WUIRL: RECONSTRUCTING WILDFIRE STRUCTURE IGNITION IN THE WILDLAND–URBAN INTERFACE WITH REINFORCEMENT LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Wildfires are increasingly burning through the wildland–urban interface (WUI), destroying homes and other structures where development meets wildland vegetation. Modeling which buildings are most at risk remains difficult because the physical mechanisms governing fire spread and ignition in the WUI are incompletely understood, data on WUI and wildland fuel are coarse, and historical fire progression is too sparsely observed to learn directly. Supervised methods currently map building, vegetation, and weather data to structure outcomes. However, these static inputs cannot capture the mechanisms by which fire reaches a structure over time, which is necessary to predict ignition and determine an appropriate method to mitigate risk. Capturing the ignition process requires temporal ignition data, which is rarely recorded, leaving the process that determines survival and guides mitigation largely undocumented. Therefore, we introduce Wildland–Urban Interface Reinforcement Learning (WUIRL), a framework that reconstructs how fire reaches and ignites buildings from limited observations by pairing proximal policy optimization (PPO) with a physical ignition model. The learned policy selects pathways through which burning vegetation, neighboring structures, and external sources expose buildings to heat, and the physical model determines how each building responds. WUIRL thus links structural vulnerability and evolving exposure, reconstructing progression from recorded damage and coarse timing without complete exposure histories. We apply WUIRL to the 2025 Eaton fire, recovering 78.9% of buildings labeled as ignited from damage records while achieving the highest balanced accuracy among the evaluated policies. The reconstruction is most sensitive to the modeled radiant heat and firebrand-proxy entry mechanisms, whose removal most degrades performance. WUIRL provides a foundation for future studies of structure loss under dynamic exposure, offering the potential to better test specific interventions' role in limiting future damage as higher resolution observations become available.

1 INTRODUCTION

The number of structures destroyed by wildfires in the United States approximately tripled between 2002 and 2022 (Carlson et al., 2025). Two compounding trends have exacerbated this rise. The number of homes in the wildland–urban interface (WUI), where urban development meets wildland vegetation, grew by 46% over the past three decades, and recent wildfires are burning larger areas with greater speed and intensity (Balch et al., 2024; Radeloff et al., 2023). As more buildings fall within reach of larger fires, understanding not only which buildings are vulnerable but how fire reaches and ignites them becomes central to reducing future losses (Cohen, 2010).

Supervised learning has been widely applied to predict structure ignition, relating building construction and surroundings to recorded damage, from tabular models to graph neural networks (Ejaz & Choudhury, 2025). However, structure ignition is governed by the latent process of fire propagation, which construction and environmental features alone do not capture. Final damage records observe only the endpoint of this process, not the exposure history that produced it, so a building that withstood heavy exposure and one that received almost none are indistinguishable in the available data. Recovering how the fire unfolded would require building-level exposure and ignition histories,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

but progression data for destructive wildfires are seldom recorded, a data limitation that rules out training a deep model of the end-to-end dynamics (Ejaz & Choudhury, 2025).

While fires in the WUI do ignite and spread via transported embers and direct flame contact between small combustibles, destructive WUI fires are dominated by a conflagration of structure-to-structure fire spread (Caton et al., 2017; Filkov et al., 2023). A burning building thermally exposes its neighbors, and those that ignite become sources in turn. Therefore, whether a structure ignites depends on what has already burned around it, not on its construction type and condition alone (Naser & Kodur, 2025). This calls for a method that couples structural vulnerability with the ignition process as it evolves, not from a single static state of the community layout. We approach this as a Markov decision process, adapting a sequential framing that prior work (Ganapathi Subramanian & Crowley, 2018; Soderlund & Kumar, 2018) applied to landscape fire spread.

We establish Wildland–Urban Interface Reinforcement Learning (WUIRL), a framework that reconstructs how fire approaches and ignites buildings from limited observations by pairing a learned decision policy with a physical ignition model. A graph-based encoder represents each building's attributes and neighborhood, and a policy trained with proximal policy optimization (PPO) (Schulman et al., 2017) selects the pathways through which burning vegetation, neighboring structures, and external sources expose a building to heat. A physical ignition model then determines whether each building ignites. The policy learns only which pathways to select rather than the full ignition process, letting WUIRL reconstruct progression from recorded damage and coarse timing without complete exposure histories.

We apply WUIRL to reconstruct the 2025 Eaton fire, which destroyed thousands of structures in a dense WUI community and is unusually well documented, with building-level damage inspections (California Department of Forestry and Fire Protection, 2025), pre-fire imagery for mapping vegetation (Coleman et al., 2020; European Union/ESA/Copernicus, 2015–2026; USDA Farm Production and Conservation, Business Center, Geospatial Enterprise Operations, 2022), and an estimated hourly perimeter timeline (Kerber et al., 2025). From this single fire's sparse, coarse observations, WUIRL recovers 78.9% of buildings recorded as ignited and attains the highest mean reconstruction reward among the evaluated policies. The model's reconstructions depend primarily on radiant heat between structures and firebrand-proxy entry. Most missed ignitions occur at buildings the simulated fire never reaches. We recover burn outcomes more reliably than ignition timing, and observe deterioration with continued training, highlighting the challenge of learning fire progression from limited data. We present WUIRL as a case study to demonstrate this technique rather than a validated predictor. It shows how damage records and coarse progression can be integrated with model-consistent ignition histories, making the possible fire spread pathways through a community explicit and testable. With further validation, WUIRL could help investigate how mitigation alters exposure and building loss. Our contributions are threefold:

- **A sequential formulation of WUI fire reconstruction.** We recast structure loss as an evolving community process, connecting each building's changing exposure with an ignition response set by its structural attributes, so reconstruction can proceed from recorded damage and coarse progression intervals rather than complete exposure histories.
- **WUIRL, a physically constrained reinforcement-learning framework.** We combine a pretrained, frozen GraphSAGE encoder with a masked, factorized PPO policy and fixed physical ignition dynamics, so the policy learns only which exposure pathways to select while the physical model determines their consequences.
- **An Eaton fire case study with mechanism, timing, and stability diagnostics.** We evaluate against Random and Greedy baselines and probe dependence on modeled exposure mechanisms, separating final-loss recovery from progression accuracy and identifying candidate-pathway availability and training stability as central limitations.

2 Related Work

Wildfire modeling and structure loss in the WUI. Physics-based fire spread models represent wildland and WUI fire progression through front-propagation rules or level-set approaches (Finney, 1998; Lautenberger, 2013; Masoudvaziri et al., 2021; Purnomo et al., 2024), spread-probability graphs (Chulahwat et al., 2022), or cellular automata formulations (Jiang et al., 2021; Mastorakos

et al., 2023). A complementary line of work lacks fire propagation but uses supervised learning to predict structure damage from building attributes, surrounding vegetation, weather, and spatial contexts (Bar et al., 2026; Jia & Opabola, 2025; Mockrin et al., 2023; Noumeur & Tohir, 2025; Zamanialaei et al., 2025). GraphFire-X (Esparza et al., 2026), for example, combines a physics-informed neighborhood graph with a structural fragility model to predict building-scale damage. These approaches simulate fire progression using fixed relationships or predict eventual structure outcomes, but do not reconstruct the sequence of structure-level exposure pathways that produced the observed damage. WUIRL instead couples evolving exposure with structure-specific vulnerability to reconstruct candidate building-level fire progressions.

Reinforcement learning for wildfire dynamics. Wildfire progression has previously been formulated as a sequential decision process by treating the spreading fire itself as an agent (Soderlund & Kumar, 2018). Ganapathi Subramanian & Crowley (2018) used reinforcement learning to model landscape fire spread, with a learned policy that obtains reward for agreement with observed burn outcomes. Other wildfire reinforcement-learning work uses the agent as a controller for firebreak placement (Murray et al., 2025), forest management (Altamimi et al., 2022), aerial suppression (Haksar & Schwager, 2018; Mathur et al., 2025), or adverse-scenario generation near the WUI (González-Villa et al., 2025). WUIRL differs in that ignition dynamics are fixed rather than learned, and the policy selects unobserved exposure pathways between structures to reconstruct progressions consistent with recorded outcomes. To our knowledge, this is the first application of wildfire reinforcement learning to retrospective building-level reconstruction.

Graph-based policy parameterization. Graph neural networks have been combined with reinforcement learning for decision making in graph-structured environments (Nie et al., 2023), including actor–critic policies that select graph nodes as actions (Antonietti et al., 2025). WUIRL uses GraphSAGE (Hamilton et al., 2017), invalid-action masking (Huang & Ontañón, 2022), and proximal policy optimization (PPO) (Schulman et al., 2017). Our contribution is the use of these methodologies within a physically constrained sequential formulation of structure-level WUI fire reconstruction.

3 Sequential Reconstruction of Community Fire Progression

A structure's outcome when exposed to a wildfire is determined by two properties the reconstruction must represent, vulnerability and exposure. A structure's *vulnerability* is how readily it ignites given its structural attributes, and *exposure* is the heat and embers that reach the structure as fire moves through the community (Young et al., 2025). We represent vulnerability using fixed structural attributes, with exposure depending on the latent order of the fire spread (Naser & Kodur, 2025). Due to data limitations, the recorded data captures final damage and coarse progression, but not the complete order of spread (Ejaz & Choudhury, 2025; Young et al., 2025). The same set of burned buildings could arise from many different spread sequences, so the order in which the fire moved through the community, and the mechanism by which each building ignited, go undocumented.

WUI structure ignition cascades, as a building that ignites becomes a source for its neighbors. Therefore, exposure at one structure depends on what has already burned, and reconstruction must follow the fire as a sequence. We construct the simulator state to retain the fire status, accumulated dose, source and vegetation histories, selection plans, external-entry clocks, and elapsed simulation time. Conditional on fixed environmental inputs, we define a modeled Markov decision process $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$, specified component by component below, extending to structural ignition the sequential-decision framing prior work applied to landscape wildfire spread (Ganapathi Subramanian & Crowley, 2018; González-Villa et al., 2025; Soderlund & Kumar, 2018). Each trajectory generated through this process is a candidate reconstruction of the latent spread order, scored by the reward against the recorded damage, so reconstruction comprises searching the constrained space of admissible trajectories for those consistent with what was observed. Our method, WUIRL, couples a learned policy that selects exposure pathways with the physical model that resolves them. The transition $\mathcal{P}$ combines literature-based physical relationships with modeling assumptions rather than learned relationships, so the policy learns only pathway selection, and agreement with outcomes alone cannot validate the physical relationships individually (Appendix B, Table 2).

3.1 COMMUNITY AND EXPOSURE PATHWAYS

We represent the community using a building-level graph, as demonstrated in prior WUI modeling approaches (Chulahwat et al., 2022; Esparza et al., 2026). Our directed graph $\mathcal{G} = (\mathbb{B}, \mathbb{L})$ contains buildings $\mathbb{B}$ connected by candidate exposure edges $\mathbb{L}$ within 30 m footprint-to-footprint distances. The graph supports neighborhood encoding and local exposure pathways, and external-entry routes provide additional action opportunities.

A local exposure pathway $p = (u, i, m)$ identifies a source structure u, receiving building i, and ignition mechanism m. Radiation, direct flame contact, and firebrands are key established ignition pathways for structures (Caton et al., 2017).

We model local thermal exposure through whole-building radiation, flames emerging from building openings, and radiation from nearby burning trees and shrubs. Vegetation patches activate on modeled flame contact and deliver a constant, distance-dependent radiant heat flux for 30 seconds. For direct building exposure, u identifies the burning source, whereas for vegetation-mediated exposure it identifies the building that activated the vegetation.

Two rule-based routes represent external entry. Wildland-interface entry represents ignition where an advancing wildland fire interacts with buildings at the community boundary. Eligibility depends on the estimated hourly progression, nearby vegetation, wind, distance, and building vulnerability, but not on observed damage. Firebrand-proxy entry parameterizes ignition by firebrands through an eligibility rule rather than an explicit transport model. A building becomes eligible after sufficient selected thermal exposure when an active source lies upwind within a fixed range. Both routes can trigger ignition without adding modeled heat flux or component dose (Appendix B.6).

3.2 BUILDING RESPONSE AND SEQUENTIAL PROGRESSION

The thermal branch of $\mathcal{P}$ maps selected local pathways to accumulated dose and ignition. The selected pathways to a receiver i, written $\mathbb{P}_i(t)$, contribute a combined incident heat flux $q''_i(t) = \sum_{p \in \mathbb{P}_i(t)} q''_p(t)$, where $q''_p(t) \geq 0$ is the flux from pathway p.

Thermal exposure alone does not necessarily result in ignition (Cohen & Butler, 1998). The model compares how strongly and how long each accessible component is heated against requirements set by construction and protective features, the building's vulnerability, using a flux–time-product formulation (Silcock & Shields, 1995). For building i and component c, $q''_{\mathrm{crit},ic}$ is the critical flux below which no progress accrues, $n_{ic} > 0$ shapes the response above it, and $\mathrm{FTP}_{ic} > 0$ is the flux–time requirement. Construction determines the assigned material parameters, including the adopted wood ignition correlation (Cohen & Butler, 1998). Material-specific values are reported in Appendix B.5. Some building components additionally require a pilot, which is a small flame or ember that initiates ignition. A pilot may be supplied by a positive contribution from building-opening flames or burning trees and shrubs, but not by whole-building radiation alone.

Normalized dose Φ_{ic} records progress toward ignition. Holding the combined flux at its interval value $\bar{q}''_i$ over a step δ_i, for an eligible component whose pilot requirement, if any, is satisfied, the dose evolves as

$$\Phi_{ic}(t + \delta_i) = \min\left\{1,\ \Phi_{ic}(t) + \frac{[\max(\bar{q}''_i - q''_{\mathrm{crit},ic}, 0)]^{n_{ic}} \delta_i}{\mathrm{FTP}_{ic}}\right\} \tag{1}$$

and thermal ignition occurs when any eligible component reaches $\Phi_{ic} = 1$. External-entry rules provide a separate ignition route. The model omits cooling effects, so accumulated dose never decays.

An ignited building emits a time-dependent heat release rate for a set duration and then extinguishes, exposing neighbors and activating nearby vegetation while it burns (Appendix B.3). Newly ignited buildings enter the next time window's exposure calculation, so earlier ignitions introduce later ignition opportunities while the dose retains prior exposure.

3.3 THE WUIRL POLICY

The preceding sections define the environment WUIRL acts within, which is the exposure-pathway action space $\mathcal{A}$ (Section 3.1) and the physical transition $\mathcal{P}$ (Section 3.2). This section defines the

agent, its state $\mathcal{S}$, the policy that acts on it, and the reward $\mathcal{R}$ it is trained to maximize. Decisions are indexed by k at physical time t_k, and several decisions may be made before time advances.

State. The state s_k comprises building fire status, component dose, source ages and vegetation histories, external-entry clocks, current selection plans, physical time, and scoring records. Recorded initial ignitions start the rollout and are excluded from scoring. Damage supplies endpoint targets, while progression proxies supply timing targets and interface conditions. Reconstruction is therefore retrospectively conditioned.

Observation and policy. The policy acts on a compressed observation $o_k = \phi(s_k)$ carrying pooled graph context and the decision masks, rather than acting on s_k directly. This observation is not assumed to retain the full simulator state, so the problem is partially observed (Appendix C). The graph context comes from a two-layer mean-aggregation GraphSAGE encoder that maps standardized building attributes to embeddings on the community graph (Hamilton et al., 2017), pretrained by graph link prediction and frozen during policy training, so the simulated fire state enters only through pooled context and pathway features. Neural-network actor heads score receivers and conditional pathway choices, while a separate critic estimates future discounted reconstruction reward from pooled context. The policy may select multiple pathways before committing exposure, or decline to add further exposure at any point, representing exposure blocked by unmodeled factors such as shielding or suppression. Acting on o_k, the policy factorizes the two-stage action into a receiver and a conditional exposure pathway or stopping action.

$$\pi_{\boldsymbol{\theta}}(a_k \mid o_k) = \pi^{\mathrm{rec}}_{\boldsymbol{\theta}}(i_k \mid o_k)\, \pi^{\mathrm{path}}_{\boldsymbol{\theta}}(b_k \mid i_k, o_k) \tag{2}$$

Here, $a_k = (i_k, b_k)$ consists of receiver i_k and pathway or stopping choice b_k, and $\boldsymbol{\theta}$ denotes the actor parameters. Each distribution is normalized over its admissible choices by action masking (Huang & Ontañón, 2022). This autoregressive factorization keeps the action space tractable, since the joint space of receiver and pathway choices grows with the number of burning sources. The full selection-and-commit procedure, including within-hour retention, the hour-boundary reset, and the external-entry thresholds, is given in Appendix C.

Reward. The reward scores cumulative progression and final outcomes. Let F_h measure cumulative ignition agreement at hour h and O final-damage agreement, using class-balanced binary log scores with double weight for boundary buildings. The reconstruction score is

$$R = \frac{1}{N_h} \sum_{h=1}^{N_h} F_h + 2O \qquad N_h = 15 \tag{3}$$

where the reward pays F_h/N_h as each hour closes and $2O$ at the end, so per-step rewards r_k sum to R. Its weights are design choices rather than calibrated observation errors, with the exact score in Appendix D.

Learning. We train the policy using clipped proximal policy optimization (Schulman et al., 2017) with invalid-action masking (Huang & Ontañón, 2022). As several decisions can occur at the same physical time, the process is semi-Markov (Bradtke & Duff, 1995; Sutton et al., 1999), so generalized advantage estimation (Schulman et al., 2015) discounts by elapsed simulated time, and selections that do not advance physical time incur no additional discount. The training objective is

$$J(\boldsymbol{\theta}) = \mathbb{E}_{\tau \sim (\pi_{\boldsymbol{\theta}}, \mathcal{P})} \left[\sum_k \gamma^{(t_k - t_0)/(1\,\mathrm{min})} r_k \right] \qquad \sum_k r_k = R(\tau) \tag{4}$$

where τ is a rollout generated by $\pi_{\boldsymbol{\theta}}$ under $\mathcal{P}$, and γ discounts reward per elapsed simulated minute. Checkpoint selection and reporting use the undiscounted score R. Optimization details are in Appendix D, and the training protocol is in Section 4.2.

4 Experiments

4.1 Eaton data and physical parameterization

We reconstruct the 2025 Eaton fire, comprising 18,384 buildings, using Microsoft Building Footprints (Microsoft, 2022), construction attributes and damage records from CAL FIRE's Damage

Inspection (DINS) program (California Department of Forestry and Fire Protection, 2025), and progression-derived timing proxies (Kerber et al., 2025). Pre-fire vegetation is mapped from 2022 NAIP and Sentinel-2 imagery (Coleman et al., 2020; European Union/ESA/Copernicus, 2015–2026; USDA Farm Production and Conservation, Business Center, Geospatial Enterprise Operations, 2022), with tree canopy geometry extracted using a pretrained segmentation model (Yoo et al., 2026). Recorded damage supplies a positive ignition proxy and no damage supplies a negative proxy, leaving 18,271 scored buildings after exclusions. Physical parameters are assigned from geometry, construction categories, and literature relationships. Data processing and parameterization are detailed in Appendices A and B.

4.2 Training and evaluation protocol

We trained three PPO policies for 500,000 actions each, where an action is one policy decision rather than a complete fire episode. Screening selected a learning rate of 10^{-4} and pathway-entropy coefficient of 0.001. Physical-time discount and GAE parameters were $\gamma = 0.9999$ and $\lambda = 0.95$. For each policy, we selected the checkpoint with the highest mean reconstruction score over three selection seeds, then evaluated it on ten disjoint reporting seeds. This separates checkpoint selection from reporting randomness, but does not hold out Eaton observations. We compare WUIRL with Random, a uniform policy over masked actions, and Greedy, a myopic one-step policy that ranks actions by incident flux relative to critical flux. Both baselines use the same environment, constraints, scorer, and reporting seeds. We report the reconstruction score, survivor recall, burn precision, recall, F1, balanced accuracy, and exact-hour recall. Screening, seed lists, and exact baseline rules are given in Appendix D.

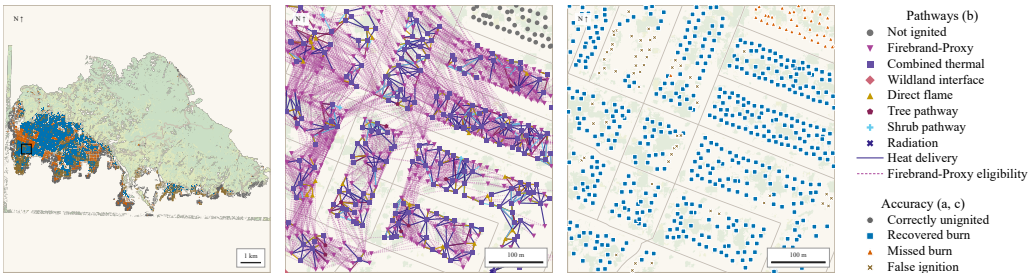
Ablations and diagnostic experiments. We individually disable radiation, direct flame, vegetation-mediated exposure, and firebrand-proxy entry in the three frozen selected policies, using ten paired reporting seeds per policy and condition. These evaluation-time ablations measure mechanism dependence without retraining. Separately, we retrain without the temporal reward and compare final checkpoints on the selection seeds, and we evaluate cumulative ignition agreement through hours 2–15 of fixed rollouts from the selected policies. A spatial holdout excludes a geographic region and buffer from training computations and checkpoint selection before evaluating the restored region. Protocols and results are detailed in Appendices D and E.

4.3 Reconstructing the Eaton fire

Learned selection gives the best balance between recovering burned and surviving structures based on reconstruction score and balanced accuracy. Across the Eaton fire, WUIRL recovers 78.9% of observed burned buildings (burn recall) while preserving 52.8% of observed survivors (survivor recall, Table 1). WUIRL's improvement in reconstruction score over Greedy (0.172) is more than twice Greedy's improvement over Random (0.075). The two baselines sit at opposite ends of this tradeoff. Greedy recovers more burns (87.4%) and has the highest burn F1, but preserves only 40.1% of survivors. Random preserves more survivors (71.3%) but only recovers 54.8% of burns. WUIRL exceeds Greedy's reconstruction reward on all ten reporting seeds for training seeds 42 and 137, and on six of ten for seed 2027. All three selected policies achieve higher mean reward than Greedy. In matched-state comparisons from one rollout, WUIRL selected 34–37% fewer pathways than Greedy at hours 2, 4, and 6 and produced fewer next-minute ignitions (Table 6). This suggests that the learned policy can adopt a more selective exposure strategy than the Greedy baseline.

Figure 1 shows what is included in a reconstruction. Across the study area (a), recovered burns, missed burns, false ignitions, and correctly unignited buildings are resolved building by building, so errors can be located spatially rather than only counted. In the enlarged neighborhood, the replay records the pathway through which each building ignited (b). Solid lines show simulated heat delivery, and dashed lines connect sources that satisfied firebrand-proxy eligibility, which are possible exposure links rather than identified ember trajectories. Comparing the same buildings with damage records (c) ties each modeled pathway to a specific building's success or error.

The same replay addresses questions that damage records alone cannot address. Among its 5,655 thermally ignited buildings, which exclude firebrand-proxy ignitions (Section 4.4), 27.0% received heat only through building radiation or opening flames, 6.3% only through vegetation pathways, and



(a) Eaton study area (b) Selected ignition pathways (c) Ignition accuracy

Figure 1: Spatial reconstruction of the Eaton fire (training seed 42 at 50k actions, evaluation seed 2001). (a) Reconstruction outcomes across the study area, with the enlarged neighborhood outlined. (b) Simulated ignition mechanisms. Solid lines show simulated heat delivery, and dashed lines show sources eligible for firebrand-proxy entry, not identified ember donors. (c) Predicted eventual ignition compared with damage records.

Table 1: Reconstruction performance on the Eaton fire. Values are means ± evaluation seed standard deviations over ten common reporting seeds. WUIRL averages three independently trained policies, and its standard deviations pool within-policy evaluation variances. Across training seeds, policy means have sample standard deviations of 0.058 for reward and 2.35 points for burn recall. Balanced accuracy is the mean of burn and survivor recall. Greedy is deterministic, so it has no standard deviation. Exact-hour recall uses observed destroyed buildings with known timing. Bold indicates the highest mean, not statistical significance.

	Reward ↑	**Survivor recall (%)** ↑	**Burn recall (%)** ↑	**Burn precision (%)** ↑	**Burn F1** ↑	**Balanced accuracy** ↑	**Exact-hour recall (%)** ↑
Random	-4.582 ± 0.142	$\mathbf{71.3 \pm 1.6}$	54.8 ± 3.4	$\mathbf{71.6 \pm 1.2}$	0.621 ± 0.025	0.631	8.1 ± 1.0
Greedy	-4.507	40.1	**87.4**	65.8	**0.751**	0.637	10.1
WUIRL	$\mathbf{-4.335 \pm 0.117}$	52.8 ± 2.0	78.9 ± 3.0	68.8 ± 0.9	0.735 ± 0.016	**0.658**	$\mathbf{10.7 \pm 0.9}$

66.7% through both. Therefore, 93.7% received at least some heat from neighboring structures. The most connected source buildings each delivered heat to eight distinct buildings that subsequently ignited thermally.

Given that all policies share the same physics and scorer, score differences reflect pathway selection. Greedy selects by immediate flux alone, so WUIRL's advantage is consistent with a benefit from sequential selection, although this comparison does not isolate it (Table 6).

4.4 Dependence on modeled exposure mechanisms

We find that reconstructions depend primarily on radiant heat transfer between neighboring structures and on firebrand-proxy entry. Removing building radiation or firebrand-proxy entry from the frozen selected policies reduces burn recall from 78.9% to 9.2% and 9.8%, respectively, whereas removing direct flame contact or radiation from burning trees and shrubs near the structure reduces it only to 67.1% and 67.9% (Figure 4). Removing building radiation also suppresses firebrand-proxy entry, as firebrand-proxy entry requires 900 seconds of prior selected thermal exposure (Appendix B.6), and building radiation is one of its sources.

Recovered ignitions divide almost evenly between direct thermal pathways (48.0%) and firebrand-proxy entry (48.6%), with wildland-interface entry accounting for the remaining 3.4%. Firebrand-proxy entry represents ignition by firebrands through an eligibility rule rather than an explicit transport model (Section 3.1), so its share indicates how often the model attributes ignition to embers without identifying a specific ember source or trajectory.

Eligibility rules also shape reconstruction. In an exploratory paired test with a single policy and evaluation seed, removing the thermal exposure prerequisite for firebrand-proxy entry raises burn recall to 99.5% but lowers survivor recall to 5.7% (Appendix E). Discrimination therefore depends on constraining when ignition is available, not only on providing it. Reconstructions thus rest largely on two routes, one of which represents embers only through an eligibility rule, which makes explicit ember generation and transport modeling a direct target for improving the environment.

4.5 Temporal reconstruction

Reconstruction recovers whether buildings burn more reliably than when they ignite. Exact-hour recall is 10.7% for WUIRL, 10.1% for Greedy, and 8.1% for Random (Table 1), so learned selection gives little timing advantage. Simulated progression may run ahead of the observed record. Across all three selected policies, cumulative burn precision is 7.0% through hour 4 and rises to 69.0% through hour 12 as observed ignitions enter the comparison (Appendix E), although these early predictions also include false ignitions of observed survivors.

Reachability limits reconstruction. Candidate availability accounts for most missed burns. For buildings observed to ignite in hours 10–15, 97.9% of those ever offered to the policy are eventually recovered, and 92.9% of the missed buildings were never offered (Appendix E). When the reconstruction offers a pathway to a building that burned, it usually recovers that ignition, so the main bottleneck is whether the simulated fire reaches the building at all. WUIRL is therefore best read as reconstructing which buildings burn, with timing as a weaker secondary signal.

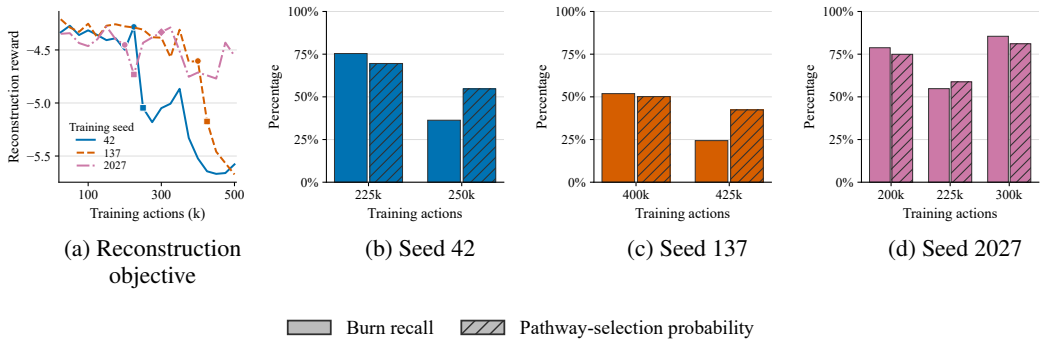
4.6 Spatial holdout

Reconstruction of a withheld region of the Eaton fire is less consistent across training seeds. We withheld the westernmost region and a 100 m buffer from training and checkpoint selection (Appendix E.5). As the holdout protocol differs from the main experiment, we compare policies within the holdout rather than against the full-fire results. Mean held-out burn and survivor recall are 55.3% and 61.2%, compared with 48.1% and 61.7% for Random and 81.9% and 41.5% for Greedy. WUIRL recovers more burns than Random at similar survivor recall, while Greedy retains higher burn recall and F1. Held-out burn recall ranges from 21.8% to 79.9% across training seeds and broadly tracks each policy's recall on the training region. For the weakest policy, 94.4% of missed held-out burns were never offered as candidates, consistent with the reachability limit in Section 4.5.

4.7 Training stability

We find that continued training degrades reconstruction. Selected checkpoints, at 50k to 150k actions, reach 78.9% burn recall on reporting seeds, compared with 22.2% at the final 500k checkpoints. The decline also appears at fixed checkpoints. From 50k to 500k actions, recall on selection seeds falls from 80.1% to 6.6%, 66.3% to 3.4%, and 75.9% to 62.5% for the three training seeds (Appendix E.6). The decline coincides with a shift in behavior. On identical saved decision states, later policies select exposure pathways less often, with partial recovery for seed 2027 (Figure 2). Continued training also worsens the reachability limit identified in Section 4.5. In reporting rollouts, the fraction of observed burned buildings never offered as candidates rises from 19.7% at selected checkpoints to 72.7% at final checkpoints. Since fewer selections activate fewer sources, reduced selection plausibly compounds through the cascade, which may explain how a modest change in selection produces a large loss in recall.

Updates remain small throughout, which suggests a gradual drift rather than unstable optimization. This pattern resembles reported loss of plasticity in deep RL (Lyle et al., 2023; Nikishin et al., 2022), although we do not test that mechanism directly. In continuation runs, update KL divergence stays below 0.0025 and gradient clipping occurs in 20 of 1,248 steps (Appendix E.8). Delayed credit is a plausible contributor, since reward settles hourly and the GAE trace retains 0.046 of its weight after one simulated hour, so many decisions depend on a critic whose explained variance is negative even at high recall (Appendix E.7). The two tested remedies did not resolve this. Setting $\lambda = 1$ worsened the decline, and adding current exposure plan inputs to the critic gave inconsistent results across continuation seeds. For reinforcement learning on this simulator, checkpoint selection on separate evaluation seeds is therefore necessary rather than optional.



(a) Reconstruction objective (b) Seed 42 (c) Seed 137 (d) Seed 2027

Figure 2: Training deterioration and checkpoint behavior. (a) Reward for each training seed, averaged over three selection seeds, with markers at the checkpoints examined in (b–d). (b–d) Burn recall (solid) and pathway selection probability on 20 fixed saved states, restricted to free path choices (hatched). Each metric is scaled to its own panel and is best compared across checkpoints rather than against the other metric. The saved states are exploratory and not a representative sample.

4.8 Limitations and Future Work

Validation. Damage records and coarse progression proxies cannot verify what ignited each building or when. A learned policy could partly compensate for errors in the model's physical parameters and still reproduce the observed outcomes, so agreement with damage records is not evidence that those parameters are individually correct. For the same reason, the mechanism removals in Section 4.4 are a sensitivity analysis of what the trained model depends on, not a validation of the underlying physics or an identification of each building's historical ignition cause. Future work could compare reconstructed routes against independent exposure observations and test alternative parameterizations, such as dose decay and explicit ember transport.

Evaluation design. Initialization, timing targets, interface conditions, and checkpoint selection all draw on Eaton, so the main results are a conditional reconstruction of a single fire rather than prospective prediction. The spatial holdout reduces this overlap within Eaton but does not test transfer across fires. We report three training seeds, and variability between them is substantial. Full physical rollouts make evaluation, not training, the main computational cost, which limits the number of training seeds. Additional training seeds and additional fires with varied building inventories, weather, and observation quality would test broader applicability.

Failure modes. False ignitions and missed burns remain substantial (Figure 1), and most missed burns are buildings the reconstruction never reaches. Separating physical model error, restrictive eligibility rules, and learned selection as sources of these misses would require controlled tests of those assumptions and finer progression data. Whether learned selection outperforms the baselines by anticipating downstream exposure or by preferring different immediate selections also remains open. The training decline likewise lacks an identified initiating cause, which controlled interventions on learning updates near a decline could address. With stronger validation, the framework could support controlled studies of how mitigation changes exposure and building loss.

5 Conclusion

We introduced WUIRL, which couples learned pathway selection with fixed ignition dynamics to reconstruct community fire progression. On the Eaton fire, it improves reconstruction score and balanced accuracy over the baselines while making modeled exposure routes inspectable. In one replay, 93.7% of thermal ignitions received heat from neighboring structures, and most missed burns were buildings the simulated fire never reached, pointing to exposure reach rather than ignition response as the main limit. This case study provides a basis for testing reconstruction on richer observations, alternative physical models, and additional fires.

AI USE STATEMENT

In this work, we used generative AI tools to provide feedback on research methodology and experiments, to implement methods, and to support data analysis and the interpretation of results. We have not used generative AI tools to generate synthetic datasets, formulate mathematical claims, write proofs, or assist with translation, and the remaining required disclosure tasks are not applicable to this work. Additionally, we used generative AI tools to create and edit software code, prepare scientific figures, draft and edit parts of the manuscript for readability, identify relevant literature, and format references. We have reviewed all AI-assisted work. All experimental results were produced by running the authors' experiments, AI-assisted code was tested and verified by the authors, and suggested citations were checked against the original sources. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Appendix A lists all data sources and describes how observations and progression proxies are constructed. Appendix B specifies the exposure and ignition model with its parameters and derivations. Appendix C describes the state, encoder, policy architecture, and action execution. Appendix D gives the reconstruction score, optimization details, training settings (Table 4), random seeds, checkpoint selection, metrics, baseline definitions, and computational requirements (Appendix D.8).

REFERENCES

Abdulelah Altamimi, Constantino Lagoa, José G. Borges, Marc E. McDill, C. P. Andriotis, and K. G. Papakonstantinou. Large-scale wildfire mitigation through deep reinforcement learning. *Frontiers in Forests and Global Change*, 5, 2022. ISSN 2624-893X. doi: 10.3389/ffgc.2022.734330.

Paola F. Antonietti, Matteo Caldana, Ilario Mazzieri, and Andrea Re Fraschini. Magnet: an open-source library for mesh agglomeration by graph neural networks. *Engineering with Computers*, 41(6):4825–4850, 2025. doi: 10.1007/s00366-025-02223-y.

Jennifer K Balch, Virginia Iglesias, Adam L Mahood, Maxwell C Cook, Cibele Amaral, Amy DeCastro, Stefan Leyk, Tyler L McIntosh, R Chelsea Nagy, Lise St. Denis, et al. The fastest-growing and most destructive fires in the us (2001 to 2020). *Science*, 386(6720):425–431, 2024.

Somnath Bar, Shu Li, and Tirtha Banerjee. Explaining building damage from wildfires in california. *Science Advances*, 12(30):eaed4197, 2026. doi: 10.1126/sciadv.aed4197.

Steven J. Bradtke and Michael O. Duff. Reinforcement learning methods for continuous-time markov decision problems. In *Advances in Neural Information Processing Systems*, volume 7, 1995.

California Department of Forestry and Fire Protection. Cal fire damage inspection (DINS) data, 2025. URL `https://gis.data.cnra.ca.gov/datasets/CALFIRE-Forestry::cal-fire-damage-inspection-dins-data`.

California Department of Forestry and Fire Protection. California fire perimeters, 2026. URL `https://data.ca.gov/dataset/california-fire-perimeters-all`. Version firep25_1, released April 2026.

Amanda R. Carlson, Todd J. Hawbaker, Miranda H. Mockrin, Volker C. Radeloff, Lucas S. Bair, Michael D. Caggiano, James R. Meldrum, Patricia M. Alexandre, H. Anu Kramer, and Paul F. Steblein. Rising rates of wildfire building destruction in the conterminous United States. *Proceedings of the National Academy of Sciences*, 122(51):e2505886122, 2025. doi: 10.1073/pnas.2505886122.

Sara E. Caton, Raquel S. P. Hakes, Daniel J. Gorham, Aixi Zhou, and Michael J. Gollner. Review of pathways for building fire spread in the wildland urban interface part I: Exposure conditions. *Fire Technology*, 53(2):429–473, 2017. doi: 10.1007/s10694-016-0589-z.

Chloe K. Chou, Amara L. Holder, Adam Nored, Glenn Walters, Wubin Bai, M. Ian Gilmour, Yong Ho Kim, and Julia E. Rager. Burning "exposhome": Deriving a mixture of combustible materials in american homes at the wildland-urban interface for health studies. *GeoHealth*, 9(12), 2025. doi: 10.1029/2025GH001439.

Akshat Chulahwat, Hussam Mahmoud, Santiago Monedero, Francisco Jośe Diez Vizcaíno, Joaquin Ramirez, David Buckley, and Adrián Cardil Forradellas. Integrated graph measures reveal survival likelihood for buildings in wildfire events. *Scientific Reports*, 12:15954, 2022. doi: 10.1038/s41598-022-19875-1.

Jack D. Cohen. The wildland-urban interface fire problem. *Fremontia*, 38(2–3):16–22, 2010.

Jack D. Cohen and Bret W. Butler. Modeling potential structure ignitions from flame radiation exposure with implications for wildland/urban interface fire management. In *Proceedings of the 13th Fire and Forest Meteorology Conference*, pp. 81–86. International Association of Wildland Fire, 1998.

Red Willow Coleman, Natasha Stavros, Vineet Yadav, and Nicholas Parazoo. A simplified framework for high-resolution urban vegetation classification with optical imagery in the Los Angeles megacity. *Remote Sensing*, 12(15):2399, 2020. doi: 10.3390/rs12152399.

Mark A. Dietenberger and Charles R. Boardman. EcoSmart Fire as structure ignition model in wildland urban interface: Predictions and validations. *Fire Technology*, 53:577–607, 2017. doi: 10.1007/s10694-016-0632-0.

Yifei Ding, Thomas Gernay, Xuechun Li, Negar Elhami-Khorasani, Susu Xu, and Xinyan Huang. Ai assessment of structural fuel load and fire risk via street house images in wildland-urban interface. *Applications in Energy and Combustion Science*, 27:100525, 2026. doi: 10.1016/j.jaecs.2026.100525.

Naveed Ejaz and Salimur Choudhury. A comprehensive survey of the machine learning pipeline for wildfire risk prediction and assessment. *Ecological Informatics*, 90:103325, 2025. doi: 10.1016/j.ecoinf.2025.103325.

Miguel Esparza, Vamshi Battala, and Ali Mostafavi. A graph neural network and decision tree modeling approach for predicting wildfire-induced building damage. *Computer-Aided Civil and Infrastructure Engineering*, 47:100085, 2026. doi: 10.1016/j.cacaie.2026.100085.

European Union/ESA/Copernicus. Harmonized Sentinel-2 MSI: Multispectral instrument, level-1c top-of-atmosphere reflectance. Google Earth Engine Data Catalog, 2015–2026. URL `https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_HARMONIZED`. Collection COPERNICUS/S2_HARMONIZED.

Alexander I. Filkov, Virginie Tihay-Felicelli, Nima Masoudvaziri, David Rush, Andres Valencia, Yu Wang, David L. Blunck, Mario Miguel Valero, Kamila Kempna, Jan Smolka, Jacques De Beer, Zakary Campbell-Lochrie, Felipe Roman Centeno, Muhammad Asim Ibrahim, Calisa Katiuscia Lemmertz, and Wai Cheong Tam. A review of thermal exposure and fire spread mechanisms in large outdoor fires and the built environment. *Fire Safety Journal*, 140:103871, 2023. doi: 10.1016/j.firesaf.2023.103871.

Mark A. Finney. FARSITE: Fire area simulator—model development and evaluation. Technical Report RMRS-RP-4, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Ogden, UT, 1998.

Sriram Ganapathi Subramanian and Mark Crowley. Using spatial reinforcement learning to build forest wildfire dynamics models from satellite images. *Frontiers in ICT*, 5:6, 2018. doi: 10.3389/fict.2018.00006.

Javier González-Villa, David Lázaro, Arturo Cuesta, Adriana Balboa, Daniel Alvear, and Mariano Lázaro. Identifying critical fire spread to the wildland–urban interface using cellular automata and reinforcement learning. *Machine Learning with Applications*, 22:100779, 2025. doi: 10.1016/j.mlwa.2025.100779.

Ravi N. Haksar and Mac Schwager. Distributed deep reinforcement learning for fighting forest fires with a network of aerial robots. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1067–1074. IEEE, 2018. doi: 10.1109/IROS.2018.8593539.

William L Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, volume 30, pp. 1024–1034, 2017.

Longhua Hu, Xiepeng Sun, Xiaolei Zhang, and Fei Ren. Facade flame height and horizontal extending distance from opening of compartment fire with external sideward wind. *Proceedings of the Combustion Institute*, 37(3):3859–3867, 2019. doi: 10.1016/j.proci.2018.06.201.

Shengyi Huang and Santiago Ontañón. A closer look at invalid action masking in policy gradient algorithms. In *Proceedings of the International FLAIRS Conference*, volume 35, 2022. doi: 10.32473/flairs.v35i.130584.

Yunzhi Huang and Krishnan Gowri. Analysis of iecc (2003, 2006, 2009) and ashrae 90.1-2007 commercial energy code requirements for mesa, az. Technical report, Pacific Northwest National Laboratory, 02 2011.

Yiming Jia and Eyitayo A. Opabola. Interpretable machine learning insights into wildfire damage drivers in california, usa. *International Journal of Disaster Risk Reduction*, 126:105610, 2025. doi: 10.1016/j.ijdrr.2025.105610.

Wenyu Jiang, Fei Wang, Linghang Fang, Xiaocui Zheng, Xiaohui Qiao, Zhanghua Li, and Qingxiang Meng. Modelling of wildland-urban interface fire spread with the heterogeneous cellular automata model. *Environmental Modelling & Software*, 135:104895, 2021.

Steve Kerber, Derek Alkonis, Daniel Gorham, and Gavin Horn. Southern california fires timeline report. Technical report, UL Research Institutes, Fire Safety Research Institute, Columbia, MD, 2025.

Chris Lautenberger. Wildland fire modeling with an eulerian level set method and automated calibration. *Fire Safety Journal*, 62:289–298, 2013. doi: 10.1016/j.firesaf.2013.08.014.

Yee-Ping Lee, Michael A. Delichatsios, and G. W. H. Silcock. Heat fluxes and flame heights in facades from fires in enclosures of varying geometry. *Proceedings of the Combustion Institute*, 31(2):2521–2528, 2007. doi: 10.1016/j.proci.2006.08.033.

Yee-Ping Lee, Michael A. Delichatsios, Yoshifumi Ohmiya, Kaoru Wakatsuki, Akito Yanagisawa, and Daisuke Goto. Heat fluxes on opposite building wall by flames emerging from an enclosure. *Proceedings of the Combustion Institute*, 32(2):2551–2558, 2009. doi: 10.1016/j.proci.2008.06.130.

Fangjun Li, Xiaoyang Zhang, Mark Cochrane, Shobha Kondragunta, and Shuai An. Fire spread, intensity, and emissions observations by multiple satellites: The Southern California wildfires of january 2025. *AGU Advances*, 7(1):e2025AV002064, 2026. doi: 10.1029/2025AV002064.

Stuart P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28 (2):129–137, 1982. doi: 10.1109/TIT.1982.1056489.

Andrew Lock, Matthew Bundy, Erik L. Johnsson, Anthony Hamins, Gwon Hyun Ko, Cheolhong Hwang, Paul Fuss, and Richard Harris. Experimental study of the effects of fuel type, fuel distribution, and vent size on full-scale underventilated compartment fires in an iso 9705 room. Technical Report NIST Technical Note 1603, National Institute of Standards and Technology, 2008.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2019.

Clare Lyle, Zeyu Zheng, Evgenii Nikishin, Bernardo Avila Pires, Razvan Pascanu, and Will Dabney. Understanding plasticity in neural networks. In *International Conference on Machine Learning*, 2023.

Nima Masoudvaziri, Fernando Szasdi Bardales, Oguz Kaan Keskin, Amir Sarreshtehdari, Kang Sun, and Negar Elhami-Khorasani. Streamlined wildland-urban interface fire tracing (SWUIFT): Modeling wildfire spread in communities. *Environmental Modelling & Software*, 143:105097, 2021. doi: 10.1016/j.envsoft.2021.105097.

Epaminondas Mastorakos, Savvas Gkantonas, Georgios Efstathiou, and Andrea Giusti. A hybrid stochastic lagrangian–cellular automata framework for modelling fire propagation in inhomogeneous terrains. *Proceedings of the Combustion Institute*, 39(3):3853–3862, 2023.

Shaurya Mathur, Shreyas Bellary Manjunath, Nitin Kulkarni, and Alina Vereshchaka. Spatiotemporal wildfire prediction and reinforcement learning for helitack suppression. In *2025 International Conference on Machine Learning and Applications (ICMLA)*, pp. 766–771. IEEE, 2025. doi: 10.1109/ICMLA66185.2025.00111.

M. McKinnon, C. Weinschenk, N. Dow, M. DiDomizio, and D. Madrzykowski. Materials and products database (version 1.0.0). Fire Safety Research Institute, UL Research Institutes, 2023. URL `https://materials.fsri.org`.

E. Gregory McPherson, Natalie S. van Doorn, and Paula J. Peper. Urban tree database and allometric equations. Technical Report PSW-GTR-253, U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station, Albany, CA, 2016a.

E. Gregory McPherson, Natalie S. van Doorn, and Paula J. Peper. Urban tree database, 2016b. URL `https://doi.org/10.2737/RDS-2016-0005`. Updated 21 January 2020.

Microsoft. Global ML building footprints. GitHub repository, 2022. URL `https://github.com/microsoft/GlobalMLBuildingFootprints`.

Miranda H Mockrin, Dexter H Locke, Alexandra D Syphard, and Jarlath O'Neil-Dunne. Using high-resolution land cover data to assess structure loss in the 2018 woolsey fire in southern california. *Journal of Environmental Management*, 347:118960, 2023.

Xareni Monroy, Evan Sluder, Faraz Hedayati, Hank Pogorzelski, and Murray Morrison. The 2025 LA conflagrations. Technical report, Insurance Institute for Business & Home Safety, 2025.

J. Muñoz Sabater, E. Dutra, A. Agustí-Panareda, C. Albergel, G. Arduini, G. Balsamo, S. Boussetta, M. Choulga, S. Harrigan, H. Hersbach, B. Martens, D. G. Miralles, M. Piles, N. J. Rodríguez-Fernández, E. Zsoter, C. Buontempo, and J.-N. Thépaut. Era5-land: a state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9):4349–4383, 2021. doi: 10.5194/essd-13-4349-2021.

Lucas Murray, Tatiana Castillo, Isaac Martín de Diego, Richard Weber, José Ramón González-Olabarria, Jordi García-Gonzalo, Andrés Weintraub, and Jaime Carrasco-Barra. Deep reinforcement learning for optimal firebreak placement in forest fire prevention. *Applied Soft Computing*, 175:113043, 2025. doi: 10.1016/j.asoc.2025.113043.

M. Z. Naser and V. Kodur. Vulnerability of structures and infrastructure to wildfires: a perspective into assessment and mitigation strategies. *Natural Hazards*, 121:9995–10015, 2025. doi: 10.1007/s11069-025-07168-5.

Harold E. Nelson and Scot Deal. Comparing compartment fires with compartment fire models. In *Fire Safety Science—Proceedings of the Third International Symposium*, pp. 719–728, 1991. doi: 10.3801/IAFSS.FSS.3-719.

Mingshuo Nie, Dongming Chen, and Dongqi Wang. Reinforcement learning on graphs: A survey. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 7(4):1065–1082, 2023. doi: 10.1109/TETCI.2022.3222545.

Evgenii Nikishin, Max Schwarzer, Pierluca D'Oro, Pierre-Luc Bacon, and Aaron Courville. The primacy bias in deep reinforcement learning. In *International Conference on Machine Learning*, 2022.

A. Noumeur and M. Z. M. Tohir. Predicting california wildfire damage to structures using machine learning: A comparative study of random forest and XGBoost. *Journal of Physics: Conference Series*, 3121(1):012029, 2025. doi: 10.1088/1742-6596/3121/1/012029.

Dwi M.J. Purnomo, Yiren Qin, Maria Theodori, Maryam Zamanialaei, Chris Lautenberger, Arnaud Trouvé, and Michael Gollner. Reconstructing modes of destruction in wildland–urban interface fires using a semi-physical level-set model. *Proceedings of the Combustion Institute*, 40(1): 105755, 2024. ISSN 1540-7489. doi: 10.1016/j.proci.2024.105755.

James G. Quintiere. *Fundamentals of Fire Phenomena*. John Wiley & Sons, Chichester, England, 2006. doi: 10.1002/0470091150.

James G. Quintiere and Bernard J. McCaffrey. The burning of wood and plastic cribs in an enclosure: Volume i. Technical Report NBS IR 80-2054v1, National Bureau of Standards, 1980.

Volker C. Radeloff, Miranda H. Mockrin, David Helmers, Amanda Carlson, Todd J. Hawbaker, Sebastian Martinuzzi, Franz Schug, Patricia M. Alexandre, H. Anu Kramer, and Anna M. Pidgeon. Rising wildfire risk to houses in the united states, especially in grasslands and shrublands. *Science*, 382(6671):702–707, 2023. doi: 10.1126/science.ade9223.

John Schulman, Philipp Moritz, Sergey Levine, Michael Jordan, and Pieter Abbeel. High-dimensional continuous control using generalized advantage estimation, 2015. URL https://arxiv.org/abs/1506.02438.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms, 2017. URL https://arxiv.org/abs/1707.06347.

Alex Schutz and Victor-Alexandru Darvariu. Using graph neural networks in reinforcement learning: A practical guide. In *ICLR Blogposts 2026*, 2026. URL https://iclr-blogposts.github.io/2026/blog/2026/rl-with-gnns/. https://iclr-blogposts.github.io/2026/blog/2026/rl-with-gnns/.

G. W. H. Silcock and T. J. Shields. A protocol for analysis of time-to-ignition data from bench scale tests. *Fire Safety Journal*, 24(1):75–95, 1995. doi: 10.1016/0379-7112(95)00003-C.

Alexander A. Soderlund and Mrinal Kumar. Markovian wildfire modeling via evidential reasoning. In *2018 IEEE Conference on Decision and Control (CDC)*, pp. 5592–5597, 2018. doi: 10.1109/CDC.2018.8619298.

Richard S. Sutton, Doina Precup, and Satinder Singh. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112(1–2):181–211, 1999.

U.S. Census Bureau. 2025 TIGER/Line shapefiles, 2025. URL https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html. Los Angeles County, California, road data.

USDA Farm Production and Conservation, Business Center, Geospatial Enterprise Operations. National agriculture imagery program (NAIP): 2022 imagery. Google Earth Engine Data Catalog, 2022. URL https://developers.google.com/earth-engine/datasets/catalog/USDA_NAIP_DOQQ. Collection USDA/NAIP/DOQQ.

Jooyoung Yoo, Yi Qi, Isaac Ashe-McNalley, Beau MacDonald, and John P. Wilson. An open and transferable deep learning framework for mapping urban tree canopy using NAIP imagery. *Remote Sensing*, 18(12):1899, 2026. doi: 10.3390/rs18121899.

Bryce A. Young, Matthew P. Thompson, Christopher J. Moran, and Carl A. Seielstad. Modeling neighborhoods as fuel for wildfire: A review. *Fire Technology*, 61(7):5049–5071, 2025. doi: 10.1007/s10694-025-01773-3.

Maryam Zamanialaei, Daniel San Martin, Maria Theodori, Dwi Marhaendro Jati Purnomo, Ali Tohidi, Chris Lautenberger, Yiren Qin, Arnaud Trouvé, and Michael Gollner. Fire risk to structures in california's wildland-urban interface. *Nature Communications*, 16:8041, 2025. doi: 10.1038/s41467-025-63386-2.

APPENDIX OVERVIEW

Appendix A describes the Eaton data and how observations are constructed. Appendix B specifies the exposure and ignition model and its parameters. Appendix C describes the policy and environment implementation, and Appendix D gives training settings, checkpoint selection, metrics, baselines, and computational requirements. Appendix E presents supplementary results, including training stability diagnostics.

A EATON DATA AND OBSERVATION CONSTRUCTION DETAILS

This appendix details the data summarized in Section 4.1.

A.1 BUILDINGS AND RECORDED OUTCOMES

Microsoft building footprints provide plan geometry (Microsoft, 2022). CAL FIRE Damage Inspection (DINS) records provide construction attributes and damage outcomes (California Department of Forestry and Fire Protection, 2025). Affected, Minor, Major, and Destroyed labels within DINS are positive ignition proxies, and No Damage is negative. The final scoring set contains 18,271 buildings, with 10,392 positive and 7,879 negative proxies. Of the burned buildings, 395 lack timing information and are evaluated only on final outcomes. The 71 initially ignited buildings, 41 buildings with unknown outcomes, and one with conflicting records remain in the simulation but are excluded from scoring. These initially ignited buildings retrospectively condition reconstruction, and assigning final damage to the 15-hour endpoint is an observation-model assumption.

Component observations are often missing in heavily damaged structures (Monroy et al., 2025, p. 70, Tables 5–6). We assign missing hardening attributes as vulnerable and missing connective-fuel indicators as available pathways. Attribute-missingness flags are retained for provenance but excluded from policy inputs.

A.2 PRE-FIRE LAND COVER AND CANOPY GEOMETRY

We construct a land-cover map from 2022 National Agriculture Imagery Program (NAIP) and Sentinel-2 imagery (European Union/ESA/Copernicus, 2015–2026; USDA Farm Production and Conservation, Business Center, Geospatial Enterprise Operations, 2022), adapting the urban-vegetation classification of Coleman et al. (2020). A pretrained NAIP canopy-segmentation model (Yoo et al., 2026) supplies canopy geometry, subdivided into effective crown units using regional tree-size priors (McPherson et al., 2016a).

A.3 GEOMETRY AND ROAD CONTEXT

Footprints determine plan dimensions, separation, and intervening-building obstruction, but not height or roof geometry. DINS structures sharing a footprint remain separate nodes, reconstructed using peer-based dimensions and a near-zero-gap partition, with geometry provenance retained. Roads are obtained from the 2025 TIGER/Line Los Angeles County dataset (U.S. Census Bureau, 2025). Missing widths are estimated normal to the centerline from the impervious raster and capped at 30 m.

A.4 WIND DATA

We use hourly ERA5-Land wind components at 10 m above ground on a $0.1^\circ \times 0.1^\circ$ grid, from the dataset accompanying Li et al. (2026), derived from ERA5-Land (Muñoz Sabater et al., 2021). Eastward and northward wind components determine speed and direction, and each building uses the nearest available grid cell. Winds update hourly and determine wall-parallel flame adjustment and downwind eligibility, without a local roughness correction.

A.5 PROGRESSION PROXIES AND CONDITIONING

Hourly fire regions from Kerber et al. (2025) are processed in EPSG:3310, accumulated across reporting windows, and clipped to the final Eaton perimeter (California Department of Forestry and Fire Protection, 2026). Each building receives the earliest window intersecting its saved node geometry. These estimated fire-arrival times are uncertain because of mapping errors, reporting limitations, and hourly grouping. Weather is taken from the starting clock hour of each window, so the window from 18:18 to 19:17 uses 18:00 PST weather.

Damage records (California Department of Forestry and Fire Protection, 2025) supply endpoint targets, progression (Kerber et al., 2025) supplies timing targets, initially ignited buildings damaged in the first hour, and interface-exposure eligibility. Timing agreement therefore does not independently validate wildfire advance. Reference outcomes are excluded from actor inputs and local heat-transfer calculations, but reconstruction remains retrospectively conditioned. Audit traces retain offers, selected pathways, fluxes, dose, and ignition crossings.

B FIRE-EXPOSURE AND STRUCTURE-IGNITION MODEL

This appendix specifies the physical model summarized in Sections 3.1 and 3.2.

The model below maps structural fuel and openings to source heat release rate (HRR), receiver flux, and component ignition. Observation conditioning is described in Appendix A.

Notation and units. Indices i, j, c, and m denote source structure, receiver structure, receiving component, and exposure family. The main text indexes the receiver by i, so its Φ_{ic} corresponds to Φ_{jc} here. In vegetation calculations, p denotes a fuel object, distinct from the main-text pathway tuple. Unless stated otherwise, units are meters, seconds, kilograms, kilowatts (HRR), and $\mathrm{kW\,m^{-2}}$ (flux). Inventory energy is in MJ and heat of combustion in $\mathrm{MJ\,kg^{-1}}$. Source and dimensionless flame calculations require the conversions specified below. Heat exposure and flux–time ignition accumulation are distinct.

The model forms a chain from each burning source to each receiving component. Structural fuel inventory (Eq. 6) and opening geometry determine a ventilation-supported reference HRR (Eq. 8). A fixed 60-minute profile distributes released energy over time (Eq. 10), and the total is split into a ventilation-supported part, which drives whole-building radiation (Eq. 13), and an excess divided among open windows, which drives opening flames (Eqs. 16–19). Flame reach also determines when nearby trees and shrubs are contacted, triggering vegetation radiation pulses (Eqs. 20–25). The policy selects which offered contributions to receiver j enter the summed flux q''_j, which accumulates dose in each accessible component (Eq. 27). A building ignites thermally when any eligible component reaches its threshold (Eq. 28), and wildland-interface and firebrand-proxy entry provide separate routes (Appendix B.6). Either way, the ignited building becomes a source, which closes the loop. Table 2 maps each model component to where it is specified.

Modeling conventions. Citations identify the source relationships, not validation at full WUI scale or for individual study buildings. Distances, durations, fractions, and thresholds are modeling choices.

B.1 STRUCTURES AND DIRECTED NEIGHBORHOOD

Building data (Appendix A) determine source fuel, receiver vulnerability, and policy features. Component mappings and encoder inputs appear in Appendices B.7 and C. The physical neighborhood is the directed graph

$$\mathcal{G} = (\mathbb{B}, \mathbb{L}) \qquad (i,j) \in \mathbb{L} \iff i \neq j \text{ and } d_{\mathrm{fp}}(i,j) \leq 30\ \mathrm{m} \tag{5}$$

Here, $\mathbb{B}$ contains structure identifiers, $\mathbb{L}$ contains candidate directed connections, and d_{fp} is minimum footprint separation. The 30 m radius is motivated by home-ignition-zone research (Cohen, 2010), which does not establish a universal exposure cutoff. Both directions are admitted, with potentially different source, receiver, wind, vegetation, obstruction, and road attributes. Edges indicate candidate connections, not active exposure or ignition.

Table 2: Physical-model provenance. Detailed values, equations, and source citations appear in the referenced sections.

Component	Basis	Details
Fuel and ventilation	Published inventories and inflow relation, with assumed count and geometry	Sections B.2, B.7
Air–fuel mixture	Derived from constituent ratios and inventory fractions	Section B.7
Source history	Fixed profile and energy normalization	Section B.3
Building radiation	Point-source relation, with assumed geometry and obstruction	Section B.4
Opening flames	Adapted correlations and transfer curves	Sections B.4, B.7
Vegetation	EcoSmart and USDA inputs, with assumed activation and shrub response	Sections B.4, B.7
Component ignition	Wood correlation and FSRI fits, with assumed accessibility and dose retention	Section B.5
Entry and integration	Fixed eligibility, acceptance, and integration rules	Section B.6

B.2 Structural fuel and heat release rate

Following Ding et al. (2026), we estimate combustible mass and energy from geometry and construction categories using their wall and roof inventories, representative heights, and timber-truss relationship. From Ding et al. (2026), single- and multi-story building heights are 3.81 and 6.55 m, and sloped roof area is $1.414 L_i W_i$ for represented length L_i and width W_i. Appendix B.7 gives material, mass, and opening assumptions. Summed wall, roof, and truss inventories give energy $Q_{\mathrm{SFL},i}$ (MJ), mass $M_{\mathrm{SFL},i}$ (kg), and effective heat of combustion,

$$H_{c,i} = Q_{\mathrm{SFL},i} / M_{\mathrm{SFL},i} \tag{6}$$

where HRR additionally depends on the ventilation and source-profile assumptions below.

For modeled openings o with area A_{io} and height H_{io},

$$\dot{m}_{\mathrm{air},i} = 0.50 \sum_o A_{io} \sqrt{H_{io}} \qquad \dot{m}_{\mathrm{fuel},i} = \dot{m}_{\mathrm{air},i} / \mathrm{AFR}_{\mathrm{eff}} \tag{7}$$

where $\mathrm{AFR}_{\mathrm{eff}} \approx 5$, which is calculated later in Appendix B.7. The average ventilation-supported reference heat release rate (HRR) is

$$\overline{\dot{Q}}_{\mathrm{vent},i} = 1000 \eta_i H_{c,i} \dot{m}_{\mathrm{fuel},i} \tag{8}$$

Each building is assigned a combustion efficiency $\eta_i \sim U(0.80, 0.90)$, seeded by its identifier so the sampled value is reproducible. The factor 1000 converts MW to kW.

Doors contribute to ventilation but not opening flames. The inflow coefficient follows $0.5A\sqrt{H}$ in Lee et al. (2007, p. 2523, Section 3), and their internal HRR of $1500A\sqrt{H}$ kW assumes 3000 kJ per kilogram of consumed air. Equation 8 instead uses the modeled inventory, efficiency, and air–fuel ratio. Likewise, where Lee et al. (2009, p. 2556, Eq. (10)) subtract $1500A\sqrt{H}$ from total HRR, we subtract the modeled ventilation-supported estimate and clip at zero.

B.3 Time-varying source fire

The quadratic-growth form has precedent in standard t^2 design fires (Quintiere, 2006). Each building source follows a fixed 60-minute profile ($T = 3600$ s), 20 minutes of quadratic growth ($T_g = 1200$ s), a 34-minute plateau, and 6 minutes of linear decay ($T_d = 360$ s). For source age a in seconds, we define

$$g(a) = \begin{cases} (a/T_g)^2 & 0 \le a < T_g \\ 1 & T_g \le a < T - T_d \\ (T-a)/T_d & T - T_d \le a < T \\ 0 & \text{otherwise} \end{cases} \qquad \Gamma = \int_0^T g(a)\,\mathrm{d}a = T - 2T_g/3 - T_d/2 = 2620 \text{ s} \tag{9}$$

The released-energy budget and instantaneous total HRR are

$$E_i = f_E \eta_i Q_{\mathrm{SFL},i} \qquad f_E \approx 0.64 \qquad \dot{Q}_{\mathrm{total},i}(a) = \frac{E_i}{\Gamma} g(a) \tag{10}$$

Here E_i is in MJ and HRR in MW, giving $\int_0^T \dot{Q}_{\mathrm{total},i}(a)\,\mathrm{d}a = E_i$ and peak E_i/Γ. The fixed factor $f_E \approx 0.64$ scales the energy released over the 60-minute source duration.

Let $Q_{\mathrm{cap},i}$ denote the ventilation-supported reference HRR from Equation 8, converted to MW. We use this reference to allocate the time-varying source HRR between two modeled contributions,

$$\dot{Q}_{\mathrm{vent},i}(a) = \min\Big\{\dot{Q}_{\mathrm{total},i}(a), Q_{\mathrm{cap},i}\Big\} \tag{11}$$

$$\dot{Q}_{o,i}(a) = \frac{\Big[\dot{Q}_{\mathrm{total},i}(a) - Q_{\mathrm{cap},i}\Big]_+}{N_{\mathrm{open},i}} \tag{12}$$

where $[x]_+ = \max(x, 0)$ and $N_{\mathrm{open},i}$ counts modeled open windows. Buildings without open windows have no opening flame contribution. The ventilation-supported HRR supplies building radiation (converted to kW in Eq. 13). Positive excess is divided equally among open windows to recompute flame geometry and transfer before applying the edge-flux cap. Per-opening HRR is converted from MW to kW before entering the flame correlations below. Both building contributions follow source age, whereas vegetation uses separate contact-triggered pulses.

B.4 Local exposure pathways

We model whole-building radiation, opening-flame, tree, and shrub exposure. Their active selected contributions enter the combined-component ignition calculation in Appendix B.5.

Whole-building radiation. Thirty percent of instantaneous ventilation-supported HRR is assigned to an isotropic point source. The inverse-square form adapts the fire-radiation point-source approximation in Purnomo et al. (2024).

$$q''_{ij,\mathrm{rad}}(t) = B_{ij} \frac{0.30 \dot{Q}_{\mathrm{vent},i}(t)}{4\pi \max(r_{ij}, 1\ \mathrm{m})^2} \tag{13}$$

Here r_{ij} is centroid distance and B_{ij} is zero for an intervening two-dimensional footprint and one otherwise. The radiant fraction 0.30, centroid representation, 1 m floor, and binary obstruction are assumed adaptations of the point-source relation. Opening flames are modeled separately.

External facade flame. Opening flame dimensions use the correlations of Lee et al. (2007) and Hu et al. (2019). Let A, W, and H denote opening area, width, and height, respectively, with lengths in meters. Define the characteristic lengths $\ell_1 = (A\sqrt{H})^{2/5}$ and $\ell_2 = (AH^2)^{1/4}$. For per-opening HRR $\dot{Q}_o$ expressed in kW, the dimensionless HRR is

$$Q^* = \frac{1000 \dot{Q}_o}{\rho_\infty c_p T_\infty \sqrt{g}\, \ell_1^{5/2}} \tag{14}$$

The zero-wind flame height is

$$Z_{f,0} = 3.3 \ell_1 (Q^*)^\beta \tag{15}$$

where $\beta = 2/3$ for $Q^* \le 1$ and $\beta = 2/5$ otherwise. We fix ambient air density $\rho_\infty = 1.204\,\mathrm{kg\,m^{-3}}$, specific heat $c_p = 1005\,\mathrm{J\,kg^{-1}\,K^{-1}}$, temperature $T_\infty = 293.15\,\mathrm{K}$, and gravitational acceleration $g = 9.8\,\mathrm{m\,s^{-2}}$. The factor 1000 converts HRR from kW to W.

For wall-parallel wind speed U_w, define $Fr_w = U_w/\sqrt{gZ_{f,0}}$ and $\chi = (Z_{f,0}/\ell_2)^{0.8} Fr_w^{1.2}$. The wind-adjusted flame height is

$$Z_{f,w} = 0.6H + \max(0, Z_{f,0} - 0.6H) \exp\Big(-\frac{\chi}{3.4}\Big) \tag{16}$$

where the height attenuation follows Hu et al. (2019). With $\zeta = (Z_{f,0}/\ell_2) Fr_w^{1.12}$, the wind-adjusted flame width is

$$W_{f,w} = W_{f,0} \frac{Z_{f,w}}{Z_{f,0}} (1 + \zeta)^{1.32} \tag{17}$$

We estimate zero-wind flame width as $W_{f,0} = 0.370167 W^{0.205969} H^{0.356300}$, with $W_{f,0}$ in meters. Application at residential scales extrapolates beyond those data.

Opposite-wall flux adapts Lee et al. (2009). Let

$$\ell_3 = (AH^{4/3})^{3/10} \qquad s = \begin{cases} d, & d \le \ell_3 \\ \ell_3, & d > \ell_3 \end{cases} \tag{18}$$

where d is footprint separation. The implemented peak flux is

$$q''_{ij,\mathrm{Lee}} = f(Z/Z_{f,w}) \frac{\dot{Q}_o}{s Z_{f,w}} \exp[-0.6(\ell_3/\ell_1)^{5.5}] \tag{19}$$

where we assume equal ground elevations and use the wind-adjusted flame height $Z_{f,w}$. Here f interpolates the near- and far-wall curves of Lee et al. (2009, Figures 10–11) in log–log space. We extend the pathway to 10 m footprint separation and cap instantaneous recomputed flux at $100\ \mathrm{kW\,m^{-2}}$, retaining uncapped values and experimental-domain flags.

Tree-mediated radiation. For crown area A_p, dry foliage mass is $M_{\mathrm{dry},p} = A_p L_f$, using regional median loading $L_f = 0.6685\ \mathrm{kg\,m^{-2}}$ from the USDA Urban Tree Database (McPherson et al., 2016a). We assume dry-basis moisture content $MC = 20\%$, giving wet mass $M_{\mathrm{wet},p} = 1.2 M_{\mathrm{dry},p}$. The EcoSmart Fire general-tree branch (Dietenberger & Boardman, 2017) gives

$$\dot{Q}_p = \frac{2 M_{\mathrm{wet},p}(700)}{1 + 0.1295 MC} \tag{20}$$

where masses are in kg and $\dot{Q}_p$ is in kW.

Crowns are subdivided into emitters below the reported 1000 kW validation ceiling while conserving area, mass, and HRR. For emitter diameter D_p and HRR $\dot{Q}_p$, we compute flame height H_p, temperature $T_{f,p}$, and emissivity ε_p using the EcoSmart Fire correlations. The temperature calculation clips diameter to $D_p^* = \min(7.9, \max(0.019, D_p))$, with diameter in meters. Radiant flux from emitter p to receiver j is

$$q''_{pj,\mathrm{tree}} = \frac{\varepsilon_p \sigma_{\mathrm{SB}} T_{f,p}^4}{1000} F_{pj} \tag{21}$$

where σ_{SB} is the Stefan–Boltzmann constant and F_{pj} is the flame-to-receiver view factor. The factor 1000 converts flux from $\mathrm{W\,m^{-2}}$ to $\mathrm{kW\,m^{-2}}$.

For a differential receiver facing the cylindrical flame, the cumulative view factor to flame extent $z > 0$ is

$$F^+(z) = \frac{R_p}{2\pi L_{pj}} [\pi - 2\theta(z) + \sin(2\theta(z))] \tag{22}$$

where $\theta(z) = \tan^{-1}(L_{pj}/z)$, $R_p = D_p/2$, and L_{pj} is the horizontal distance from the receiver to the flame centerline. This is the facing-wall specialization of Dietenberger & Boardman (2017, Equations (4) and (7), pp. 584–585). Defining $G(z) = \mathrm{sgn}(z) F^+(|z|)$ for $z \neq 0$ and $G(0) = 0$, the finite flame-segment factor is

$$F_{pj} = |G(z_{\mathrm{top},p} - z_j) - G(z_{\mathrm{base},p} - z_j)| \tag{23}$$

where $z_{\mathrm{top},p}$ and $z_{\mathrm{base},p}$ are flame-top and flame-base heights, and z_j is receiver height.

We evaluate facing receivers at three facade heights, taking each patch's maximum before summing concurrent patches. This is a scalar approximation, not the combined flux at one height. Unresolved near-field emitters are excluded, so the subtotal is not a guaranteed lower bound on actual exposure.

A tree pathway $i \to p \to j$ requires

$$d_{\mathrm{fp}\to p}(i,p) \le W_{f,w,i} \tag{24}$$

where $d_{\mathrm{fp}\to p}(i,p)$ is the distance from building i's footprint to patch p, and $W_{f,w,i}$ is its maximum wall-specific flame extension. This reach is applied around the footprint and recomputed from the time-varying building source. A patch activates at first modeled flame contact, independently of policy selection, and supplies constant precomputed flux during a 30 s pulse. Tree and shrub patches both activate automatically on modeled flame contact, and the policy selects which available vegetation-to-building pathways deliver heat to receivers. Therefore, activation alone does not imply that a patch's radiation enters a building's accumulated ignition dose.

Table 3: Reference component ignition parameters. FSRI fits use nine observations at three tested fluxes and are provisional (McKinnon et al., 2023).

Proxy	q''_{crit} (kW m^{-2})	n	FTP	Use
Wood proxy	13.100	1.828	11501.0	roof, eave, interior, other siding
Asphalt shingle	7.881	1.314	3661.4	asphalt roof
Pine siding	6.303	2.130	34039.3	wood siding
Vinyl siding	16.395	1.234	2563.4	vinyl siding

Shrub-mediated radiation. Shrub cover is identified from the same land-cover map. The same no-extra-buffer flame-reach rule as Eq. 24 determines eligibility. For shrub-to-receiver distance d_{pj},

$$q''_{pj,\text{shrub}} = 100 \left(\frac{1\ \text{m}}{\max(d_{pj}, 1\ \text{m})} \right)^2 \text{kW m}^{-2} \tag{25}$$

applied for 30 s after first flame contact, using the maximum concurrent patch flux summed over selected sources. The reference flux and attenuation do not come from an EcoSmart shrub correlation.

B.5 Construction-dependent ignition accumulation

Construction and hardening features determine accessible components $c \in \mathbb{K}_j$ and map them to

$$g_c(\boldsymbol{x}_j) = \left(q''_{\text{crit},jc}, n_{jc}, \text{FTP}_{jc}, \kappa_{jc}\right) \tag{26}$$

where q''_{crit} is critical incident flux, n is the exponent, FTP is the threshold, and $\kappa_{jc} \in \{0, 1\}$ records whether a pilot is required. A component is a modeled combustible target, including a material surface or an interior-wood proxy, rather than a separate structure node. Let $q''_j(t) = \sum_m q''_{jm}(t)$ be the sum of concurrent selected local thermal contributions at the effective receiver. Let $\mathbb{K}_j^{\text{elig}}(t) \subseteq \mathbb{K}_j$ contain accessible components whose pilot requirements are satisfied. For eligible components, the model updates one accumulated dose I_{jc} per component.

$$I_{jc}(t + \Delta t) = \min\left\{\text{FTP}_{jc},\ I_{jc}(t) + [\max(\bar{q}''_j - q''_{\text{crit},jc}, 0)]^{n_{jc}} \Delta t\right\} \tag{27}$$

Here $\bar{q}''_j$ is the combined flux held at its interval value (Appendix B.6). The progress in Eq. 1 is $\Phi_{jc} = I_{jc}/\text{FTP}_{jc}$, starting at zero, and the model includes no cooling or loss of accumulated dose.

A positive flux from any pilot-capable family supplies the shared pilot condition. Components requiring a pilot receive no increment when that condition is absent. The complete component inventory is determined by construction and accessibility mappings. This is not a resolved pilot-transport model. Thermal ignition occurs at

$$t^{\text{ign}}_{j,\text{thermal}} = \inf\{t : \exists c \in \mathbb{K}_j^{\text{elig}}(t), I_{jc}(t) \geq \text{FTP}_{jc}\} \tag{28}$$

For eligible component c at constant combined above-critical flux and prior dose I_0, the remaining time is

$$t_{\text{ig}} = \frac{\max(\text{FTP}_{jc} - I_0, 0)}{(q''_j - q''_{\text{crit},jc})^{n_{jc}}} \tag{29}$$

At or below critical flux, no new dose accumulates, and a building ignites at the earliest component crossing. Concurrent families contribute to the same effective component flux, a scalar approximation, so a combined ignition record does not identify a unique causal pathway. Table 3 lists the reference ignition parameters for each component.

Wood parameters use the Cohen Douglas-fir proxy (Cohen & Butler, 1998, Eq. (2), p. 84), not a whole-building threshold. Combustible exterior and exposed-eave routes require a pilot supplied by opening-flame, tree, or shrub exposure. Building radiation alone cannot supply it. Every present window is assumed failed, exposing interior wood without a pilot requirement. This override permits radiation-only ignition but is not a measured unpiloted threshold. No glass-failure timing, flux transmission, or vent or ember penetration is modeled. "No Windows" removes the window route.

B.6 CHRONOLOGY AND EXTERNAL ENTRY

Wind and estimated progression update hourly (Appendix A). Selected exposure is integrated in 60-second windows, subdivided at source-profile and vegetation-pulse boundaries. Nonvegetation flux uses interval midpoints and vegetation pulses have constant flux. Threshold crossings are resolved within subintervals, and newly ignited sources enter the next window. Structures ignite once and remain active for their modeled duration.

Local offers require an active source, eligible receiver, directed edge, and positive pathway flux. Thermal ignition follows Eq. 28. Declining additions retains earlier selections. Appendix C specifies features, masks, and persistence.

Wildland-interface entry uses hourly progression, land cover, wind, distance, woody contact, and receiver vulnerability, without observed damage, to identify exposure at the wildland–urban boundary. Ignition requires 300 seconds of continuous selected interface exposure.

Firebrand-proxy entry parameterizes ember ignition without explicit ember transport or penetration. Eligibility requires 900 seconds of accumulated positive selected thermal exposure, an active upwind source within 100 m, and no current interface eligibility. It may coexist with local offers without requiring their failure. A selected firebrand-proxy entry attempt ignites a still-unignited receiver at the next update-window endpoint, with no additional random acceptance step.

These gates are assumptions, not calibrated physical thresholds. Thermal events are processed before external events, which add neither heat flux nor component dose. Either ignition route activates the building as a potential source.

B.7 PARAMETER DERIVATIONS AND IMPLEMENTATION DETAILS

Assembly and opening values. The wall energy values are 153.4 (wood siding), 27.9 (other siding, represented by an EPS layer), and 65.7 (studs), in $\mathrm{MJ\,m^{-2}}$. Roof values are 340.7 (wood) and 288.3 (other). These follow Tables 3–4 of Ding et al. (2026). The implemented areal masses are 11.287 and 0.712 for siding, $65.7/12.3$ for studs, and 18.937 and 14.875 for roof assemblies, in $\mathrm{kg\,m^{-2}}$. These are project calculations from layer data, not separately measured building properties. Timber density is $355\ \mathrm{kg\,m^{-3}}$ and heat of combustion is $12.3\ \mathrm{MJ\,kg^{-1}}$. Residential and utility window-to-wall ratios are taken to be 0.15, and commercial is 0.33. Window heights are 1.2192 and 1.3076 m for residential and commercial classes, respectively. Four representative windows of width 0.9144 m are used, with total window area limited to the smaller of their combined area and the window-to-wall-ratio area. Doors are 0.9144×2.1336 m, with two for residential and commercial structures and one for utility structures. Modeled openings are fully open. Records classified as having no windows receive zero window ventilation. The residential ratio and height are adopted from the four-story mid-rise multifamily prototype in Huang & Gowri (2011, Appendix A, Table A-1, p. 10). The commercial values come from its medium-office prototype (Table A-2, p. 12). Applying them to broader DINS classes is a project approximation. Window count, width, placement, opening state, and transfer to utility structures are project assumptions. Multi-story records are assigned two stories, and institutional and mixed-use records without a story count are assigned one.

Receiver material mapping. Wood siding uses the pine fit and vinyl siding the vinyl fit, while stucco, brick, cement, and metal siding have no combustible-surface FTP route. Other or missing siding uses the wood proxy. Wood or combustible roofs use the wood proxy, asphalt roofs use the asphalt fit, and tile, metal, concrete, or noncombustible roofs have no outer-surface FTP route. Other or missing roofs use the wood proxy. Enclosed or absent eaves close the eave route, while other or missing eaves expose a provisional wood and cavity proxy. All present windows, including multipane, are assumed failed with full flux reaching interior wood, and "No Windows" removes that route. Vents create no physical pathway in this model.

Air–fuel mixture. The implemented calculation is

$$\mathrm{AFR_{eff}} = \frac{81}{86}(4.92) + \frac{5}{86}\left[0.8871(6.34) + 0.1129(1/0.076)\right] \approx 5 \tag{30}$$

The inventory's 81% wood and 5% plastic fractions (Chou et al., 2025, Section 4) are renormalized after excluding other categories. Wood uses the sugar-pine air-to-volatile-fuel ratio 4.92 (Quintiere

& McCaffrey, 1980, Table 6-2, p. 90) as a bulk-mass proxy. Polystyrene uses fuel-to-air ratio 0.076 (Lock et al., 2008, Table 6.2, p. 125), and polyurethane uses the stoichiometric ratio 6.34 from rigid-polyurethane crib tests (Nelson & Deal, 1991, p. 725), distinct from their experimental 9.03 and best-fit 5.8. The plastic split normalizes Table 1 spray-polyurethane and polystyrene-board masses as $1273/(1273+162)$ and $162/(1273+162)$, rounded to four decimals. These transfers are project approximations. The ventilation coefficient 0.50 has units $\mathrm{kg\,s^{-1}m^{-5/2}}$.

FSRI observations and estimation. The asphalt-shingle, pine-siding, and vinyl-siding material pages of the FSRI database supply three ignition-time observations at each of 25, 50, and 75 $\mathrm{kW\,m^{-2}}$ (McKinnon et al., 2023). The fit minimizes the sum of squared residuals $\log t - \{\log \mathrm{FTP} - n\log(q'' - q''_{\mathrm{crit}})\}$. Bounds are $q''_{\mathrm{crit}} \in [0, 24.99]$, $n \in [0.1, 5]$, and $\log \mathrm{FTP} \in [-20, 30]$. Initialization is $(8, 1.5, \log 5000)$, with at most 10000 function evaluations. The reproduced log-time RMSEs are 0.03322 (asphalt), 0.08969 (pine), and 0.08781 (vinyl). Three distinct flux levels constrain three parameters, so low residuals do not establish predictive accuracy. Fluxes below the lowest tested level are extrapolations.

Foliar loading. For each usable USDA tree record, dry leaf mass is leaf area times the matched species–region dry mass per leaf area, divided by 1000 to convert grams to kilograms. Crown area is $\pi D_{\parallel} D_{\perp}/4$, with average-diameter circular fallback. The median ratio of dry leaf mass to crown area gives $0.6684682\ \mathrm{kg\,m^{-2}}$ for the Inland Empire (740 records) and $0.3725648\ \mathrm{kg\,m^{-2}}$ for Southern California Coast (543 records). The former is the Eaton value. These calculations were reproduced from USDA tables TS3 and TS5 (McPherson et al., 2016a;b). Matching, positive-value filtering, and regional aggregation are project processing choices.

Opening flame implementation. Height ratios are clamped to the fitted curve limits, and effective gaps replace shared-footprint zero gaps. Building footprints screen both opening flame and whole-building radiation.

Tree view factor. Equation 22 specifies the view factor and its finite-segment application. The three receiver heights are $\min(0.5, h_j/2)$, $h_j/2$, and $\max(h_j - 0.5, h_j/2)$. This project approximation excludes unresolved near-field emitters and does not reproduce every EcoSmart shadowing or ground-reflection treatment. Candidate vegetation associations use a 5 m corridor buffer, distinct from the no-extra-buffer ignition-reach rule. The Stefan–Boltzmann coefficient is $5.6696 \times 10^{-8}\ \mathrm{W\,m^{-2}K^{-4}}$.

C Policy and environment implementation

This appendix details the decision process in Section 3.3.

C.1 State and policy observations

The environment retains fire states, component dose, source and vegetation histories, external-entry clocks, selections, and scoring records. Decision k occurs at physical time t_k, and multiple choices may precede time advance. The compressed policy observation o_k, including masks, is not assumed Markov-sufficient. Pathway features include source and receiver embeddings, mechanism, flux, distance, wind alignment, road crossing, uncertainty, dose, critical flux, and vulnerability. External offers have zero source embeddings. Context pooling is defined below, and conditioning appears in Appendix A.

C.2 Encoder and policy architecture

A two-layer mean-aggregation GraphSAGE encoder (Hamilton et al., 2017) maps standardized attributes to 32-dimensional building embeddings. Link-prediction pretraining precedes PPO, during which the encoder is frozen. Fire state enters through context and pathway features.

Static inputs and preprocessing. The 26 inputs comprise eight geometry attributes (height, area, perimeter, length, width, aspect ratio, compactness, orientation), year built and occupancy units,

and geometry-imputation and shared-footprint flags. They also include six resilience indicators (roof, siding, window, eave, vent, fence), an absent-window flag, three combustible-pathway indicators (deck, patio or carport, fence), core-four and system-six conservative resilience fractions, a known-weak-link flag, and nearest-structure distance. Booleans are numeric, invalid or missing values become zero, and columns are standardized using full-Eaton means and population standard deviations, replacing deviations below 10^{-6} with one.

Pretraining and execution. The original graph has 18,384 nodes and 161,846 directed edges within 30 m. Incoming-neighbor means enter learned self and neighbor transformations with ReLU between layers. Ten pretraining epochs use graph edges and random node pairs for link prediction. Exact-local evaluation computes required inbound neighborhoods without sampling. Interface and firebrand-proxy entry opportunities extend beyond this embedding graph.

Context and heads. Context concatenates embedding means and maxima over the computation subgraph, burning buildings, and frontier receivers, plus population and time scalars. The computation pool is not community-wide. Receiver and pathway heads combine embeddings, physical features, and context, whereas the critic receives context alone. Pool membership and pathway features vary with simulation state.

C.3 Initialization and action execution

The environment initializes the 71 initially ignited buildings with source clocks at zero and an ignition hour of zero that marks initialization. Their ignition is taken from the record rather than obtained by accumulating component dose, and all other buildings start with zero accumulated dose.

Selection-and-commit procedure. A pathway links a receiver to an exposure mechanism and, for local exposure, to a source. An offer is a pathway currently available for selection, and a building is a candidate when it has at least one offer. Candidate availability measures whether an observed burned building ever becomes a candidate during a rollout. Selection uses a receiver distribution followed by a receiver-conditional pathway distribution. Local offers require positive exposure from an active building or vegetation pulse. Opening flames also require separation within 10 m. Within each hour, selection proceeds as follows.

1. Before selecting any new pathway, the policy may decline new exposure.
2. Once it selects a pathway, it remains on that receiver and can either add another or commit the selection. Already selected pathways are masked out, and selecting all available offers triggers an automatic commit.
3. Committed contributions persist within the hour while available. New options can be added, loss of a selected source allows reconsideration, and declining an addition leaves retained contributions intact without guaranteeing survival.
4. The hour boundary resets the selection plan while leaving accumulated dose unchanged. Once the current receivers are resolved, the environment advances and prepares the next offers.

Action masks (Huang & Ontañón, 2022) enforce the modeled eligibility rules directly.

External-entry eligibility, acceptance rules, and numerical update timescales are specified in Appendix B.6.

D Training, selection, and evaluation details

This appendix details the protocol summarized in Section 4.2.

D.1 Reconstruction score

This subsection gives the full scoring terms of Eq. 3.

Let y_{ih} and $\widehat{y}_{ih}$ denote the observed-proxy and simulated cumulative ignition states of building i by hour h. For mismatch probability ϵ, define

$$\ell(\widehat{y}, y; \epsilon) = \begin{cases} \log(1-\epsilon) & \widehat{y} = y \\ \log \epsilon & \widehat{y} \neq y \end{cases} \tag{31}$$

For a scored population S, let $S_c = \{i \in S : y_i = c\}$ and let C contain only the observed classes with nonempty S_c. The class-balanced, boundary-weighted average is

$$\mathcal{B}_S(\ell, y) = \frac{1}{|C|} \sum_{c \in C} \frac{\sum_{i \in S_c} w_i \ell_i}{\sum_{i \in S_c} w_i} \qquad w_i = \begin{cases} 2 & i \text{ is a boundary building} \\ 1 & \text{otherwise} \end{cases} \tag{32}$$

Boundary buildings have centroids within 30 m of a building with the opposite known, nonconflicting outcome. The reference population includes initially ignited buildings. The scored population excludes them. This is an outcome-derived boundary, not an official fire perimeter.

For a positive building with assigned observation hour τ_i, the hourly mismatch probability is

$$\epsilon_{ih} = \begin{cases} 0.4 & \tau_i - 1 \leq h < \tau_i + 1 \\ 0.05 & \text{otherwise} \end{cases} \tag{33}$$

Negative buildings use 0.05 throughout. The half-open interval describes uncertainty in cumulative state rather than a symmetric inclusive window over hourly states. The separate timing-recall metric evaluates whether a predicted ignition is within one hour of its observation proxy.

Let S_{time} contain buildings with known hourly states and S_{out} contain those with known final outcomes y_i. Then

$$F_h = \mathcal{B}_{S_{\text{time}}}\big(\ell(\widehat{y}_{ih}, y_{ih}; \epsilon_{ih}), y_{ih}\big) \tag{34}$$

$$O = \mathcal{B}_{S_{\text{out}}}\big(\ell(\widehat{y}_{iN_h}, y_i; 0.01), y_i\big) \tag{35}$$

$$R = \frac{1}{N_h} \sum_{h=1}^{N_h} F_h + 2O \qquad N_h = 15 \tag{36}$$

Missing timing labels contribute only to O. Initially ignited buildings and buildings with unknown or conflicting outcomes are excluded from scoring, whereas buildings never offered to the policy remain scored. Probabilities and weights are design choices, not calibrated measurement errors. The reward pays F_h/N_h at hour closure and $2O$ at the endpoint, so undiscounted increments sum to R. Selection-only actions receive no separate reward.

D.2 Policy optimization

Our graph-policy design follows the practical guidance of Schutz & Darvariu (2026), and the reported experiments use a custom PyTorch training loop and simulation interface. PPO (Schulman et al., 2017) uses the factorized probability of Eq. 2, $\pi_{\boldsymbol{\theta}}(a_k \mid o_k) = \pi^{\text{rec}}_{\boldsymbol{\theta}}(i_k \mid o_k)\pi^{\text{path}}_{\boldsymbol{\theta}}(b_k \mid i_k, o_k)$. Here $\boldsymbol{\theta}$ and ψ denote actor and critic parameters, o_k is the policy observation, and r_k is the transition reward. Let $\Delta_k = (t_{k+1} - t_k)/60$ be elapsed simulated minutes and d_k indicate episode termination. The implemented physical-time GAE (Schulman et al., 2015) is

$$\delta_k = r_k + \gamma^{\Delta_k}(1 - d_k)V_\psi(o_{k+1}) - V_\psi(o_k) \tag{37}$$

$$\widehat{A}_k = \delta_k + (\gamma\lambda)^{\Delta_k}(1 - d_k)\widehat{A}_{k+1} \tag{38}$$

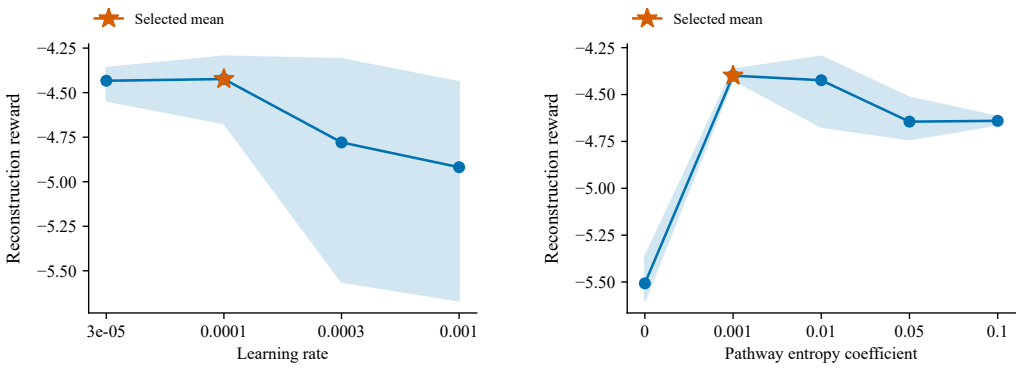
$$\widehat{R}_k = \widehat{A}_k + V_\psi(o_k) \tag{39}$$

The advantage recursion starts at zero at each buffer end. Nonterminal boundaries bootstrap from the next-state value, and termination suppresses bootstrapping and cross-episode advantages. Zero-time choices have unit discount factors. With $\gamma = 0.9999$ per minute, $\gamma^{900} \simeq 0.914$, so training differs from maximizing the undiscounted reported score.

With $\rho_k = \pi_{\boldsymbol{\theta}}(a_k \mid o_k)/\pi_{\boldsymbol{\theta}_{\text{old}}}(a_k \mid o_k)$ and normalized advantages $\widetilde{A}_k$, the policy maximizes

$$L_{\text{clip}} = \mathbb{E}_k\Big[\min\Big(\rho_k \widetilde{A}_k, \text{clip}(\rho_k, 1 - \epsilon_{\text{clip}}, 1 + \epsilon_{\text{clip}})\widetilde{A}_k\Big)\Big]. \tag{40}$$

Training also fits the critic to $\widehat{R}_k$ and adds separate receiver and conditional-pathway entropy bonuses.



(a) Learning rate

(b) Pathway entropy coefficient

Figure 3: Sequential hyperparameter screening at 70k actions. Lines show mean full-fire score across three training seeds, evaluated with seed 2001. Learning-rate screening fixes pathway entropy at 0.01, and entropy screening fixes learning rate at 10^{-4}. Stars mark selected settings. The condition shared by both sweeps is fitted once.

Original training settings. Table 4 lists the original training settings. Advantages are normalized across collected records. The separate $\lambda = 1$ experiment changes advantage estimation and is not the original configuration.

Table 4: Original WUIRL training settings.

Setting	Value
Optimizer	AdamW (Loshchilov & Hutter, 2019)
Learning rate	10^{-4}
Weight decay	10^{-4}
Actions per rollout	2,048
PPO epochs per update	2
Optimization batch size	2,048 actions (full rollout)
Maximum gradient norm	0.5 (global ℓ_2 norm)
Clipping range ϵ_{clip}	0.2
Value loss coefficient	0.10
Receiver entropy coefficient	0.002
Pathway entropy coefficient	0.001
Discount γ (per simulated minute)	0.9999
GAE λ (per simulated minute)	0.95
Total actions per training seed	500,000
Checkpoint interval (nominal)	50,000 actions
Training seeds	42, 137, 2027
Selection seeds	2001, 4001, 6007

D.3 Screening and training

Sequential screening used training seeds 42, 137, and 2027, 70,000 actions per fit, and evaluation seed 2001. Learning rates $\{3 \times 10^{-5}, 10^{-4}, 3 \times 10^{-4}, 10^{-3}\}$ were screened at pathway entropy 0.01, followed by entropy coefficients $\{0, 0.001, 0.01, 0.05, 0.1\}$ at the selected rate. The condition shared by both sweeps was fitted once, giving 24 fits in total. Mean full-fire score selected learning rate 10^{-4} and entropy coefficient 0.001 (Figure 3). Main policies use fresh heads and the shared pretrained encoder (Appendix C).

D.4 Checkpoint selection and reporting

Each training seed (42, 137, 2027) received 500,000 actions, collected in 2,048-action updates, with a shorter final update to reach 500,000 actions. Original checkpoint selection uses nominal 50,000-action intervals, whereas separate diagnostic reruns use a 25,000-action grid. Intermediate checkpoints occur at completed update boundaries. Selection maximizes mean score over seeds 2001, 4001, and 6007, breaking ties toward earlier targets. Selected nominal targets were 50k, 100k, and 150k for training seeds 42, 137, and 2027.

Reporting seeds are 10007, 12007, 14009, 16001, 18013, 20011, 22003, 24001, 26003, and 28001. They are disjoint from selection seeds, not a held-out observation set. The 30 rollouts are nested within three training replicates, so their standard deviation is not a confidence interval across training seeds. When the same checkpoint appears in more than one comparison, its evaluation results are reused.

D.5 Evaluation metrics

Evaluation reports both classes and confusion counts. Exact-hour and within-one-hour recall count missing ignitions as misses among timed positives. Timing MAE conditions on matches between observed and predicted ignitions and reports that denominator. We also report offered-conditional recall, the fraction of observed burned buildings offered at least one pathway at any time during a rollout that eventually ignite, regardless of timing. Prefix summaries (hours 1–2 through 1–15) reuse full rollouts, exclude outcome-only records, and apply the endpoint score only at hour 15. They do not test reduced-observation training.

D.6 Baseline definitions

Random selection samples an unresolved receiver uniformly, then an admissible unselected pathway or applicable decline or commit action uniformly. Greedy selects the receiver with the largest summed positive ratio of remaining flux to critical flux, then its highest-scoring pathway. The denominator floor is $1\,\mathrm{kW\,m^{-2}}$, and interface and firebrand-proxy fallback scores are 10^{-6}. Ties are broken in the existing order. Both baselines use ten reporting seeds, although identical deterministic Greedy runs are not independent fits.

D.7 Mechanism removal and training comparisons

Evaluation-time mechanism removals disable vegetation, firebrand-proxy, direct flame, or building radiation in the three fixed selected policies, with ten paired reporting seeds per policy. They are not retrained ablations.

The thermal-gate diagnostic removes only the 900-second firebrand-proxy entry prerequisite, retaining range and wind criteria, in a single paired evaluation of the frozen seed-42 50k policy with evaluation seed 2001. Results appear in Appendix E.2.

D.8 Computational requirements

Training WUIRL for 500,000 actions took 75.1, 70.8, and 66.3 minutes for seeds 42, 137, and 2027, respectively, using one NVIDIA L4 per worker. These times include initialization and checkpoint saving but exclude evaluation and final archive uploads. The complete training and evaluation workflow, including 180 checkpoint evaluations, took 15.8 hours with two concurrent GPU workers.

E Supplementary reconstruction results

Subsections follow the order of Sections 4.3–4.7, with an additional subsection on separate reward and range experiments after the temporal results.

Plot conventions. Unless noted, lines average the three training-policy means and shaded bands span their minimum and maximum, not confidence intervals. Each policy mean uses ten reporting seeds

Table 5: Supplementary reconstruction performance. Values are means, and ± denotes sample standard deviation across three training-policy means, each based on ten reporting evaluations. Baselines have no training-policy standard deviation.

	Reconstruction score	Burn precision (%)	Burn recall (%)	Burn F1	Within-1-h recall (%)
Random	-4.582	71.6	54.8	0.621	17.9
Greedy	-4.507	65.8	87.4	0.751	21.9
WUIRL, selected	-4.335 ± 0.058	68.8 ± 0.6	78.9 ± 2.3	0.735 ± 0.009	23.3 ± 0.8
WUIRL, final 500k	-5.292 ± 0.607	71.1 ± 0.4	22.2 ± 30.5	0.268 ± 0.317	8.1 ± 9.3
$\lambda = 1$, selected	-4.350 ± 0.076	70.2 ± 0.6	71.3 ± 6.1	0.706 ± 0.029	22.8 ± 1.1

Table 6: Matched-state comparison of sampled WUIRL and Greedy exposure selection. Each pair starts from the same simulated state and compares selections within one 60-second update window. Results use WUIRL training seed 42 at 50,000 actions and evaluation seed 2001. Ignition counts include all newly ignited structures, regardless of observed outcome. These exploratory comparisons demonstrate different exposure selections and immediate physical consequences. They do not establish a delayed reward advantage.

Decision window	**WUIRL pathways selected**	**Greedy pathways selected**	**Receivers where WUIRL declined new exposure**[1]	**WUIRL ignitions in next minute**	**Greedy ignitions in next minute**
Hour 2	394	598	61	39	52
Hour 4	590	935	106	58	85
Hour 6	767	1,211	98	79	141

[1] Receivers with at least one eligible exposure offer where WUIRL selected no pathway.

(Appendix D). Prefix summaries reuse fixed rollouts, whereas interventions rerun frozen policies under modified conditions.

E.1 Main results and training variability

Table 5 extends the main comparison with final-checkpoint and $\lambda = 1$ results. The reporting hierarchy is specified in Appendix D. WUIRL improves the reconstruction objective over Greedy, while Greedy retains higher burn recall and F1. On average, WUIRL makes about 1,002 fewer false ignition predictions but misses 888 additional burned buildings per rollout.

From matched physical states, WUIRL selected 34–37% fewer pathways than Greedy at hours 2, 4, and 6 and produced fewer next-minute ignitions than Greedy at each window. This suggests that the learned policy adopts a more selective exposure strategy than the Greedy baseline. Table 6 reports the underlying counts.

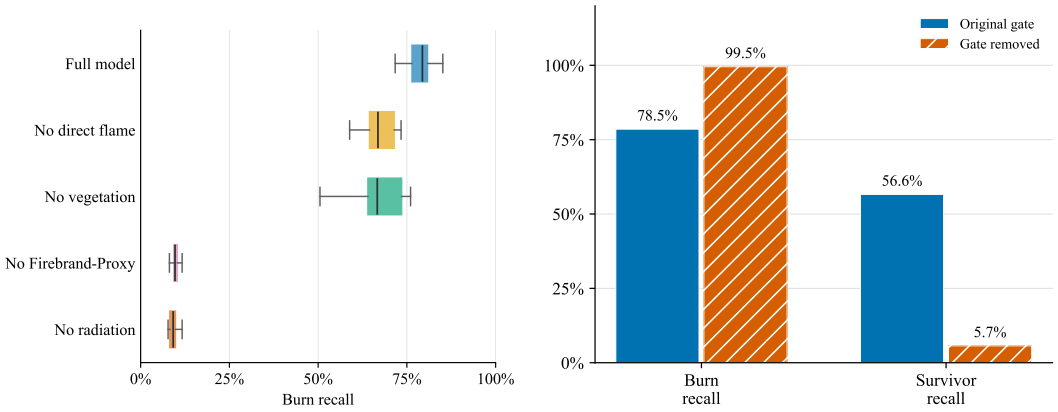
E.2 Mechanism removal and firebrand-proxy

Burn recall depends most on building radiation and firebrand-proxy. Figure 4 reports frozen-policy mechanism removals and the paired thermal-gate test, following the protocol in Appendix D.7.

E.3 Temporal progression and candidate availability

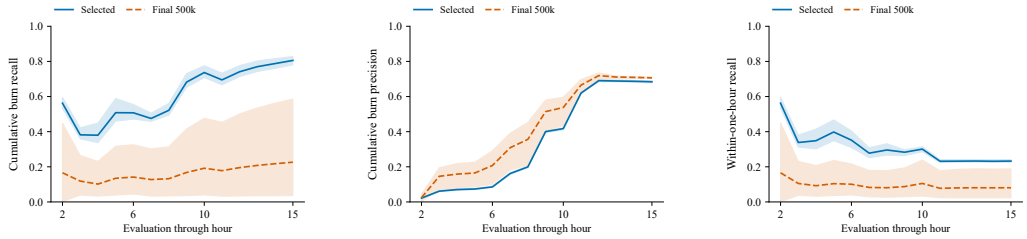
Figure 5 shows cumulative agreement as evaluation extends from hours 1–2 through 1–15 of fixed rollouts. Burn precision is 7.0% through hour 4 and rises to 69.0% through hour 12, as observed ignitions accumulate and fewer predictions precede their observed hour.

For observed ignition cohorts in hours 10–15, eventual recovery averages 76.6%, but within-one-hour recall is 15.0%. Among buildings ever offered to the policy, eventual recovery is 97.9%, and



(a) Mechanism removal (b) Thermal prerequisite removal

Figure 4: Dependence on exposure mechanisms and thermal eligibility. (a) Evaluation-time removals from three frozen selected policies, each evaluated on ten common reporting seeds. Boxes summarize 30 evaluations nested within three policies, ordered by median burn recall. Lines indicate medians, boxes the interquartile range, and whiskers observations within 1.5 interquartile ranges. Outlier markers are omitted. These removals measure model dependence, not historical causal contributions. (b) A single paired evaluation of the seed-42 policy at 50k actions on evaluation seed 2001 removes the 900-second thermal-exposure prerequisite while retaining range and wind criteria. Additional ignition capacity increases burn recall but reduces survivor recall.



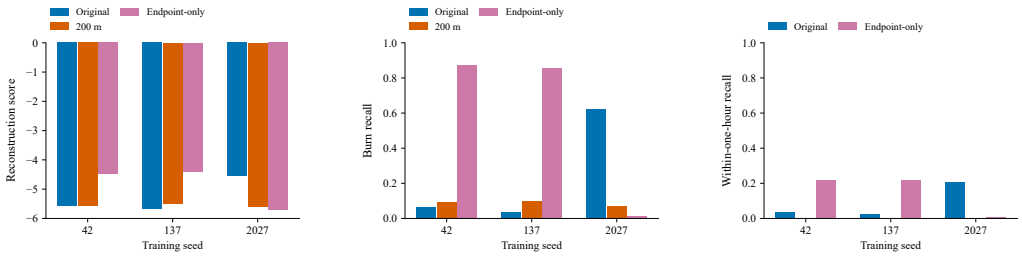
(a) Cumulative burn recall (b) Cumulative burn precision (c) Within-one-hour recall

Figure 5: Cumulative reconstruction agreement as the evaluation includes hours 1–2 through 1–15 of fixed 15-hour rollouts. Lines show means over three training policies, each averaged over ten disjoint reporting evaluation seeds. Selected and final 500k checkpoints use matched reporting seeds. Metrics use the 17,876 timing-known buildings, and observed cumulative ignitions grow from 45 at hour 2 to 9,997 at hour 15. Prefix precision counts predicted ignitions before their observed hour as disagreements at that cutoff, even if the building eventually burns. Within-one-hour recall requires ignition to have occurred by the evaluated cutoff. These metrics have different denominators from full-population endpoint metrics.

92.9% of missed buildings in these cohorts were never offered. Never offered does not imply that no physical pathway exists, and ever offered does not imply availability at the correct time.

E.4 Separate reward and range experiments

Separate experiments remove the temporal reward or extend firebrand-proxy entry range from 100 to 200 m with the original reward. Each experiment retrains seeds 42, 137, and 2027 for 500k actions. Figure 6 compares final checkpoints using the common scorer and selection seeds 2001, 4001, and 6007, rather than selected-checkpoint reporting.



(a) Common reconstruction score (b) Burn recall (c) Within-one-hour recall

Figure 6: Separate reward and range experiments at final 500k checkpoints. Bars show the common reconstruction score, burn recall, and within-one-hour recall, averaged over selection seeds 2001, 4001, and 6007 for each training seed. Timing estimates for the 200 m condition are unavailable and omitted.

Table 7: Spatial holdout performance. WUIRL averages three policy means, each over ten reporting seeds, and each baseline averages ten evaluations.

	Burn recall (%)	Survivor recall (%)	Burn F1
Random	48.1	61.7	0.564
Greedy	81.9	41.5	0.761
WUIRL	55.3	61.2	0.587

Removing the timing reward improves final burn and within-one-hour recall for seeds 42 and 137 but worsens both for 2027, showing no consistent timing benefit from the temporal term. Range-experiment timing estimates are unavailable.

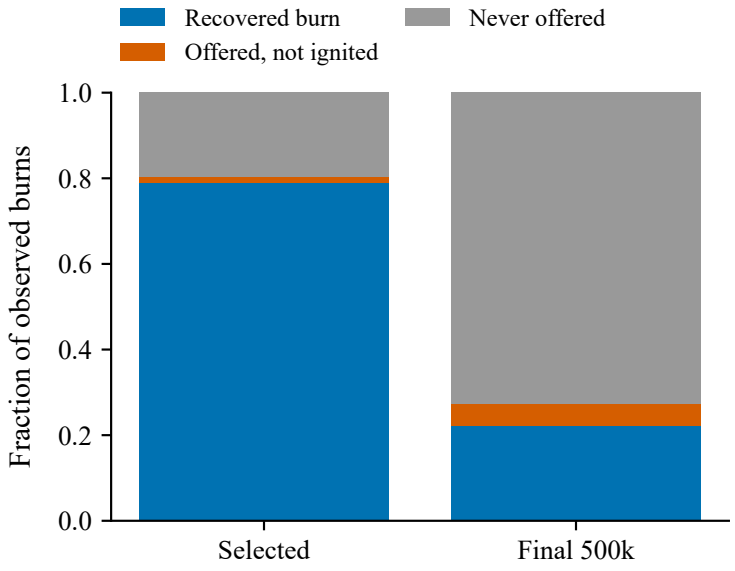
E.5 Spatial holdout

We clustered building coordinates into five regions in EPSG:3310 using K-means (Lloyd, 1982) (seed 19021, 20 initializations), withholding the westernmost region and a 100 m buffer. The region was chosen with knowledge of progression geography, before policy results, so this is an exploratory within-fire test. It contains 5,648 nodes, including 5,636 scored buildings (3,590 positive and 2,046 negative proxies). Training uses 12,469 nodes, including 12,368 scored buildings, and 267 buffer nodes are excluded. All 71 initially ignited buildings remain in the training area.

Feature scaling and a fresh ten-epoch link-prediction encoder were fitted only on the training subgraph. Three fresh policies (seeds 42, 137, 2027) trained for 500k actions over the full 15-hour horizon. Test and buffer nodes were excluded from training graph computations, physical source and receiver tables, rewards, and checkpoint selection. The simulation backend also loads a table of supervised damage probabilities fitted on full Eaton, inherited from earlier work. WUIRL's offer construction and policy features do not read this table, but as a precaution the holdout replaced it with a constant 0.5. Boundary reward weights were also set to one. These protocol differences prevent attributing changes from the original experiment solely to spatial exclusion.

Checkpoints were selected by mean training-region score over seeds 2001, 4001, and 6007, before test evaluation. Selected targets were 250k, 50k, and 100k for seeds 42, 137, and 2027. Evaluation restored the full graph and used the frozen training-region encoder and scaling. Upstream ignitions were simulated rather than supplied from later observed outcomes. Each policy and both baselines used the ten reporting seeds listed in Appendix D, and only held-out buildings were scored.

Table 7 compares WUIRL with both baselines on held-out buildings. Held-out burn recall was 64.2%, 21.8%, and 79.9% across the three training seeds, and the corresponding training-region selection recall was 64.6%, 28.9%, and 84.5%. Seed 137 was already weak on the training region. Among its missed held-out burns, 94.4% were never offered as candidates, which reflects limited



(a) Candidate availability

Figure 7: Candidate availability among 10,392 scored observed burned buildings. Never-offered fractions increase from 19.7% to 72.7% as recall falls. Availability is specific to the simulated trajectory, not permanent graph reachability. This comparison does not establish the initiating cause of deterioration.

opportunity along the simulated trajectory rather than permanent graph disconnection. The result tests conditional within-fire transfer, with fixed interface conditions and other retrospective inputs retained, not prospective prediction across fires.

E.6 Training stability

Reconstruction deteriorates with continued training for all three training seeds. On the reporting evaluations, mean burn recall decreases from 78.9% at the selected checkpoints to 22.2% at 500k actions. A fixed 50k–500k comparison on the three selection seeds also shows deterioration, as recall falls from 80.1% to 6.6%, 66.3% to 3.4%, and 75.9% to 62.5% for training seeds 42, 137, and 2027, respectively. Thus, the decline is also observed without comparing against selected optima, although it is not monotonic. As recall falls, the fraction of observed burned buildings never offered as candidates rises from 19.7% at selected checkpoints to 72.7% at final checkpoints (Figure 7).

Setting GAE $\lambda = 1$ did not prevent deterioration and instead worsened it. Mean burn recall falls from 71.3% at selected checkpoints to 11.0% at final checkpoints, with lower final scores for all three training seeds. This setting retains critic bootstrapping at rollout boundaries.

E.7 Hourly reward and credit audit

We audit original seed-42 action and update logs. Checkpoint recall averages evaluation seeds 2001, 4001, and 6007, whereas training quantities describe on-policy rollouts, not matched evaluation states.

None of three candidate explanations consistently tracks deterioration (Table 8). Reward-free buffers become less common, critic explained variance is already poor at high recall and becomes less

Table 8: Original seed-42 deterioration against candidate explanations. Each row uses the checkpoint's preceding interval of logged updates. Buffers without direct reward are those with no nonzero immediate reward, relying on critic bootstrapping. EV denotes explained variance. Intervals are descriptive comparisons for one seed.

Checkpoint	Burn recall	Reward-free buffers	Mean pathway entropy	Median critic EV
50k	80.1%	16/25	0.859	−16.07
200k	62.3%	16/24	0.912	−6.44
250k	36.3%	14/25	0.905	−5.50
350k	41.0%	10/24	0.906	−5.09
400k	9.8%	3/25	0.858	−1.86
450k	3.3%	0/24	0.713	−0.13
500k	6.6%	0/25	0.695	−0.28

negative, and pathway entropy changes little during the first large decline. Changing states and bootstrapped targets limit interpretation of critic variance, and these patterns establish no causal diagnosis.

Hourly reward settlement creates a credit-resolution limitation. In total, 92 of 245 buffers contain no nonzero immediate reward and depend on bootstrapping. With $\gamma = 0.9999$ and $\lambda = 0.95$ per minute, the GAE residual trace has a 13.5-minute half-life and an hour-later coefficient of 0.046 within a rollout. The critic cannot distinguish intermediate selection plans within a fixed frontier, yielding identical raw advantages for consecutive zero-time, zero-reward decisions with equal values. These limitations do not establish a cause of deterioration or a remedy from $\lambda = 1$. Isolating update directions would require controlled gradient measurements.

E.8 Critic-input follow-up

From the same seed-42 checkpoint at nominal 225k actions (actual 225,280), we compared the original critic with one augmented by current exposure-plan information. Actor, physics, reward, and PPO settings were retained. Each condition used continuation seeds 49001 and 49002, fresh environments, and five complete 15-hour episodes, matching episodes rather than actions or updates. Episode checkpoints used selection seeds 2001, 4001, and 6007.

Starting recall was 75.4%. After five episodes, original-critic runs reached 66.3% and 84.6%, and augmented runs reached 15.4% and 86.1%, with intermediate deterioration and recovery. Added inputs were not a consistent remedy. Update KL stayed below 0.0025, gradient clipping occurred in 20 of 1,248 steps, and the failing augmented run retained high normalized free-choice entropy. These diagnostics neither identify the cause nor exclude cumulative policy drift or advantage errors.