://dx.doi.org/10.1093/jmammal/gyae032.

Strannegård, C., Engsner, N., 2025. Reward functions for agent-based ecosystem modeling. In: ALIFE 2025 Workshop on Mortal Agents. Kyoto, Japan, pp. 1–6, URL: https://openreview.net/forum?id=tE1s2hkg9C.

Strannegård, C., Engsner, N., Lindgren, R., Olsson, S., Endler, J., 2024a. AI tool for exploring how economic activities impact local ecosystems. In: Arai, K. (Ed.), Intelligent Systems and Applications. Springer Nature Switzerland, Cham, pp. 690–709.

Strannegård, C., Engsner, N., Ulfsbäcker, S., Andreasson, S., Endler, J., Nordgren, A., 2024b. Survival games for humans and machines. Cogn. Syst. Res. 86, 101235.

Sunehag, P., Lever, G., Liu, S., Merel, J., Heess, N., Leibo, J.Z., Hughes, E., Eccles, T., Graepel, T., 2019. Reinforcement learning agents acquire flocking and symbiotic behaviour in simulated ecosystems. In: Artificial Life Conference Proceedings. MIT Press, pp. 103–110.

Sutton, R.S., Barto, A.G., 2018. Reinforcement Learning: An Introduction. MIT Press.

Swartzman, G.L., 1979. Simulation modeling of material and energy flow through an ecosystem: methods and documentation. Ecol. Model. 7, 55–81.

Vieira, V.M., Engelen, A.H., Huanel, O.R., Guillemin, M.L., 2022. An individual-based model of the red alga agarophyton chilense unravels the complex demography of its intertidal stands. Front. Ecol. Evol. 10, 797350.

Walters, C., Christensen, V., 2007. Adding realism to foraging arena predictions of trophic flow rates in ecosim ecosystem models: shared foraging arenas and bout feeding. Ecol. Model. 209, 342–350.

Weiskopf, S.R., Myers, B.J.E., Arce-Plata, M.I., Blanchard, J.L., Ferrier, S., Fulton, E.A., Harfoot, M., Isbell, F., Johnson, J.A., Mori, A.S., Weng, E., Harmáčková, Z.V., Londoño Murcia, M.C., Miller, B.W., Pereira, L.M., Rosa, I.M.D., 2022. A conceptual framework to integrate biodiversity, ecosystem function, and ecosystem service models. BioScience 72, 1062–1073. http://dx.doi.org/10.1093/biosci/biac074.

Welch, P., 1967. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. IEEE Trans. Audio Electroacoust. 15, 70–73.

Wilensky, U., Rand, W., 2015. An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo. The MIT Press, URL: http://www.jstor.org/stable/j.ctt17kk851.

Wong, A., de Nobel, J., Bäck, T., Plaat, A., Kononova, A.V., 2024. Solving deep reinforcement learning benchmarks with linear policy networks. arXiv preprint arXiv:2402.06912.

Yamada, J., Shawe-Taylor, J., Fountas, Z., 2020. Evolution of a complex predator-prey ecosystem on large-scale multi-agent deep reinforcement learning. In: 2020 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 1–8.

Zhang, W., Valencia, A., Chang, N.B., 2021. Synergistic integration between machine learning and agent-based modeling: A multidisciplinary review. IEEE Trans. Neural Networks Learn. Syst..