Multi-Objective Inverse Preference Inference for Archaeological Equifinality: A Forward Pareto Sweep on a Settlement Agent-Based Model

Simon Jaxy[1] , **Luca Vincke**[1] , **Bram Silue**[1] , **Dries Daems**[3] , **Ralf Vandam**[2] and **Pieter Libin**[1]
[1]AI Lab, Department of Computer Science, Vrije Universiteit Brussel
[2]MGC (Archaeology, Environmental Changes & Geo-Chemistry), Vrije Universiteit Brussel
[3]Spatial Digital Humanities, Vrije Universiteit Amsterdam
simon.jaxy@vub.be

Abstract

Different subsistence priorities can produce indistinguishable spatial settlement patterns, an effect known as *equifinality* that neither correlative archaeological predictive models nor agent-based simulations resolve. We re-pose this as multi-objective inverse preference inference, whose forward pass enumerates Pareto-optimal policies under each preference and whose inverse pass recovers the preference that best explains the held-out record. We deliver the forward half: **SAGAscape-RL** learns a policy family on a Sagalassos settlement agent-based model (southwest Turkey, 801 years) by sweeping a resource preference simplex. Under stationary settlement, the held-out site record concentrates posterior mass on roughly three of seven swept preferences, demonstrating equifinality; switching the evaluation target to a correlative predictive surface shifts the posterior to a different region of the simplex, a metric-dependent identifiability. Allowing communities to move, abandon, and colonise unlocks a wooded western cluster at a food-security and held-out-fit cost, the latter qualified by sparse site data and heuristic movement mechanics. The Stationary and Mobile Pareto fronts are mutually non-dominated on the shared weight vectors: action-space extension shifts the reachable Pareto region rather than improving along it.

1 Introduction

Two families of model dominate computational archaeology: correlative archaeological predictive models (APMs) such as the Locally-Adaptive Model of Archaeological Potential (LAMAP [Carleton *et al.*, 2012; Rondeau *et al.*, 2022]), which forecast site locations from terrain covariates; and agent-based simulations (ABMs) [Kohler and Gumerman, 2000], which encode explicit decision rules. APMs forecast accurately but do not represent the decision processes that generate the observed pattern. ABMs are mechanistic but their hand-tuned parameters are brittle: small perturbations can collapse simulated populations in ways inconsistent with the archaeological record [Axtell *et al.*, 2002]. Neither resolves the equifinality of spatial archaeological signatures [Verhagen and Whitley, 2012]: distinct underlying processes can produce indistinguishable site distributions, so the mechanism that generated a given pattern cannot be recovered from the pattern alone.

We re-pose equifinality as a multi-objective inverse-decision problem. Classical inverse reinforcement learning [Ng and Russell, 2000; Abbeel and Ng, 2004] treats *degeneracy* (the existence of many rewards consistent with the same observed policy) as the central obstacle to recovery. Archaeological equifinality is its analogue under a known multi-objective basis [Roijers *et al.*, 2013], with the weighting $\mathbf{w}$ as the inferred quantity [Kishikawa and Arai, 2021]. For equifinality, degeneracy becomes the object of interest: the breadth of the equivalence class of preferences consistent with the observed pattern is what we set out to characterise. Each weighting expresses an archaeological preference whose optimal policy is a Pareto-optimal behaviour over a per-resource viability basis (formally defined in Section 3.1). The forward problem is to enumerate the Pareto front and score each policy against the held-out record; the inverse problem is to recover the weighting that best explains it. This study delivers the forward half.

We instantiate this with **SAGAscape-RL**,[1] an agent-based simulation of resource competition spanning the Iron Age Archaic to Achaemenid–Hellenistic transition (801 years) at Sagalassos (southwest Turkey). The model inherits the environmental landscape of SAGAscape [Boogers and Daems, 2022]: landscape rasters, community structure, and period activation schedule are carried over, and the resource regeneration mechanics follow the original up to one reserve constant (see supplementary material); the spatial actions described in Section 3.2 extend the environment. Each community pursues three resources each year to meet its consumption requirements (formalised in Section 3.1); the resulting per-resource adequacy scores form the three-dimensional viability basis over which a preference $\mathbf{w}$ ranges. In place of SAGAscape's hand-coded resource priorities, we learn the per-resource workday allocations with a parameter-shared multi-agent policy [Schulman *et al.*, 2017; Yu *et al.*, 2022]: each community distributes its annual workday budget across the three resources, isolating allocation as the locus of be-

[1]Code: https://github.com/simomoxy/sagascape-rl.

havioural learning. Where SAGAscape uses a single hand-coded allocation rule across its parameter sweeps, we sweep the preference $\mathbf{w}$ itself over the simplex and score each resulting Pareto-optimal policy against the held-out Achaemenid–Hellenistic site record (training uses Iron Age Archaic only), placing that rule-based policy in a broader family of alternative subsistence strategies. Communities form a sequential common-pool resource dilemma [Perolat *et al.*, 2017; Leibo *et al.*, 2017]: they compete over regenerating food and wood stocks together with a non-regenerating clay stock, with individual harvesting imposing collective costs.

Our central contributions are: (i) a re-framing of archaeological equifinality as multi-objective inverse preference inference, with the forward Pareto sweep delivering the held-out likelihood surface any inverse procedure must consume; (ii) a multi-objective Markov-game formalisation of an established archaeological ABM, with a forward Pareto sweep that traces the front by training one policy per preference $\mathbf{w}$; (iii) empirical evidence on real archaeological data: under stationary settlement the held-out record concentrates posterior mass on roughly three of seven swept weight vectors (ESS $\approx 2.85/7$), an equifinality signature, while evaluation against the LAMAP suitability surface prefers a different region of the simplex (metric-dependent identifiability); extending to a mobile action space traces a second Pareto front mutually non-dominated with the stationary front, extending the reachable behavioural manifold.

2 Related Work

2.1 Multi-objective decision-making

A multi-objective Markov decision process replaces a scalar reward with a reward vector $\mathbf{r} \in \mathbb{R}^d$ and the solution concept becomes a Pareto front of non-dominated policies [Roijers *et al.*, 2013; Van Moffaert and Nowé, 2014]. Selecting a point on the front requires a preference; linear scalarisation $r = \mathbf{w}^\top \mathbf{r}$ with $\mathbf{w}$ on the simplex recovers the convex hull of the front [Van Moffaert and Nowé, 2014]. Forward MORL traces the front by training one policy per preference, by conditioning a single policy on the preference vector, or by evolving a population of policies; quality is measured by hypervolume [Roijers *et al.*, 2013] and by the ε-indicator. In one such instance, Reymond et al. [2024] apply forward MORL to COVID-19 mitigation policy.

The forward problem asks which policies are Pareto-optimal under a given preference; the inverse problem asks which preference rationalises observed behaviour. Inverse reinforcement learning [Ng and Russell, 2000; Abbeel and Ng, 2004] recovers a reward from demonstration trajectories; maximum-entropy IRL [Ziebart *et al.*, 2008] addresses strict-optimality degeneracy through a softmax-rational behaviour model. Multi-objective extensions recover a preference *weighting* over a given basis: Kishikawa et al. [2021] estimate weight vectors from trajectories whose reward decomposition is known. Trajectory-based methods are not directly applicable to archaeological data, which supplies spatial outcomes (site distributions) rather than agent trajectories. In such settings, the Pareto front itself, scored against held-out outcomes, functions as a likelihood surface over preferences.

Beyond the single-agent picture, parameter-sharing PPO (MAPPO) [Schulman *et al.*, 2017; Yu *et al.*, 2022] trains a single shared policy conditioned on each agent's local observation across cooperative agents. When agents are self-interested and share a regenerating resource, the setting becomes a sequential common-pool resource (CPR) dilemma [Perolat *et al.*, 2017; Leibo *et al.*, 2017], whose four social-outcome metrics [Perolat *et al.*, 2017] (efficiency, equality, sustainability, peace) standardise evaluation. Archaeology combines these threads as a multi-agent CPR observed only through spatial outcomes; its dominant computational traditions, reviewed next, stop short of preference recovery.

2.2 Computational archaeology

Two computational traditions address the location problem with different aims: prediction (APMs) and process simulation (ABMs). Correlative archaeological predictive models (APMs) such as LAMAP [Carleton *et al.*, 2012; Rondeau *et al.*, 2022] forecast site probability from terrain covariates; recent extensions add deep representations under label scarcity to address data-side brittleness [Jaxy *et al.*, 2025]. APMs nonetheless do not explicitly represent the decision processes that generate site patterns, even if they implicitly capture them in learned mappings; data-side improvements do not close that gap. Agent-based simulations (ABMs) [Kohler and Gumerman, 2000] instead posit that macro-level settlement patterns emerge from micro-level decisions of individual agents. Axtell et al. [2002] validate Kayenta Anasazi simulations against the archaeological record across time periods but report that hand-tuned parameters are highly sensitive, with small perturbations producing population collapse inconsistent with the record. Altaweel et al. [2015] show that settlement hierarchies emerge from a gravity model with social and kinship attractors and external-benefit multipliers. SAGAscape [Boogers and Daems, 2022] models the Sagalassos survey area (southwest Turkey) as a resource-only landscape under a rule-based policy in which communities harvest resources from a local territory; parameters are hand-tuned. In contrast to standard ABM practice, we do not tune parameters to fit a rule-based policy to the archaeological evidence; instead we treat each community's behaviour as optimal under a latent preference, which the inverse step seeks to recover. This disentangles the decision rule from what each community actually valued.

3 Methodology

3.1 Decision problem

We formalise the forward problem as a partially observable multi-objective Markov game with N self-interested agents, $\mathcal{M} = (N, S, \{O_n\}, \{A_n\}, P, \{r_n \in \mathbb{R}^3\}, \mathbf{w} \in \Delta^2, \gamma)$. Agents share no reward, do not communicate, and are coupled only through the common-pool resource dynamics in P; explicit inter-agent interaction is left to future work. For each fixed $\mathbf{w}$, a policy $\pi_\theta(a \mid o)$ is trained to maximise expected discounted scalarised return; sweeping $\mathbf{w}$ over a discretisation of the simplex traces a Pareto front whose held-out fit is the forward likelihood surface.

The global state S comprises per-patch resource fields, per-community stocks and home positions, and the tick counter that drives the period activation schedule (per-patch and per-community symbols are defined in Section 3.2). Observations $O_n = \mathbb{R}^{12}$ summarise own-state and local-environment features around the agent's territory; the global state is hidden, hence partial observability. Actions $A_n = \mathbb{R}^7$ are continuous: four allocation components (three workday weights, one per resource, plus a global aggressiveness factor) scale the harvest mechanism and the food-security target; three further components score spatial decisions (move, colonise, abandon). The environment P squashes each component through the logistic $\sigma(\cdot)$ before applying it to the harvest mechanism; condition-specific masks disable subsets of components (Section 4.3). Per-resource viability is

$$v_k(n) = \mathrm{clip}\left(\frac{\mathrm{stock}_k(n) - \mathrm{req}_k(n)}{\mathrm{req}_k(n)},\ -3,\ 3\right) \in [-3, 3] \tag{1}$$

for $k \in \{\mathrm{food}, \mathrm{wood}, \mathrm{clay}\}$ (the -3 floor prevents gradient saturation in deep deficit). Collecting the per-resource scores into the viability triple $\mathbf{v}_n = (v_{\mathrm{food}}, v_{\mathrm{wood}}, v_{\mathrm{clay}}) \in [-3, 3]^3$, we instantiate the vector reward of the MDP as $r_n := \mathbf{v}_n$, giving the scalarised multi-objective reward $r_n^{\mathbf{w}} = \mathbf{w}^\top \mathbf{v}_n$ with $\mathbf{w} \in \Delta^2 = \{\mathbf{w} \in [0,1]^3 : \sum_k w_k = 1\}$. Full state, observation, and per-component action definitions are in the supplementary material. The three resources compete for shared land and a single annual workday budget, so no allocation maximises all of them at once and the trade-off surface is a Pareto front. Each weight vector $\mathbf{w}$ encodes a distinct subsistence priority and selects one Pareto-optimal policy; sweeping $\mathbf{w}$ over the simplex enumerates a family of competing archaeological hypotheses rather than positing how any community internally weighs its resources, and the spread of weights that fit the record equally well is the equifinality we set out to quantify.

3.2 Environment instantiation

SAGAscape-RL re-implements the SAGAscape [Boogers and Daems, 2022] environment as a JAX-accelerated grid-world, following the resource mechanics, regeneration dynamics, and period activation schedule of the original NetLogo model (one regeneration constant deviates; see supplementary material). Tick-1 aggregate stocks agree with the NetLogo reference (full validation protocol and table in the supplementary material); the spatial actions described below extend beyond NetLogo. Communities are self-interested and share a regenerating landscape without inter-community trade, a sequential common-pool resource dilemma [Perolat *et al.*, 2017; Leibo *et al.*, 2017]. We deliberately couple communities only through the shared resource, testing whether resource optimisation alone reproduces the observed distribution. Future work will add richer social dynamics such as trade or kinship-driven attractors [Altaweel, 2015].

Patches. The Sagalassos survey area is discretised into an $H \times W$ raster $\mathcal{P}$ at 100 m/px, one patch per raster cell. Each patch p carries food fertility $f_p \in [0, K_p^{\mathrm{food}}]$ (with K_p^{food} the soil-quality cap inherited from SAGAscape), wood stock $w_p \in [0, K_p^{\mathrm{wood}}]$ regenerating against its own forest-cover cap K_p^{wood}, clay quantity $q_p \geq 0$ (non-regenerating), and a harvest yield multiplier $\lambda_p \in [0, 1]$. The choice of λ is an experimental switch (Section 4.3).

Communities. Time is discrete; one tick is one calendar year. Up to $N_{\max}$ communities coexist, each with a home patch, a Dijkstra-reachable territory $\mathcal{T}_n$ (the shortest-walking-time disc on the patch graph, with edge weights from Tobler's slope-aware hiking function [Tobler, 1993]), stocks $(F_n, W_n, C_n) \in \mathbb{R}^3_{\geq 0}$, and an activation tick at which the community joins under the period schedule.

Step. Each tick, every active community n harvests food, wood, then clay sequentially, scoring its territory patches and exhausting its workday budget WD_n on the highest-scoring ones. The food target is $\mathrm{target}_n^{\mathrm{food}} = \mathrm{req}_n^{\mathrm{food}} \cdot \mathrm{gpg} \cdot (1 + u_n\, \sigma(a_n^{\mathrm{aggr}}))$, with $u_n \sim \mathrm{Uniform}(0, 1)$ and post-threshing yield ratio $\mathrm{gpg} = 6/5$ (one harvested unit yields $5/6$ units of consumable food, the remaining $1/6$ retained as seed grain). The target floors at $\mathrm{req}_n^{\mathrm{food}} \cdot \mathrm{gpg}$ when aggression is zero, and $\sigma(a_n^{\mathrm{aggr}}) \in [0, 1]$ scales the random security margin. The raw patch yield y_p is f_p, w_p, or q_p depending on the resource queue; for food and wood it is modulated by λ_p as $\tilde{y}_p = y_p \cdot \lambda_p$, while clay extraction is geological and not scaled by λ_p ($\tilde{y}_p = y_p$). Within each queue, patches are ranked by $s_p = \tilde{y}_p / c_{n,p}$, where $c_{n,p}$ is the Tobler walking time [Tobler, 1993] from patch p to community n's home patch, and exploited in descending order. After exploitation, annual consumption is deducted from (F_n, W_n, C_n) (full update rules in the supplementary material). Food fertility and forest stock recover by saturating Verhulst logistic growth [Vogels *et al.*, 1975] against K_p^{food} and K_p^{wood} respectively. Bad harvests follow a Poisson distribution with rate $1/T_{\mathrm{bh}}$ ($T_{\mathrm{bh}} = 5$ years), halving food stocks for one tick; wildfires ignite from a fire-return-rate raster with stochastic drought.

Spatial transitions. *The thresholds and cost coefficients below are engineering choices rather than theory-derived constants; their status and the future work needed to ground them are discussed in Section 6.4.* The policy emits three spatial scores $a_n^{\mathrm{move}}, a_n^{\mathrm{col}}, a_n^{\mathrm{abn}}$ interpreted through $\sigma(\cdot)$. *Move* relocates the community to the highest-scoring unoccupied patch within radius $R = 20\,\mathrm{px}$ at workday cost $\kappa_m = 0.15\,\mathrm{WD}_n$. *Colonise*, gated on $\sigma(a_n^{\mathrm{col}}) > 0.7$, food surplus $F_n \geq 1.5\,\mathrm{req}_n^{\mathrm{food}}$, and a 50-tick cooldown, founds a daughter community, transferring 30% of stocks at cost $\kappa_c = 0.30\,\mathrm{WD}_n$. *Abandon*, gated on $\sigma(a_n^{\mathrm{abn}}) > 0.8$ and a 3-tick food-deficit streak, deactivates the community.

3.3 Implementation

All N agents share a single Gaussian actor-critic with parameters θ; at each step every active agent n samples $a_n \sim \pi_\theta(\cdot \mid o_n)$ from its own observation. This is parameter sharing across agents [Yu *et al.*, 2022]: a single network is trained on stacked per-agent transitions. The critic shares the trunk and predicts a decentralised value $\hat{V}^{\mathbf{w}}(o_n)$; communities have no real-time access to distant resource stocks, so a decentralised critic is archaeologically appropriate. The trunk is a two-layer multilayer perceptron of width 64 with $\tanh$ activations and orthogonal initialisation. Training uses clipped PPO [Schulman *et al.*, 2017] in JAX with JIT-compiled simulation steps

and vmap'd rollouts [Bradbury *et al.*, 2018]; discount factor $\gamma = 0.999$ gives meaningful credit across the 801-year episode ($0.999^{800} \approx 0.449$). Full PPO hyperparameters are in the supplementary material.

4 Experimental Details

4.1 Data

The simulation covers the Sagalassos survey area (southwest Turkey), an archaeologically rich Mediterranean region with steep elevation gradients and heterogeneous vegetation. The survey spans over 30 years of fieldwork identifying settlements across multiple periods, from Iron Age through Roman Imperial [Poblome, 2023]. The 396×799 raster at 100 m/px is derived from a Digital Elevation Model and its terrain covariates (slope, aspect, curvature, water proximity), resampled to the simulation grid. Confirmed site locations come from the Sagalassos survey project [Poblome, 2023], extended by Willett's held-out survey additions [2022]. Community parameters (population, resource requirements, start ticks) are drawn from SAGAscape's settlement-type distributions [Boogers and Daems, 2022] (initial size classes in the supplementary material).

4.2 Baseline

The rule-based SAGAscape policy [Boogers and Daems, 2022] fixes all resource weights and aggressiveness at 1.0 with security factor $\eta_n \sim \mathrm{Uniform}(1, 2)$, and the inherited greedy patch-selection mechanism exploits highest-scoring patches until targets or workday budgets are exhausted. We run 50 stochastic replicates for the baseline.

4.3 Conditions

We compare two main conditions and one ablation. **Stationary** fixes communities to their home patches: only the four allocation components of a_n are active; the three spatial components are masked, and the suitability multiplier is disabled ($\lambda_p = 1$ for all p). **Mobile** activates all seven action components and instantiates the suitability multiplier from LAMAP [Carleton *et al.*, 2012; Rondeau *et al.*, 2022], a correlative model that scores cells by terrain covariates (DEM, slope, curvature, water proximity). LAMAP enters only as the harvest-yield multiplier λ_p; it never appears in the per-resource viability triple $\mathbf{v}_n$ that defines the reward. The **LAMAP control** holds the action space of Stationary fixed and switches the suitability multiplier on, isolating environmental priming from the action-space extension that distinguishes Stationary from Mobile.

The Iron Age Archaic LAMAP raster is computed on the same grid as the DEM at 100 m/px; no interpolation is applied. The Achaemenid–Hellenistic LAMAP surface and the corresponding confirmed sites are withheld and used only for post-hoc evaluation.

4.4 Experimental Design and Evaluation

All training runs simulate the Iron Age Archaic to Achaemenid–Hellenistic transition on the Iron Age Archaic landscape: 16 Iron Age communities are active from year 0, 6 Achaemenid communities join at year 450, and 10 Hellenistic communities join at year 650 (a total of $N = 32$ communities; $N_{\max} = 2N = 64$ reserves capacity for colonisation under Mobile). Episodes are 801 years long. The reference learned policy trains with equal resource weights $\mathbf{w} = (1/3, 1/3, 1/3)$ for 200 episodes across five seeds (42–46); each Pareto-sweep point trains the same five seeds under a different $\mathbf{w} \in \Delta^2$. Mean viability $\bar{v}$ over the last 20 episodes is the logging summary; per-episode community-level metrics use the same window.

Spatial metrics. A *harvest density map* is a per-cell aggregate of mean harvest activity (summed across resources) accumulated within an episode, mean-averaged across the last 20 training episodes and five seeds, summed across communities and normalised to $[0, 1]$. Under Stationary each community's footprint is bounded by its fixed home territory; under Mobile the footprint reflects post-relocation positions because the move action updates patch indices. Held-out F_1 is computed between the harvest density and confirmed Achaemenid–Hellenistic sites (binary mask dilated by three cells for positional tolerance); a uniform random-over-land map gives $F_1 = 0.024$ as the chance baseline. Within-period diagnostics include per-community Spearman ρ and per-resource Soft-IoU against the training-period LAMAP surface; Soft-IoU against the held-out Achaemenid–Hellenistic LAMAP surface is reported as a secondary cross-condition spatial metric.

Spatial agreement. We compare the policy's harvest density h and the Iron Age Archaic LAMAP surface λ by overlaying their binarised maps at a heuristic threshold $\tau = 0.4$ (normalised units; threshold sensitivity in the supplementary material). Each land cell falls into one of four overlap classes: *agreement* (both above τ), *LAMAP-only* (only LAMAP above τ), *LAMAP-missed* (only harvest density above τ), and *no prediction* (both at or below τ). We use the overlay as a comparison instrument, identifying candidates for more targeted surveys in future field work.

Pareto-front quality. Hypervolume (HV) [Roijers *et al.*, 2013], the standard MO front-quality measure, is the volume of the dominated region between the front and a reference point in the worst direction; we use $(-3, -3, -3)$, the per-resource viability clip floor of Equation 1, so every reachable policy dominates the reference and HV is non-degenerate across the simplex. The additive ε-indicator,

$$I_\varepsilon(A, B) = \max_{b \in B} \min_{a \in A} \max_i (b_i - a_i), \tag{2}$$

is the smallest scalar shift that lets every point of B be matched by a point of A in every objective; $I_\varepsilon(A, B) \leq 0$ iff $A \succeq B$.

CPR social outcomes. We report three of Pérolat et al.'s four social-outcome metrics [Perolat *et al.*, 2017]: efficiency (mean reward), equality ($1 - \mathrm{Gini}$ over per-community mean viability), and sustainability (mean per-resource viability across active communities); peace is not applicable in the absence of explicit conflict actions.

All runs execute on a single NVIDIA A100 GPU.

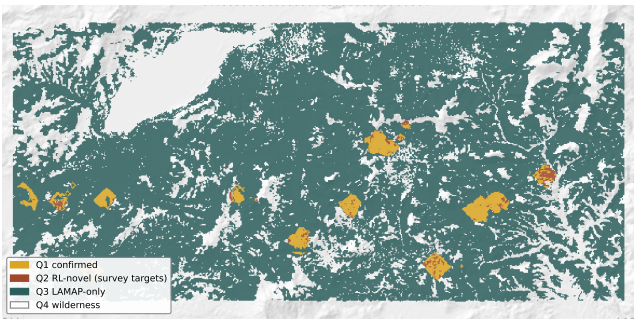


Figure 1: Stationary spatial agreement classes (seed 43, last 20 episodes, threshold $\tau = 0.4$) on the Sagalassos DEM, computed by overlaying the binarised harvest density and the binarised Iron Age Archaic LAMAP surface: *agreement* (1.7%); *LAMAP-missed*, agent-active and LAMAP-silent (0.3%, cells the correlative predictor does not flag); *LAMAP-only* (59.8%); *no prediction* (38.2%).

5 Experimental Results

5.1 Stationary: Raw Landscape

Learned policy and its spatial footprint. Table 1 compares the rule-based SAGAscape baseline against the PPO equal-weights policy (mean over five seeds, last 20 episodes). PPO improves food viability by $+0.189$ and wood viability by $+0.112$ over the baseline; clay is essentially unchanged. Mean reward improves from -0.117 to -0.020 but does not break even: the Iron Age Archaic landscape structurally under-supports the full 32-community population once the Achaemenid and Hellenistic waves activate at years 450 and 650. The substantive spatial finding is the per-community Spearman correlation $\rho = +0.522^{**}$ between viability and the LAMAP suitability surface (Table 1): the policy never sees LAMAP, so the alignment is mechanistic evidence that landscape attractors of resource efficiency co-vary with those of archaeological suitability. Wood correlates negatively because food-rich (open) and wood-rich (forested) terrain are spatially incompatible; clay correlates positively despite being unresponsive to behavioural choice. The overlap classes (Figure 1) split the land mask as follows: the policy is active on only a fraction of LAMAP-suitable terrain (*LAMAP-only*, 59.8%); the policy-active and LAMAP-silent cells form the substantive output (*LAMAP-missed*, 0.3%), cells the correlative predictor itself does not flag; *agreement* accounts for 1.7% and *no prediction* for 38.2%.

Pareto-front quality. Hypervolume, seed CV, and the cross-condition ε-indicator are reported in Table 2; the Stationary-vs-Mobile contrast is discussed in Section 5.2.

Pareto sweep and held-out likelihood. Across seven simplex points (enumerated in Section 5.2) the agent traces a behavioural Pareto front whose training-reward maximum is the food vertex (the only positive mean reward); food+clay collapses because food and clay zones are spatially incompatible. Held-out F_1 on Achaemenid–Hellenistic sites is compressed in absolute units ($[0.130, 0.136]$, $\sim$5$\times$ the random-over-land baseline of 0.024) but discriminating at the seed-noise scale ($\sigma_{\text{within}} \approx 0.0006$, giving $\sim$11 standard deviations of separability; full likelihood distribution in Figure 2 and the discussion).

5.2 LAMAP Control and Mobile

LAMAP control: yield modulation alone is null. The LAMAP control holds the Stationary action space and equal-weights reward fixed and modulates only harvest yields by the Iron Age Archaic LAMAP surface, so any behavioural change between Stationary and the LAMAP control must attribute to the environmental modulation. Across five seeds every metric is statistically indistinguishable from Stationary: mean reward -0.020 ± 0.028 vs -0.017 ± 0.029 (Mann–Whitney p_{bonf}=1.0, n=5 each), F_1 against held-out Achaemenid–Hellenistic sites 0.131 ± 0.001 vs 0.130 ± 0.002. The null result is the contribution: by ruling out a LAMAP–actions interaction, the LAMAP control lets every Stationary $\rightarrow$ Mobile behavioural change attribute cleanly to the spatial action-space extension alone. Without this control, the wood-niche reorientation of Mobile would remain confounded between two simultaneous environmental and action-space changes. We report no further LAMAP-control numbers.

Mobile: action space broadens the front. Mobile retains LAMAP modulation, extends the action space from four to seven actions (move, abandon, colonise), and runs nine weight vectors: the Stationary sweep minus wood+clay, plus three dense vectors at 0.8 (food, wood, or clay). Mobile occupies a region of viability space disjoint from Stationary's. The substantive finding is a sign-flip on wood: every Mobile weight vector reaches positive wood viability ($+0.18$ to $+0.30$), whereas no Stationary vector does (-0.27 to -0.20); per-condition wood-niche maps are in the supplementary material. The cost is food, which collapses into $[-0.77, -0.58]$ regardless of the food weight (Stationary spread: $[-0.67, +0.25]$). The additive epsilon-indicator on the six weight vectors common to both conditions (food-only, wood-only, clay-only, food+wood, food+clay, equal) confirms mutual non-dominance ($I_\varepsilon(\text{Stat}, \text{Mob})$=$+0.51$, $I_\varepsilon(\text{Mob}, \text{Stat})$=$+0.84$): neither front dominates the other, so Mobile is a genuine extension of the reachable behavioural manifold rather than an improvement along the same direction. The asymmetry $I_\varepsilon(\text{Mob}, \text{Stat}) > I_\varepsilon(\text{Stat}, \text{Mob})$ indicates Mobile displaces the front farther into wood-positive regions than Stationary displaces it back into food-positive ones.

The hypervolume CV rises from 2.7% under Stationary to 59% under Mobile: the extended action space exposes a larger and more diverse policy manifold but is sampled less reliably under the same training budget; we treat Mobile as evidence that spatial actions broaden the reachable Pareto region, with full-coverage characterisation of that region left to longer-budget or hierarchically-structured training (Section 6.4). Spatially, communities abandon Stationary's scattered ring-territories and concentrate on a wood-rich cluster in the western landscape: a many-to-one migration onto a single forested attractor that stationary patch selection cannot reach (per-condition and difference maps in the supplement).

Table 1: Stationary viability (mean $\pm$ std, five seeds, last 20 episodes) and spatial agreement with the Iron Age Archaic LAMAP surface (seed 43, last 20 episodes). $^{**}p < 0.01$ on the Spearman correlation (asymptotic).

Policy	Food	Wood	Clay	Mean reward
Rule-based	$+0.023$	-0.363	-0.011	-0.117 ± 0.001
PPO $\mathbf{w}$=(1/3, 1/3, 1/3)	$+0.212$	-0.251	-0.015	-0.020 ± 0.028
Soft-IoU vs. LAMAP	0.034	0.001	0.001	—
Per-community ρ vs. LAMAP	$+0.495^{**}$	-0.296	$+0.474^{**}$	$+0.522^{**}$

Table 2: Pareto-front quality across conditions. HV reference point $(-3, -3, -3)$, the per-resource viability clip floor; mean$\pm$std over five seeds. ε is computed on the six weight vectors common to both sweeps; $A \not\succeq B$ denotes $I_\varepsilon(A, B) > 0$ (i.e. A does not weakly dominate B).

Metric	Stationary	Mobile
HV ($\times 10^{-3}$)	8.5 ± 0.2	6.4 ± 3.8
σ/μ_{HV}	2.7%	59%
# vectors	7	9
$I_\varepsilon(\mathrm{Stat}, \mathrm{Mob})$=$+0.51$	(Stat $\not\succeq$ Mob)	
$I_\varepsilon(\mathrm{Mob}, \mathrm{Stat})$=$+0.84$	(Mob $\not\succeq$ Stat)	

Table 3: Mobile social-dilemma metrics (mean over five seeds, last 20 episodes). Efficiency: mean reward. Equality: $1-$Gini over per-community mean viability. Sustainability: mean per-resource viability across active communities. Peace omitted (per-step crowding not logged).

$\mathbf{w}$ (food, wood, clay)	Effic.	Equal.	Sust.
$(0.8, 0.1, 0.1)$	-0.60	0.90	-0.37
$(0.33, 0.33, 0.33)$	-0.13	0.91	-0.24
$(0.1, 0.8, 0.1)$	$+0.09$	0.90	-0.34

Spatial generalisation. Held-out F_1 against Achaemenid–Hellenistic sites falls from 0.131 under Stationary to 0.033 under Mobile (random-uniform-over-land baseline: 0.024). Soft-IoU against the held-out Achaemenid–Hellenistic LAMAP surface follows the same pattern ($0.071 \rightarrow 0.060$). The spatial actions diffuse harvest rather than concentrating it on the locations the held-out record validates; the sign-flip on wood is therefore a behavioural finding about what mobility unlocks under SAGAscape's resource geometry, not a generalisation gain.

CPR social outcomes. We report Pérolat's social-dilemma metrics [Perolat *et al.*, 2017] along the Mobile Pareto sweep (Table 3): efficiency (mean reward), equality ($1-$Gini over per-community mean viability), and sustainability (mean per-resource viability across active communities). At the equal-weights anchor, Stationary has equality 0.887; the LAMAP control matches; Mobile has equality 0.905. The food-viability collapse under Mobile (food $[-0.77, -0.58]$ regardless of weight) reflects the wood-niche concentration: communities that relocate onto the western forested cluster exhaust food everywhere else, and the 59% HV variance (Table 2) confirms that 200 episodes under-sample the seven-action policy manifold; Mobile's social outcomes are therefore best read as evidence that mobility unlocks the wood niche at a food cost, not as a converged characterisation. The equality increase is the egalitarian-collapse pattern of CPR dilemmas [Perolat *et al.*, 2017; Leibo *et al.*, 2017]: equality rises because most communities share the same low viability, not because the policy distributes gains. Sustainability tracks efficiency closely and is uniformly negative, consistent with Iron Age Archaic landscape under-supply once the Achaemenid–Hellenistic waves activate. The sweep makes the trade-off explicit: wood-dominant policies achieve positive efficiency at the cost of catastrophic food sustainability, while food-dominant policies are efficiency–sustainability dominated. No swept weight vector escapes the food/wood Pareto trade-off under the seven-action policy, indicating that fairness–efficiency reconciliation in this CPR requires either inter-community coordination signals (currently absent) or a non-linear scalarisation that penalises the egalitarian-collapse regime explicitly.

6 Discussion

6.1 Attribution via the LAMAP-Control Ablation

Our Stationary / LAMAP-control / Mobile design (Section 5.2) cleanly attributes the wood-niche reorientation to the spatial action space alone, ruling out a LAMAP–actions interaction as the cause.

6.2 Behavioural Headroom and the Wood Niche

The Stationary weight ranking shows food has the greatest behavioural headroom (food-only is the only positive mean reward); wood- and clay-dominant policies are uniformly worse, consistent with SAGAscape's structural scarcity of Iron Age wood and the geographically determined nature of clay. PPO at equal weights improves over the rule-based policy in the directions the landscape allows: food gains the most ($+0.023 \rightarrow +0.212$), wood partially recovers ($-0.363 \rightarrow -0.251$), clay is flat. Mobile's added mobility flips wood from $[-0.27, -0.20]$ to $[+0.18, +0.30]$ at the same weights. This explains why the rule-based SAGAscape parameter sweeps [Boogers and Daems, 2022], which hold settlement positions fixed, never escape the food/clay regime. The wood niche is a mobility-locked attractor that requires action-space enrichment, not preference change, to unlock.

6.3 Stationary Equifinality and Metric-Dependent Identifiability

The RL policy discovers resource-efficient landscape use, not archaeological site locations per se, yet aligns with

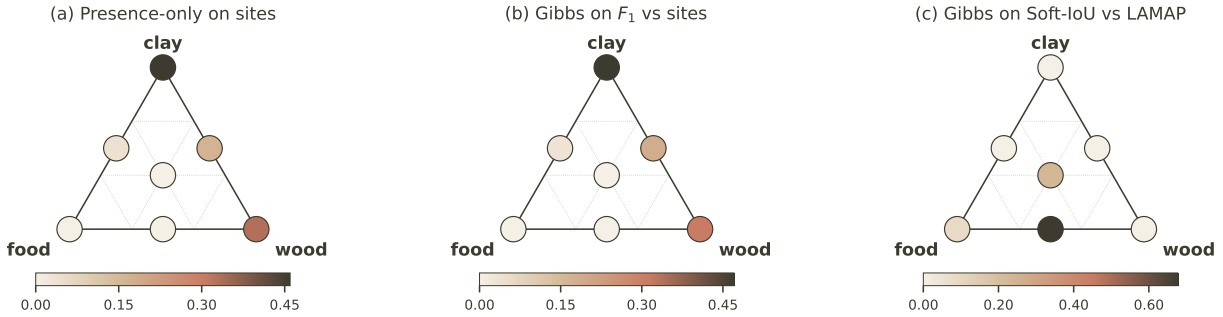


Figure 2: Likelihood-weighted distributions over the seven Stationary weight vectors; colour encodes probability mass (per-panel colorbar). (a) Presence-only on held-out sites: $\mathbb{E}[\mathbf{w}] \approx (0.02, 0.42, 0.56)$. (b) Temperature-scaled $\exp(\beta\, F_1)$ on the same data agrees with (a): $\mathbb{E}[\mathbf{w}] \approx (0.02, 0.40, 0.58)$. (c) Temperature-scaled $\exp(\beta\, \text{Soft-IoU})$ vs. held-out Achaemenid–Hellenistic LAMAP: $\mathbb{E}[\mathbf{w}] \approx (0.51, 0.42, 0.08)$. The choice of *data*, not likelihood form, drives the disagreement.

LAMAP suitability on average (per-community Spearman $\rho = +0.522$; the policy was never trained for this). The Stationary *LAMAP-missed* cells (0.3%, agent-active but LAMAP-silent) are candidates for non-resource attractors [Altaweel, 2015] that the correlative LAMAP predictor itself does not flag.

The Stationary F_1 surface is compressed across the simplex ($F_1 \in [0.130, 0.136]$, ∼5% relative range), so the held-out site record cannot uniquely pick a preference even though per-seed $\sigma_{\text{within}} \approx 0.0006$ makes the differences statistically distinguishable. Treating each fit score as a likelihood under a uniform prior, the presence-only point-process likelihood (Figure 2a, standard for sparse-site data [Phillips *et al.*, 2006; Elith *et al.*, 2011; Renner *et al.*, 2015]) concentrates mass on roughly three of seven vectors (ESS $\approx 2.85/7$, $\mathbb{E}[\mathbf{w}] \approx (0.02, 0.42, 0.56)$): the equifinality content of the experiment. The temperature-scaled $p(\mathbf{w}) \propto \exp(\beta\, F_1)$ [Ziebart *et al.*, 2008] (Figure 2b) agrees ($\mathbb{E}[\mathbf{w}] \approx (0.02, 0.40, 0.58)$), confirming wood/clay-dominant concentration as a property of the data, not the likelihood form. Evaluating against the held-out LAMAP surface via Soft-IoU (Figure 2c) instead shifts the posterior to food/wood-balanced policies ($\mathbb{E}[\mathbf{w}] \approx (0.51, 0.42, 0.08)$): the choice of *data*, not likelihood form, drives the disagreement. This is metric-dependent identifiability.

6.4 Limitations

The agent learns which resource queue receives workday effort, not which patch within a queue is selected; replacing the inherited greedy patch-selection with a learned spatial policy is a natural next step. Under Mobile this same food-based rule also picks the move destination, and that inherited simplification proves more limiting than expected: communities learn *whether* to relocate but not *where*, so $\mathbf{w}$ cannot steer them. Mobile's wood recovery is then a generic side effect of relocating onto fresh forest rather than a preference-driven choice, identical across weights. We expect that scoring destinations by all resource weights, not food alone, would restore spatial expressivity; the reported seed variance is a secondary under-sampling effect, not the cause, and held-out-period generalisation under the extended action space remains open. The Mobile action set (move, abandon, colonise) is a reasonable parameterisation rather than a theory-derived one: the move radius (R=20 px), the colonisation surplus threshold ($1.5\times$ food requirement), and the cooldown (50 ticks) are engineering choices that future work should ground in archaeological theory or post-hoc fit to data. Inactive communities contribute zero reward, so abandon replaces a negative with zero; this soft incentive may inflate Mobile's per-active metrics. The action space is also asymmetric across resources: the aggressiveness factor can drive the food target into surplus, but wood and clay are harvested only to a fixed requirement. The recovered preference is therefore informative on the food-versus-rest axis but cannot separate wood from clay; the wood/clay-dominant posterior should be read as a lump, not a wood-over-clay or clay-over-wood preference. We sweep the simplex as discrete policies (each weight vector reads as a distinct archaeological hypothesis); the conditioned-policy alternative is left to future work. The remaining limitations are inherited from SAGAscape [Boogers and Daems, 2022]: no interannual climate variability beyond stochastic bad harvests, fixed active-percentage population, no inter-community trade. The learned policy is stationary across the 801-year span; a recurrent or period-conditioned architecture could express qualitatively different strategies across sub-periods. Generalisation beyond this setting is untested: results come from a single site (Sagalassos) and a single forward model, and other transitions (e.g., Roman Imperial) require retraining; external validity to other regions and simulation models is the natural next benchmark axis.

7 Conclusion

We re-posed archaeological equifinality as multi-objective inverse preference inference and delivered its forward half. SAGAscape-RL trains a parameter-shared multi-agent policy across a discrete resource preference simplex on the Iron Age Archaic to Achaemenid–Hellenistic transition at Sagalassos (801 years), and returns a Pareto front of subsistence strategies whose held-out fit is the forward likelihood surface any inverse procedure must consume.

Three empirical findings emerged. Under stationary settle-

ment, the held-out site record fits several distinct preferences nearly equally well (ESS $\approx 2.85/7$): the equifinality content of the experiment. Point-process and temperature-scaled likelihoods agree on a wood/clay-dominant posterior, while a Soft-IoU evaluation against LAMAP prefers food/wood-balanced weights: metric-dependent identifiability, with the choice of *data*, not likelihood form, driving the disagreement. Where harvest density and LAMAP diverge, agent-active but LAMAP-silent cells are candidates the correlative LAMAP predictor itself does not flag. Allowing communities to move, abandon, and colonise traces a second front mutually non-dominated with Stationary on the shared weight vectors, unlocking a wooded western attractor at a food-security cost; the wood niche is mobility-locked, not preference-locked.

Archaeological data manifest through spatial outcomes, not the trajectories on which inverse RL typically relies. By scoring the Pareto front against the archaeological evidence, we obtain a likelihood surface over preferences. Its inversion seeks to recover the latent preference itself, a natural next step we pursue.

Future work. The natural inverse-pass tool is maximum-entropy IRL [Ziebart *et al.*, 2008]: its softmax-rational behaviour model relaxes strict optimality, an assumption ill-suited to archaeological agents, and recovers a posterior over $\mathbf{w}$ whose breadth quantifies equifinality directly. A preference-conditioned MORL formulation would replace the discrete sweep with a single conditioned policy and enable inverse preference inference proper. Because the move destination ignores the preference (Section 6.4), spatial actions as currently formulated cannot sharpen the likelihood; a weight-conditioned spatial action that scores destinations by the preference vector is the step required, and is ongoing work. An activity map combining harvest density with movement trajectories (entry, exit, settlement duration) would distinguish policies with identical footprints but different histories, useful both archaeologically and as a target for the inverse pass. A distance-decay centralised critic over Tobler walking-time radii would reduce value-estimate variance, and inter-community trade would complete the SAGAscape roadmap and unlock specialist strategies currently unreachable on the swept Pareto front.

Full derivations, hyperparameters, and additional results are in the supplementary material.

Acknowledgements

S.J. gratefully acknowledges support from Fonds Wetenschappelijk Onderzoek (FWO) via FWO PhD Fellowship strategic basic research, Belgium 1SHHV24N. P.J.K.L. wishes to express gratitude for the support received from the research council of the Vrije Universiteit Brussel (OZR-VUB via grant number OZR3863BOF). This work was further supported by the Vrije Universiteit Brussel Interdisciplinary Research Program IRP34, ARCHAI: Archaeological Prospection through an AI-driven approach. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation Flanders (FWO) and the Flemish Government department EWI. Lastly, we want to thank the HPC administration and support service of Vrije Universiteit Brussel.

References

[Abbeel and Ng, 2004] Pieter Abbeel and Andrew Y. Ng. Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the Twenty-First International Conference on Machine Learning*, ICML '04, page 1, New York, NY, USA, 2004. Association for Computing Machinery.

[Altaweel, 2015] Mark Altaweel. Settlement dynamics and hierarchy from agent decision-making: A method derived from entropy maximization. *Journal of Archaeological Method and Theory*, 22(4):1122–1150, 2015.

[Axtell *et al.*, 2002] Robert L Axtell, Joshua M Epstein, Jeffrey S Dean, George J Gumerman, Alan C Swedlund, Jason Harburger, Shubha Chakravarty, Ross Hammond, Jon Parker, and Miles Parker. Population growth and collapse in a multiagent model of the kayenta anasazi in long house valley. *Proceedings of the National Academy of Sciences*, 99(suppl_3):7275–7279, 2002.

[Boogers and Daems, 2022] Stef Boogers and Dries Daems. Sagascape: Simulating resource exploitation strategies in iron age to hellenistic communities in southwest anatolia. *Journal of Computer Applications in Archaeology*, 5(1), 2022.

[Bradbury *et al.*, 2018] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs, 2018.

[Carleton *et al.*, 2012] W Chris Carleton, James Conolly, and Gyles Ianonne. A locally-adaptive model of archaeological potential (lamap). *Journal of Archaeological Science*, 39(11):3371–3385, 2012.

[Elith *et al.*, 2011] Jane Elith, Steven J. Phillips, Trevor Hastie, Miroslav Dudík, Yung En Chee, and Colin J. Yates. A statistical explanation of maxent for ecologists. *Diversity and Distributions*, 17(1):43–57, 2011.

[Jaxy *et al.*, 2025] Simon Jaxy, Anton Theys, Patrick Willett, W. Chris Carleton, Ralf Vandam, and Pieter Libin. Needles in the landscape: Semi-supervised pseudolabeling for archaeological site discovery under label scarcity, 2025.

[Kishikawa and Arai, 2021] Daiko Kishikawa and Sachiyo Arai. Estimation of personal driving style via deep inverse reinforcement learning. *Artificial Life and Robotics*, 26(3):338–346, 2021.

[Kohler and Gumerman, 2000] Timothy A Kohler and George G Gumerman. *Dynamics in human and primate societies: Agent-based modeling of social and spatial processes*. Oxford University Press, 2000.

[Leibo *et al.*, 2017] Joel Z Leibo, Vinicius Zambaldi, Marc Lanctot, Janusz Marecki, and Thore Graepel. Multi-agent reinforcement learning in sequential social dilemmas. *arXiv preprint arXiv:1702.03037*, 2017.

[Ng and Russell, 2000] Andrew Y. Ng and Stuart J. Russell. Algorithms for inverse reinforcement learning. In *Proceedings of the Seventeenth International Conference on Machine Learning*, ICML '00, page 663–670, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.

[Perolat *et al.*, 2017] Julien Perolat, Joel Z Leibo, Vinicius Zambaldi, Charles Beattie, Karl Tuyls, and Thore Graepel. A multi-agent reinforcement learning model of common-pool resource appropriation. *Advances in neural information processing systems*, 30, 2017.

[Phillips *et al.*, 2006] Steven J. Phillips, Robert P. Anderson, and Robert E. Schapire. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3):231–259, 2006.

[Poblome, 2023] Jeroen Poblome. *Introducing the Sagalassos Archaeological Research Project*, pages 7–30. Leuven University Press, 2023.

[Renner *et al.*, 2015] Ian Renner, Jane Elith, Adrian Baddeley, William Fithian, Trevor Hastie, Steven Phillips, Gordana Popovic, and David Warton. Point process models for presence-only analysis. *Methods in Ecology and Evolution*, 6, 02 2015.

[Reymond *et al.*, 2024] Mathieu Reymond, Conor F. Hayes, Lander Willem, Roxana Rădulescu, Steven Abrams, Diederik M. Roijers, Enda Howley, Patrick Mannion, Niel Hens, Ann Nowé, and Pieter Libin. Exploring the pareto front of multi-objective covid-19 mitigation policies using reinforcement learning. *Expert Systems with Applications*, 249:123686, 2024.

[Roijers *et al.*, 2013] Diederik M Roijers, Peter Vamplew, Shimon Whiteson, and Richard Dazeley. A survey of multi-objective sequential decision-making. *Journal of Artificial Intelligence Research*, 48:67–113, 2013.

[Rondeau *et al.*, 2022] Rob Rondeau, W Christopher Carleton, Mark Collard, and Jonathan Driver. Does the locally-adaptive model of archaeological potential (lamap) work for hunter-gatherer sites? a test using data from the tanana valley, alaska. *Plos one*, 17(3):e0265597, 2022.

[Schulman *et al.*, 2017] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[Tobler, 1993] Waldo Tobler. Three presentations on geographical analysis and modeling. Technical Report 93-1, National Center for Geographic Information and Analysis, 1993.

[Van Moffaert and Nowé, 2014] Kristof Van Moffaert and Ann Nowé. Multi-objective reinforcement learning using sets of pareto dominating policies. In *Journal of Machine Learning Research*, volume 15, pages 3483–3512, 2014.

[Verhagen and Whitley, 2012] Philip Verhagen and Thomas G Whitley. Integrating archaeological theory and predictive modeling: a live report from the scene. *Journal of Archaeological Method and Theory*, 19(1):49–100, 2012.

[Vogels *et al.*, 1975] Maurice Vogels, Rita Zoeckler, Donald M. Stasiw, and Lawrence C. Cerny. P. f. verhulst's "notice sur la loi que la populations suit dans son accroissement"from correspondence mathematique et physique. ghent, vol. x, 1838. *Journal of Biological Physics*, 3(4):183–192, 1975.

[Willett, 2022] Patrick Willett. Transforming landscapes of southwest anatolia: Modeling social and environmental change from the middle to late holocene using predictive land-use and cropland reconstructions, 2022.

[Yu *et al.*, 2022] Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in neural information processing systems*, 35:24611–24624, 2022.

[Ziebart *et al.*, 2008] Brian D. Ziebart, Andrew Maas, J. Andrew Bagnell, and Anind K. Dey. Maximum entropy inverse reinforcement learning. In *Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 3*, AAAI'08, page 1433–1438. AAAI Press, 2008.

Supplementary Material
Multi-Objective Inverse Preference Inference for Archaeological Equifinality: A Forward Pareto Sweep on a Settlement Agent-Based Model

Simon Jaxy[1] , **Luca Vincke**[1] , **Bram Silue**[1] , **Dries Daems**[3] , **Ralf Vandam**[2] and **Pieter Libin**[1]
[1]AI Lab, Department of Computer Science, Vrije Universiteit Brussel
[2]MGC (Archaeology, Environmental Changes & Geo-Chemistry), Vrije Universiteit Brussel
[3]Spatial Digital Humanities, Vrije Universiteit Amsterdam
simon.jaxy@vub.be

S1 Step Equations

Food fertility on cleared patches recovers toward the soil-quality cap K_p^{food} via Verhulst logistic growth:

$$\delta_p = \frac{1 - f_p/K_p^{\text{food}}}{1/K_p^{\text{food}} + g_p/[f_p(1-g_p)]}, \quad f_p \leftarrow \min(f_p + \delta_p,\, K_p^{\text{food}}), \tag{1}$$

with per-patch growth rate g_p, calibrated from a reserve fertility f_{res} following SAGAscape [Boogers and Daems, 2022]; fully depleted patches restart at the flat reserve $f_p = f_{\text{res}}$. SAGAscape sets $f_{\text{res}} = 0.1$; the experiments reported here used $f_{\text{res}} = 0.5$, which makes agricultural recovery more forgiving than the original. As f_{res} is held fixed across every condition, it shifts fertility trajectories uniformly and does not confound the cross-condition comparisons; the tick-1 validation of Table 1 precedes any regeneration step and is unaffected. Subsequent work adopts $f_{\text{res}} = 0.1$. Forest regrowth follows:

$$w_p \leftarrow \text{clip}(A_p \exp(B_p \cdot \text{age}_p),\, w_p,\, K_p^{\text{wood}}), \quad \text{age}_p \mathrel{+}= 1, \tag{2}$$

with per-patch shape parameters A_p, B_p derived from local conditions in SAGAscape [Boogers and Daems, 2022] and forest-cover cap K_p^{wood}, subject to a six-year recolonisation lag after field abandonment.

S2 Initial Community Sizes

Communities are initialised in three size classes inherited from SAGAscape [Boogers and Daems, 2022], drawing populations from per-class Gaussian distributions:

Class	Population draw	Description
Hamlet	$\mathcal{N}(50,\, 10)$	small farmstead
Village	$\mathcal{N}(500,\, 100)$	mid-tier nucleated settlement
Town	$\mathcal{N}(1000,\, 200)$	major regional centre

Per-class assignment is read from the Sagalassos survey typology of Willett [Willett, 2022] and held fixed across replicates; only the intra-class population draw is stochastic. Per-capita food, wood, and clay requirements scale linearly with population, inherited verbatim from SAGAscape [Boogers and Daems, 2022].

S3 NetLogo Validation

We validate the JAX environment against the original NetLogo model [Boogers and Daems, 2022] by comparing aggregate stocks after one simulated year under deterministic mode: security factor fixed at 1.5 (midpoint of NetLogo's uniform range), bad-harvest modifier fixed at 1.0, and community populations seeded to match NetLogo's total population of 12,492 (inhabitants summed across the community agents).

Table 1: Tick-1 aggregate stock comparison (deterministic mode).

Resource	JAX	NetLogo	Error
Population	12,492	12,492	0.0%
Clay (t)	127.54	127.54	0.0%
Wood (m^3)	−9,625	−9,599	−0.3%
Food (t)	2,865	3,435	16.6%

Population, clay, and wood stocks align within measurement precision (wood is negative because tick-1 demand exceeds standing stock; the sign matches NetLogo by construction). The 16.6% food discrepancy arises because, while global population totals match, the two implementations draw independent community-level samples from the same distributions, producing different per-community workday budgets and patch-selection sequences. This divergence is expected when comparing independently seeded stochastic ABMs.

S4 Soft-IoU Threshold Sensitivity

The four-quadrant analysis in the main text (Figure 1 of the main paper) binarises both the harvest density and the Iron Age Archaic LAMAP surface at 0.4 (normalised units). Table 2 reports the binarised IoU at thresholds 0.3, 0.4, 0.5 across all seven Stationary weight vectors. The ranking is preserved across thresholds, but absolute IoU contracts roughly by half between 0.3 and 0.5 as the binarised mask shrinks. The qualitative finding (food/wood edge and equal-weights interior dominate) is robust to threshold choice; the absolute IoU values should be read as threshold-conditional, not as universal alignment scores.

Table 2: Binarised IoU between Stationary harvest density and the Iron Age Archaic LAMAP surface, evaluated at three binarisation thresholds. Mean over five seeds, last 10 episodes.

Weight vector	τ=0.3	τ=0.4	τ=0.5
food_only	0.029	0.021	0.005
wood_only	0.016	0.011	0.003
clay_only	0.015	0.010	0.003
food_wood	0.033	0.023	0.006
food_clay	0.020	0.014	0.003
wood_clay	0.017	0.012	0.003
equal	0.031	0.022	0.005

S5 Stationary Simplex Panels (Raw Signals)

Figure 1 shows the Stationary sweep on the simplex. The food vertex dominates mean reward; held-out F_1 is compressed in absolute units ($[0.130, 0.136]$) but statistically discriminating across $\sim 11\sigma$ at the seed-noise scale (see Figure 2 of the main paper for the resulting posterior). Soft-IoU shows visible differentiation in absolute units too.

S6 Wood-Niche Spatial Maps

Figure 2 shows wood harvest density maps at the equal-weights anchor under Stationary and Mobile, with the many-to-one migration map between them.

S7 Observation Vector

Notation: $\hat{f}_n, \hat{w}_n, \hat{c}_n$ are the food, wood, and clay requirements of community n; $\mathcal{T}_n$ is its territory of size up to $T_{\max}$; $s_p = r_p/c_p$ is the resource-to-cost score at patch p; d_n^{food} is the consecutive-tick food deficit count; $T_{\text{ep}} = 801$ is the episode length in ticks; ℓ_n is the home patch.

Table 3: 12-D observation vector. All dimensions are present in every condition; under Stationary, dim 11 is identically 1 (no signal).

Dim	Feature	Definition
0–2	Stock surplus (f/w/c)	$F_n/\hat{f}_n$, $W_n/\hat{w}_n$, $C_n/\hat{c}_n$
3–5	Territory richness (f/w/c)	$\overline{r_p/c_p}$ over $\mathcal{T}_n$, normalised
6	Territory fullness	$\lvert\mathcal{T}_n\rvert/T_{\max}$
7	Active fraction	$\sum_n \mathbf{1}[\text{active}_n]/N$
8	Deficit streak	$d_n^{\text{food}}/T_{\text{ep}}$
9	Population fraction	$\text{pop}_n/\max_{n'} \text{pop}_{n'}$
10	Best patch score	$\max_{p\in\mathcal{T}_n} s_p$
11	LAMAP at home	λ_{ℓ_n} ($\equiv 1$ under Stationary)

S8 PPO Hyperparameters

Table 4: PPO training hyperparameters.

$\gamma = 0.999$	GAE $\lambda = 0.99$
Clip $\epsilon = 0.2$	Entropy coeff. 0.01
Learning rate 1×10^{-4}	Adam ($\epsilon_{\text{Adam}} = 10^{-5}$)
Max grad. norm 0.5	Hidden layers 2×64 (tanh)
Minibatches/update 4	Epochs/episode 4
Episode length 801 ticks	Training episodes 200
Replicates 50 (baselines)	Seeds 5 (42–46)

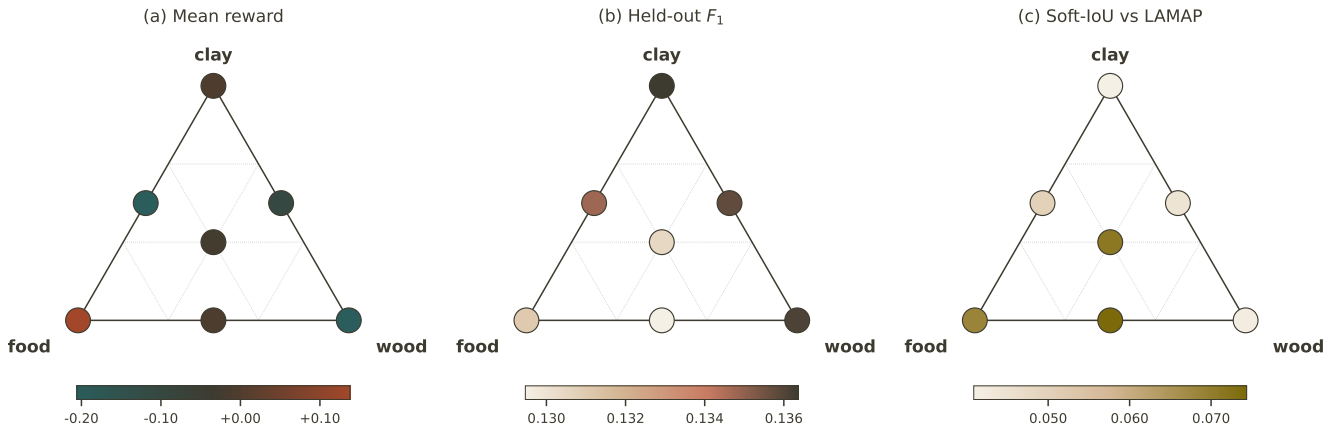
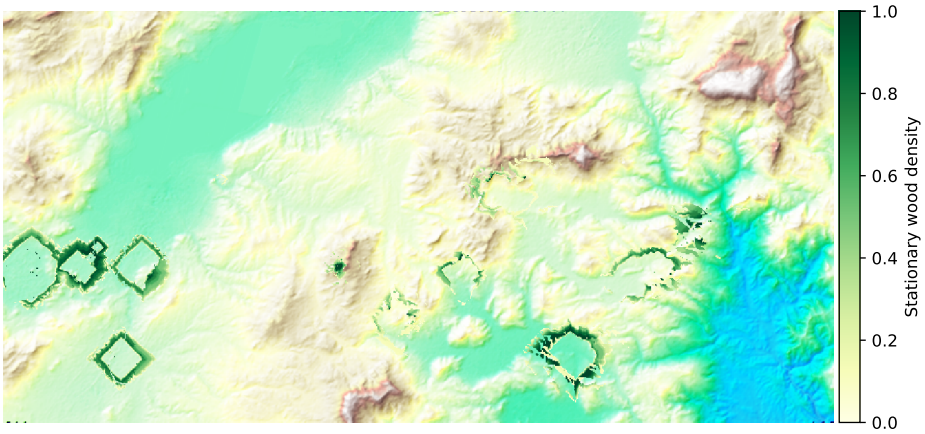
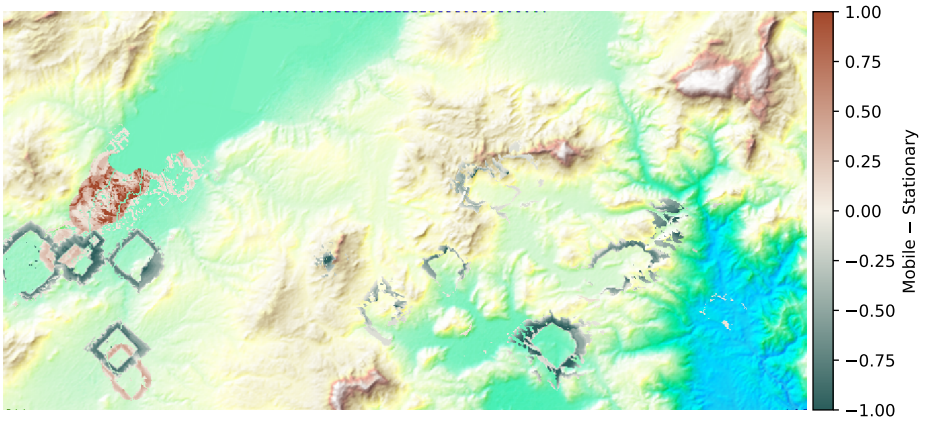


Figure 1: Stationary weight sweep over the simplex: (a) mean training reward, (b) held-out F_1 against Achaemenid–Hellenistic sites, (c) Soft-IoU against the held-out Achaemenid–Hellenistic LAMAP surface. Each dot is a trained policy at its weight vector.



(a) Stationary

(b) Mobile



(c) Mobile − Stationary

Figure 2: Per-condition wood harvest density maps at the equal-weights anchor, averaged over five seeds and the last 20 episodes; values clipped at the shared 98th percentile. Stationary's harvest forms ring-shaped territory shells around fixed community centres; Mobile collapses most communities onto a single wood-rich cluster in the western landscape. Panel (c) is the many-to-one migration map: teal cells lose harvest activity under Mobile (Stationary territories abandoned), rust cells gain it (the wood-rich plain colonised under mobility).

References

[Boogers and Daems, 2022] Stef Boogers and Dries Daems. Sagascape: Simulating resource exploitation strategies in iron age to hellenistic communities in southwest anatolia. *Journal of Computer Applications in Archaeology*, 5(1), 2022.

[Willett, 2022] Patrick Willett. Transforming landscapes of southwest anatolia: Modeling social and environmental change from the middle to late holocene using predictive land-use and cropland reconstructions, 2022.