DIGITAL TWINS: A SIMULATION-BASED REALTIME DEEP REINFORCEMENT LEARNING APPROACH FOR FIGHTING WILDFIRES

Jose Tupayachi[a] and Xueping Li[a]

[a]Ideation Laboratory, The University of Tennessee, Knoxville, USA
jtupayac@vols.utk.edu x.li@utk.edu

ABSTRACT

This research introduces a methodology for wildfire management, employing digital twins in conjunction with a real-time deep reinforcement learning (DRL) framework. The experimentation involves parameterized scenarios within simulated environments, the developed digital twin operates under two stages. Initially, fire stations are strategically selected based on their capacities, followed by a collaborative engagement of clustered stations in firefighting activities. The deep reinforcement approach guides this collaboration, optimizing the decision-making process under the apperance of multiple to-be-controlled scenarios. An allocation approach is applied to assign firefighting entities, utilizing real-time satellite imagery and historical data, a routing engine, and the fire stations parameters analyzed. This ensures a coordinated and prompt response to evolving wildfire situations. In the subsequent stage, selected stations are grouped into clusters, representing independent agents, while fire events serve as the dynamic environment influencing agent actions. Agents aim to derive optimal policies in real time to minimize the impact of wildfires on the population. The holistic approach presented in this study strives to synergize resources efficiently, using real-time data, to create an adaptive and effective wildfire response system.

Keywords: Wildfires, Deep Reinforcement Learning, Resource Allocation, Digital Twins, Realtime Decision Making.

1 INTRODUCTION

Wildfires are one of the most expensive and deadly natural catastrophes in the world, particularly in the western United States. In 2018, a wildfire known as the Camp Fire was recorded as one of the worst wildfires in the history of California. Preventive measures could have lessen the effects of this catastrophic event and those subsequents that may occur due to terrein conditions or human initiated. Yet these catastrophic fires are expected increase by the year 2050 [1] if preventive actions are not taken nor improvement of response services are studied.

Fire events are responsible of large-scale evacuations, infrastructure being destroyed, millions of hectares of forest resources damaged, and most critically, human lives being in danger.

Wildfires impact populations differently communities marked by challenges—financial limitations, cultural and institutional obstacles, restricted mobility, and compromised physical health. are at heightened risk of bearing a disproportionate burden when confronted with the effects of wildfires. Literature, highlights preventive measures as effective and highly cost effective [2]. Although reducing hazard activities should be priorotized: Controlling wildland fuels, constructing houses with enough separation to historically affected areas and fire resistant infraestructure. Federal reports such as "On Fire: The Report of the Wildland Fire

Proc. of the 2024 Annual Simulation Conference (ANNSIM'24), May 20-23, 2024, American University, DC, USA
P.J. Giabbanelli, I. David, C. Ruiz-Martin, B. Oakes and R. Cárdenas, eds.

Mitigation and Management Commission" elaborates a set of guidelines for wild fire control and prevention. Fire detection and monitoring technologies have evidenced a rapid deployment including ground sensors, unmaned aerial vehicles, and satellite imagery. Nonetheless, current systems may encounter drawbacks due to delayed detection or false negative recogniztion that, in conjuction, may jeopardize preventive measures. Due to this preventive measures must be paired with methods that desicion-makers can benefit from under instances where pre fire events mitigations plans are not enough to properly contrarest the effect of fire events. Resource allocation can be promoted as an initial phase given a fire emergency. Then how the fire will be contained poses as a second problem. For the proposed tasks we

This enables community leaders to pinpoint locations where older housing may require retrofitting to enhance resistance against wildfires and customize outreach initiatives to engage nonresident landowners effectively. The relevance lies in understanding wildfire risk can assist communities in prioritizing preventative and mitigation efforts in order to target the most vulnerable individuals. Initiatives and data based tools like those developed for the USDA Forest Service if a community is directly threatened by wildfire from adjacent combustible vegetation or indirectly threatened by embers. Wildfires are considered complex scenario when it comes to simulation and proper rource ahndling due to that these tend ti vary This research is motivated on the baisis of the complexity of resource allocation. As Wildfire containment involves the allocation of various resources, including firefighting personnel, equipment, and supplies. High complexity due to factors such as the dynamic nature of wildfires, varying terrain, and resource constraints This poses a rich environment for technologies paired with simulation based platforms such as deep reinformcement learning approaches to tackle effectively the decision making process in high complexity scenarios. Research on simulators utilizing deep reinforcement learning for wildfire detection has primarily focused on firefighting strategies with limited emphasis on resource allocation and decision-making, despite the significant societal benefits. In a notable study [3], a decentralized deep reinforcement learning strategy is employed for a fleet of Unmanned Aerial Vehicles (UAVs) to autonomously combat forest fires. The authors opt for a deep RL approach due to the computational challenges of exact and approximate solutions using dynamic programming. Monte Carlo simulations demonstrate the superior performance of the deep RL policy compared to a manually crafted heuristic, showcasing adaptability to varying forest sizes, UAV quantities, and model parameter variations.

In another study [4], deep reinforcement learning is applied to address complex conservation decision problems, providing a conceptual and technical introduction along with annotated code. The study showcases the effectiveness of deep RL in solving sequential decision-making challenges related to fisheries management and ecological tipping points. In our approach, parameterized experiments unfold across simulated scenarios using a digital twin. The twin operates in two stages: firstly, selecting fire stations based on their capacity, and then, clustered stations collaboratively engage in firefighting guided by the optimal policy determined by the deep reinforcement model. An allocation approach strategically assigns firefighting entities to neighboring areas, leveraging real-time data from satellite imaging and a routing engine for a coordinated and prompt response. Subsequently, the objective is to group selected stations into clusters and fire events, treating firefighting clusters as independent agents and fire events as dynamic environments influencing agent actions. Agents seek the optimal policy to minimize wildfire impact on the population. This holistic approach aims to synergize real-time data-driven resources into an effective and adaptive response system. The primary contributions of our work lie in these innovative approaches to resource allocation, decision-making, and firefighting strategy within the context of wildfire management.

- We introduce a digital twin framework for wildfire management with parameterized experiments to strategically allocate firefighting entities to neighboring areas based on a proposed fire station index.
- Utilize deep reinforcement learning to determine an optimal policy for clustered fire stations in firefighting scenarios by seeking optimal policies to minimize the impact of wildfires on the population

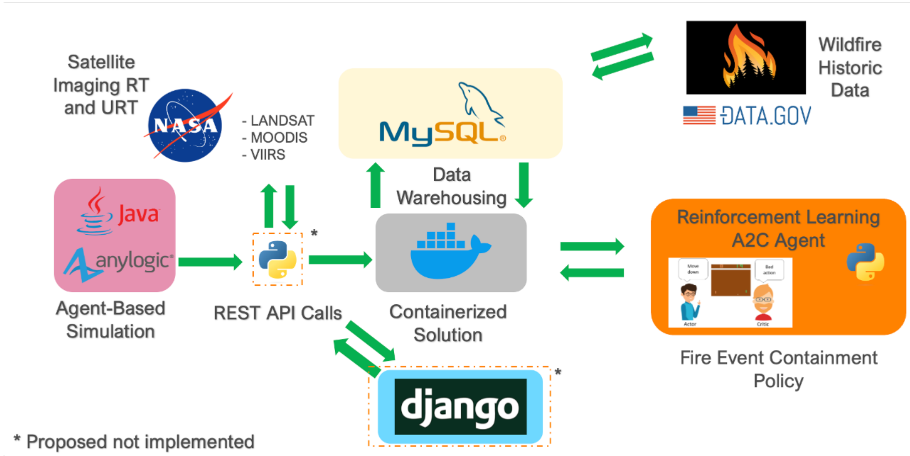

Figure 1: System Architecture: The illustration depicts the fundamental components of this digital twin. Initially, satellite imaging is employed from three sources—Landsat, MODIS, and VIIRS. Each dataset is standardized, offering a confidence level for the likelihood of a fire event at each represented heat point in the future. This data is presented in tabular format and forms the basis for the backend, comprising a Django-based API and a containerized MySQL platform. The API operates as a RESTful interface for socket communication and serves as an executor of operations within the inner backend. Real-time traffic information is sourced from TomTom live data, communicating with the OSRM routing engine to generate a realistic cost matrix for transportation in optimizing neighboring station placement. Following optimization, the results are visualized in the AnyLogic GUI. The second stage involves a reinforcement learning-based policy aiming to select the policy that minimizes the economic impact caused by a fire event. An advantage actor-critic is utilized to train independent clusters for each agent, empowering them to decide on the necessary containment actions for fire events.

- Incorporation of real-time data from satellite imaging and a routing engine for enhanced decision-making.
- Development of a holistic and adaptive response system, integrating deep reinforcement learning for effective wildfire management.

The structure of this article is as follows: Section 1 provides an overview of the novel contribution. Section 2 introduces the necessary environment to develop the digital twin system. We introduce the problem in section 3. Under Section 4 the main aim of this article is expanded and the three methodological points are explored. Section 5 presents the proposed parameterized scenarios. This is followed by an analysis of the policy optimization process and the results obtained, that validate the effectiveness of the presented methods. Finally the conclusion and further discussion explores potential directions for future research. The developed code is freely accessible to the public at https://github.com/josetup123/Advanced_Simulation_Paper

2 ARCHITECTURE AND INFORMATION COLLECTION

3 PROBLEM DESCRIPTION

critical challenge is the limited effectiveness of traditional strategies in rapidly evolving and dynamic fire scenarios. Conventional approaches often struggle to adapt to the unpredictable nature of wildfires, leading to suboptimal resource allocation, delayed decision-making, and increased environmental impact Additionally, the lack of real-time decision support systems hampers the ability of firefighting teams to respond

θ: Parameters of the policy function.
ϕ: Parameters of the state-value function.
$\pi(a|s;\theta)$: Policy function representing the probability of taking action a in state s with parameters θ.
R_t: Return at time t (sum of discounted rewards over time).
γ: Discount factor.
A_t: Advantage function at time t, representing the advantage of taking action a_t in state s_t.
$V(s;\phi)$: State-value function with parameters ϕ.

The A2C objective function is given by:

$$L(\theta,\phi) = \sum_{t=0}^{T} \left(\log \pi(a_t|s_t;\theta) \cdot A_t + \beta \cdot \mathrm{H}(\pi(\cdot|s_t;\theta)) \; - \alpha \cdot \frac{1}{2}(V(s_t;\phi) - R_t)^2 \right) \tag{1}$$

Where:

$$\mathrm{H}(\pi(\cdot|s_t;\theta)) \text{ is the entropy regularization term,} \tag{2}$$

$$\beta \text{ is the entropy regularization coefficient,} \tag{3}$$

$$\alpha \text{ is the coefficient for the value function loss.} \tag{4}$$

swiftly and effectively Develop an innovative solution that seamlessly integrates digital twins, simulation-based methodologies, and real-time deep reinforcement learning to address the shortcomings of existing wildfire management approaches. This solution aims to provide actionable insights, improve resource utilization, and minimize the environmental impact of firefighting operations, ultimately contributing to a more adaptive and proactive wildfire management paradigm

4 METHODOLOGY

5 EXPERIMENTATION

5.1 Data Sources

The system's information architecture is enriched through the judicious integration of diverse datasets, notably leveraging FIRMS (Fire Information for Resource Management System). This repository provides real-time and pertinent information on active fires, affording the system a dynamic understanding of ongoing fire events. Complementing this, the incorporation of data from Open Fire Stations enhances the spatial awareness of existing fire station locations, thereby facilitating the optimization of neighboring station placements. Furthermore, the system draws insights from original sources, notably the United States Forest Service, establishing a foundation rooted in authoritative forestry data. This comprehensive amalgamation of information from FIRMS, Open Fire Stations, and original sources contributes to the system's scholarly rigor, ensuring a nuanced understanding of the multifaceted variables inherent in fire event management and response. The system enhances its data foundation through integration with Data.gov, a reputable source providing diverse datasets. By incorporating historical fire event registries from Data.gov, the system gains valuable insights into past incidents, enabling a nuanced analysis of patterns and trends for more informed decision-making in fire event management.

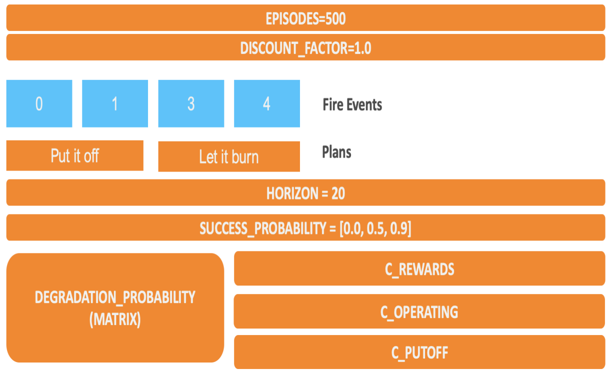

Figure 2: Deep Reinformcement Leanring Parameters: The experimental parameters are configured as follows: `EPISODES=50000` dictates the number of agent simulations for policy improvement; `DISCOUNT_FACTOR=1.0` emphasizes the full consideration of future rewards; `seed_value = 42` ensures reproducibility through a set seed for the random number generator. `num_states = Max_FirePoints` signifies the count of environmental states, often representing diverse fire scenarios or locations. `num_plans = 2` denotes the available action choices for the firefighter, likely relating to deciding whether to address a fire. `T = 20` defines the time horizon, indicating the number of time steps before updating policies or the simulation duration. `success_pr = [0.0, 0.5, 0.9]` characterizes the success probability of extinguishing a fire, with values indicating varying success levels. The `degrade_pr` matrix portrays degradation probabilities for each state, reflecting the likelihood of environmental deterioration. `C_reward = 10` denotes the reward for fire prevention per time step, incentivizing a safe environment. `C_Operating` encapsulates operating costs, suggesting state-dependent variations. `C_Putoff` captures put off costs, contingent on the chosen action, likely reflecting the expenses associated with fire suppression strategies.

5.2 Neighboring Station

The determination of neighboring fire stations is derived through the following formulation, structured around a supply-demand schema. Its objective is to ascertain the optimal number of fire stations capable of responding to a fire event based on a designated index referred to as the fire_index. This index integrates various factors obtained from independent data sources, which have been precomputed for experimental accuracy. The formulation details are provided in the subsequent lines:

5.3 Fire Containment

6 CONCLUSION & FURTHER DIRECTIONS

The presented approach outlines a comprehensive digital twin framework for wildfire management that leverages various innovative technologies. The integration of parameterized experiments facilitates the strategic allocation of firefighting entities to neighboring areas, guided by a novel fire station index. The utilization of deep reinforcement learning is a key highlight, as it determines optimal policies for clustered fire stations, aiming to minimize the impact of wildfires on the population. This not only enhances the efficiency of firefighting scenarios but also demonstrates a forward-looking approach to addressing dynamic and complex challenges.

The incorporation of real-time data from satellite imaging and a routing engine further enhances decision-making capabilities, allowing for a more accurate and timely response to evolving wildfire situations. By

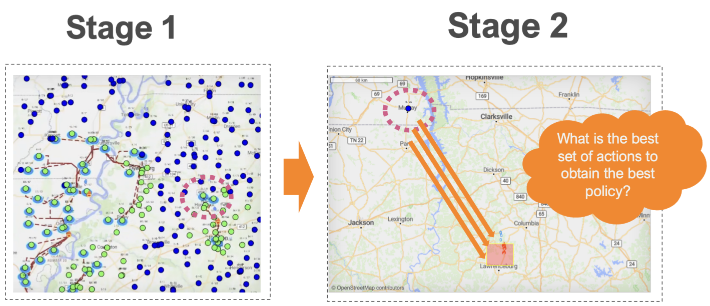

Figure 3: Stages Transfering: We illustrate the process of transitioning from the initial stage to the subsequent deep reinforcement learning (DRL) model in a two-stage framework. Initially, our Neighboring approach strategically allocates the requisite number of fire stations to address the prevailing number of fire events in a given area. The fire stations surrounding each fire event are then clustered, constituting Stage two. In this stage, the individual fire stations and associated resources are treated as a unified entity. The DRL approach subsequently formulates a policy, determining the optimal actions for each fire event—whether to contain, extinguish, or withhold resource allocation—at discrete time intervals, starting at $t = 0$. Visualized through three arrows, these actions represent the various states the agent (comprising a cluster of fire stations and resources) can assume during the second stage. Our temporal horizon spans 20 time steps, with each step equivalent to an hour in Anylogic units, providing a comprehensive representation of the evolving dynamics.

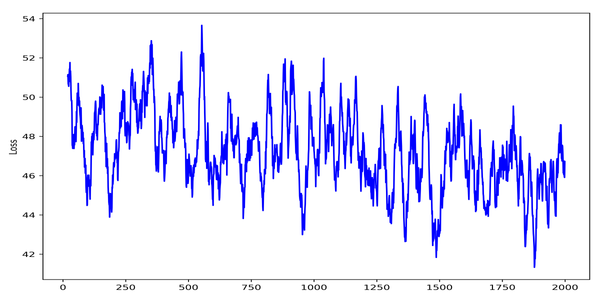

Figure 4: Actor Loss: It indicates the cost of training the policy network. The actor loss leads policy modification to raise the likelihood of actions leading to better returns and decrease the likelihood of actions leading to lower returns. It is a component of the overall loss function, which combines actor and critic losses with the goal of maximizing the predicted cumulative reward by changing both the policy and value functions.

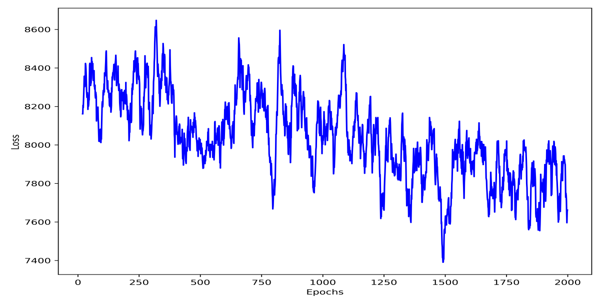

Figure 5: Critic Loss: In Advantage Actor-Critic (A2C), the critic loss indicates the difference between the projected state value and the actual return seen during reinforcement learning training. The critic is in charge of estimating the state-value function, which calculates the predicted cumulative reward from a particular condition in A2C. The critic loss, which is frequently expressed as the mean squared error between the anticipated state value and the actual return, is a measure of how well the critic's predictions match the genuine rewards received in the environment.

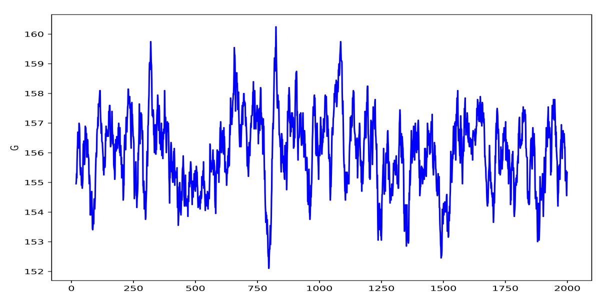

Figure 6: This section elucidates the profound impact of actions on the overarching policy return. A well-converged model is anticipated to showcase a discernibly smoother pattern, contrasting with the inherent noise observed during earlier stages. The graphical representation serves to illustrate the pivotal role actions play in shaping the cumulative return of the policy, offering insights into the stability and effectiveness achieved through convergence.

	epochs	reward	state	action	return
0	0	10	state: (0, 0)	0	NaN
1	1	9	state: (1, 1)	0	NaN
2	2	8	state: (2, 1)	1	NaN
3	3	8	state: (3, 2)	0	NaN
4	4	7	state: (4, 3)	0	NaN
5	5	6	state: (5, 3)	1	NaN
6	6	7	state: (6, 3)	0	NaN
7	7	7	state: (7, 3)	0	NaN
8	8	6	state: (8, 3)	1	NaN
9	9	7	state: (9, 3)	0	NaN
10	10	6	state: (10, 3)	1	NaN
11	11	9	state: (11, 1)	0	NaN
12	12	8	state: (12, 2)	0	NaN
13	13	6	state: (13, 3)	1	NaN
14	14	9	state: (14, 1)	0	NaN
15	15	8	state: (15, 2)	0	NaN
16	16	7	state: (16, 3)	0	NaN
17	17	7	state: (17, 3)	0	NaN
18	18	7	state: (18, 3)	0	NaN
19	19	7	state: (19, 3)	0	NaN
20	20	7	state: (20, 3)	0	156.0

Figure 7: Action Schema: This comprehensive depiction provides a detailed examination of state rewards, actions, and individual states. To navigate the intricacies of the data, it is advisable to implement a state-action filter to prevent unnecessary traversal through events where fires have already been extinguished. Notably, an observable pattern emerges, highlighting the effective control measures applied to fire event 3, also denoted as state 3, which demonstrate a successful containment strategy deployed multiple times.

Table 1: This table represents the results of various test events conducted to assess system performance. Five distinct tests were executed, systematically increasing the number of fire events to evaluate their impact on the overall system return. Notably, our analysis focused on a carefully selected cluster of data points. For this specific cluster, it became evident that exceeding eight fire events would likely surpass the capacity of the fire points cluster, inevitably leading to repercussions on the population. It is crucial to emphasize that the policies implemented during these tests are still in the developmental phase. Ongoing efforts are directed towards further refinement and completion in the subsequent stages of this research. These findings underscore the need for cautious consideration and ongoing optimization as we progress in the development of robust policies for managing fire events.

Test	Fire Events	Return
1	4	156
2	8	97
3	80	-961
4	100	-1224
5	800	-11465

integrating these technologies, the framework creates a holistic and adaptive response system. The emphasis on deep reinforcement learning throughout the system signifies a commitment to effective wildfire management through continuous learning and optimization. This approach represents a significant step towards leveraging cutting-edge technologies for more efficient, data-driven, and adaptive wildfire management strategies.

REFERENCES

[1] U. Nations. As wildfires increase, integrated strategies for forests, climate and sustainability are ever more urgent. Publisher: United Nations. [Online]. Available: https://www.un.org/en/un-chronicle/wildfires-increase-integrated-strategies-forests-climate-and-sustainability-are-ever-0

[2] P. Baylis and J. Boomhower, "The economic incidence of wildfire suppression in the united states," vol. 15, no. 1, pp. 442–473. [Online]. Available: https://www.aeaweb.org/articles?id=10.1257/app.20200662

[3] R. N. Haksar and M. 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, ISSN: 2153-0866. [Online]. Available: https://ieeexplore.ieee.org/document/8593539

[4] Deep reinforcement learning for conservation decisions - lapeyrolerie - 2022 - methods in ecology and evolution - wiley online library. [Online]. Available: https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13954

AUTHOR BIOGRAPHIES

JOSE TUPAYACHI, UTK graduate student, with expertise in digital transformation, simulation, and machine learning. With a background in information technologies and two years in the industry, Jose has consistently engineered cutting-edge solutions that are both scalable and efficient, effectively tackling complex technical challenges. His areas of interest extend across a diverse spectrum of programming languages and development frameworks, including but not limited to Python, Java, and SQL.

XUEPING LI is a Professor and Dan Doulet Faculty Fellow in the Department of Industrial and Systems Engineering, Co-Director of the Health Innovation Technology and Simulation (HITS) Lab, and the Director of the Ideation Laboratory (iLab) at the University of Tennessee - Knoxville. His research areas include complex system modeling, simulation, and optimization with broad applications in supply chain logistics, healthcare, and energy systems. His research has been sponsored by federal agencies including NSF, NIH, DOE, and HRSA, and other industry partners. He has over 150 peer-reviewed publications and 10 invention disclosures. He is an IISE Fellow and a member of IEEE and INFORMS, and the Past President of the Modeling of Simulation Division of IISE.