Aerial Coverage Path Planning in Nuclear Emergencies

Johann Blake[1,2,3] , **Matthias Schubert**[1,3]
[1]LMU Munich
[2]German Federal Office for Radiation Protection
[3]Munich Center for Machine Learning (MCML)
johanndavidblake@gmail.com, schubert@dbs.ifi.lmu.de

Abstract

We formulate a Coverage Path Planning (CPP) problem for a helicopter or a UAV tasked with mapping ground-level radiation while avoiding radiation that is too strong. We introduce a simulation environment that incorporates digital elevation models, altitude-dependent measurement footprints and realistic flight constraints, as well as state-of-the-art radiation scenario simulations, such as nuclear explosions, provided by the German Federal Office for Radiation Protection. We highlight the complexity of radiological survey missions and demonstrate the necessity for new CPP approaches that address these unique challenges. The code to our simulation environment can be found under https://github.com/JohannBlake/Aerial-Coverage-Path-Planning-in-Nuclear-Emergencies.

1 Introduction

Nuclear power is becoming increasingly relevant as a source of clean energy. However, nuclear emergencies, such as the explosion at the Fukushima nuclear power plant, serve as stark reminders of the risks associated with nuclear technologies. Radiation poses significant hazards to both people and the environment, necessitating effective monitoring and intervention strategies.

1.1 Motivation

Following incidents involving radioactive contamination, a critical task for government agencies is to map radiation levels in affected areas. This is currently done using helicopters equipped with radiation measurement devices, which collect two key data types: (1) the *local dose rate*, i.e., the radiation intensity reaching the helicopter at its position, and (2) the *ground radiation level*, i.e., the radiation intensity at the surface.

The goal is to cover the area of interest to create a map of ground radiation levels while avoiding excessive local dose rates, which can contaminate equipment and endanger personnel. Existing flight paths are not optimized for radiation avoidance.

Our research enables the German Federal Office for Radiation Protection to plan real-time flight paths for helicopters and UAVs, ensuring effective and safe deployment during nuclear emergencies. While UAVs can withstand higher radiation exposure than manned helicopters, limits remain due to contamination risks.

1.2 Simulation Requirements for CPP in Nuclear Emergencies

Our simulation models CPP in nuclear emergencies with high realism, extending into three dimensions to allow a helicopter to adjust its altitude while surveying and avoiding radiation. Flight dynamics are modeled through an action-based system, where the agent controls its heading and altitude via continuous angular adjustments. This ensures that movement adheres to realistic aerial constraints.

A downward-facing measurement cone defines the surveyed area, with its size determined by altitude and terrain. Higher altitudes expand coverage and lower radiation exposure but reduce measurement quality.

The helicopter's path should minimize sharp turns, as frequent course changes are difficult to execute and degrade measurement accuracy. To enhance realism, we integrate detailed terrain data and radiation maps, including state-of-the-art simulations of nuclear incidents from the German Federal Office for Radiation Protection.

Our system accounts for the dynamic nature of radiation exposure, which varies with time and altitude and is only known at explored positions. This requires analyzing radiation gradients experienced at the already explored positions to adjust flight paths in real-time. In mountainous regions, optimizing altitude is challenging due to sparse rewards, which are only given for measuring new areas at the correct elevation, complicating policy development.

1.3 Distinction From Existing Work

Although our problem shares similarities with CPP in unknown environments—such as the one described in [Jonnarth *et al.*, 2024], which includes applications like autonomous vacuum navigation—existing CPP methodologies do not directly apply to our setting. This is because of the different dynamics of radiation avoidance (Radiation gradients instead of strict borders), height maintenance, the constraint on turns and the flight paths one can follow due to the three-dimensional nature of the environment.

2 Definition of the Task

The primary objective of our flight planning task is to maximize the covered area while strictly adhering to radiation limits. Radiation avoidance can be approached in two ways: making sure the local dose rate at any given time does not exceed a legally enforced limit or ensuring that the cumulative dose over the entire mission remains below thresholds vital for health and contamination control.

Beyond radiation considerations, the flight path's simplicity is crucial. Excessive maneuvering makes piloting more challenging, increases fuel consumption, and can undermine measurement accuracy due to tilt-induced changes in the orientation of both the helicopter and the measurement equipment. Furthermore, maintaining the correct flight altitude—typically around 90m above ground— is necessary to ensure optimal measurement quality. Finally, the helicopter must always return to the airport or base within a predetermined time frame, which further constrains the design of the flight plan. Last but not least minimizing the overall flight time is paramount in CPP missions.

2.1 Definition as a Markov Decision Process (MDP)

We formulate the flight planning task as an MDP defined by the tuple (S, A, T, R).

The **State Space** S includes the helicopter's measured area, area of interest, history of local dose rates with their positions, helicopter's current position and direction, elevation data, and remaining time.

The **Action Space** A is defined as the set of continuous two-dimensional vectors in the range $[-1, 1]^2$, where each component specifies a directional adjustment along the horizontal and vertical axes (see Section 3).

The **Transition Function** T is governed by the mechanics of the environment, including the helicopter's movement dynamics, altitude-dependent radiation exposure, and interactions with the terrain and radiation scenarios detailed in Section 3.

The objective consists of 5 components.

- R_1: Maximize the measured area at the optimal altitude of 90m.
- R_2: Adhere to radiation exposure limit, ensuring safety limits are not exceeded.
- R_3: Minimize the amount of turns.
- R_4: Minimize the overall flight time to enhance operational efficiency.
- R_5: Ensure the helicopter returns to the base within the specified time frame.

We define the **Reward Function** R as:

$$R(s,a) = \sum_{i=1}^{5} w_i R_i(s,a)$$

where w_i are weighting factors that balance the importance of each objective.

Alternatively, we can use constrained RL by treating radiation exposure as a constraint.

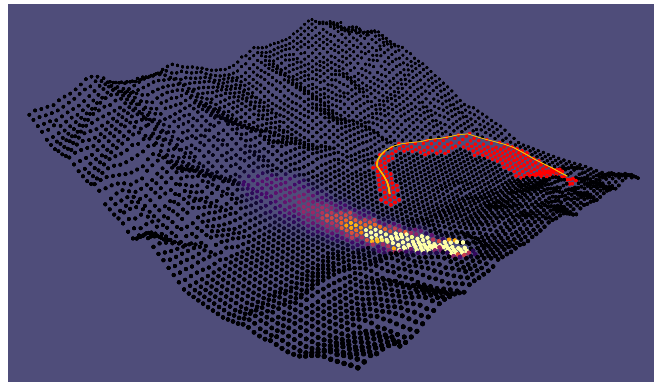

Figure 1: The height of points corresponds to the digital elevation model, red indicates the measured area, black represents the unmeasured area, the "Inferno" color map depicts the radiation scenario, and the orange line shows the helicopter's flight path.

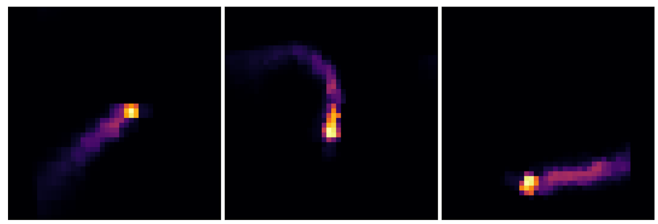

Figure 2: One original (left) with two derived radiation scenarios.

3 Simulation Environment

We use OpenAI's standardized Gymnasium interface due to its strong open-source support, which simplifies development [Towers *et al.*, 2024]. Our simulation environment features the following components.

Geo data. Terrain data is obtained from the digital elevation model available as GeoTIFF files from the ALOS Global Digital Surface Model (AW3D30) project [Japan Aerospace Exploration Agency (JAXA), 2024]. The helicopter's position, orientation, and altitude are represented as geospatial data, ensuring accurate integration with the terrain model.

Radiation data. We received a limited number of radiation scenarios from the German Federal Office for Radiation Protection and generated alternative scenarios based on the originals to improve the model's generalization capabilities, ensuring robust performance across a variety of tasks. This was done by translating and distorting the original images. Figure 2 displays examples of an original and two generated radiation scenarios.

Local Dose Rate. Given a radiation scenario as the one depicted in Figure 1 where the color of each point x_k, y_k in the inferno coloring scheme represents the level of the radiation $\text{LDR}_{source}(x_k, y_k)$ for this single point at the ground. The local dose rate $\text{LDR}(x, y, 0)$ at a point (x, y) and height 0 is calculated using inverse distance weighting

$$\mathrm{LDR}(x, y, 0) = \frac{\sum_k \frac{\mathrm{LDR}_{source}(x_k, y_k)}{d_k^2}}{\sum_k \frac{1}{d_k^2}},$$

where d_k is the distance between (x, y) and the k-th known point. We adjust the result for height h by scaling $\mathrm{LDR}(x, y, 0)$ in an inverse logarithmic manner, ensuring that the local dose rate decreases as altitude increases.

The measured area. The measured area is defined as the intersection of the downward-facing measurement cone and the terrain. We have simulated the radius of the measuring cone in dependence of the type of radiation (energy) according to [Grasty *et al.*, 1979].

Impact of actions on environment. At each point in time, the agent's current direction is given by a vector in $\mathbb{R}^3$. The agent can change this direction by adjusting the horizontal angle $\theta \in [-\alpha, \alpha]$, and the vertical angle $\phi \in [-\beta, \beta]$. After applying θ and ϕ, the agent follows the updated direction vector for a fixed duration T. The parameters α, β, and T must be chosen to reflect the flight dynamics of the specific aerial vehicle (e.g., drone or helicopter). In our Gymnasium environment, we choose a 2-dimensional continuous action space $[-1, 1]^2$ representing these angle changes.

Visualization via "DECK.GL". We use DECK.GL[1] for its GPU acceleration capabilities, enabling efficient rendering of large datasets. It supports map integration, facilitating visualization in a geospatial context, and is well-suited for real-world deployment.

4 Demonstration Setup

Our MDP features a large state space and continuous actions, making reinforcement learning (RL) more suitable for the task than classical methods like a boustrophedon path (systematic back-and-forth movements). As shown in [Jonnarth *et al.*, 2024], RL effectively addresses similar CPP problems and it allows integration of additional objectives, such as minimizing turns, by incorporating reward penalties for direction changes.

We trained several models under varying environmental conditions using the Twin Delayed Deep Deterministic Policy Gradient algorithm [Fujimoto *et al.*, 2018] with Stable Baselines3 [Raffin *et al.*, 2021].

The Gymnasium environment's observation space includes the measured area, area of interest, history of local dose rates with their positions, helicopter's current position and direction, elevation data, and remaining time. The system is built to allow further extension if necessary.

In our demonstration, we present example flight paths from these models in flat and mountainous environments, with and without radiation. The example policies can be found at https://linktr.ee/johannblake, where you can find visual representations of the paths discussed below.

[1]https://deck.gl/

4.1 Policies in a Flat Environment

No radiation, straightforward coverage reward. This policy prioritizes direct traversal through unradiated regions but leads to fragmentation of unmeasured areas. As a result, flight time increases due to the need to revisit measured regions.

No radiation, coverage with TV reward. Leveraging the total variance (TV) reward from [Jonnarth *et al.*, 2024], this policy improves coverage efficiency while reducing fragmentation. However, despite turn penalties, excessive maneuvering remains.

With radiation, coverage with TV reward. This policy covers the area efficiently but lacks penalties for directional changes, leading to too many turns to navigate high-dose areas while respecting exposure limits.

With radiation, coverage with TV reward, punish turns. By penalizing directional changes, this policy reduces turns compared to the previous approach. However, fragmentation and excessive turns are still present, highlighting the challenge of balancing radiation avoidance with path efficiency, as shown in Figure 3.

4.2 Policy in a Mountainous Environment

When operating in mountainous terrains, the policies exhibit even more pronounced deficiencies. Because of the steep slopes and irregular terrain outside of the flat valley, the agent struggles to sustain the optimal altitude consistently, leading to sparse rewards, missing exploration and policy stagnation.

4.3 Impact of Reward Function and Environment

These examples underscore how the reward function and environmental factors—such as radiation levels and terrain topology—significantly influence the policies. The interplay between maximizing coverage, minimizing turns, maintaining optimal height and avoiding radiation creates a complex optimization problem. Balancing these objectives poses substantial challenges, particularly in dynamic and uneven environments like mountainous regions. The difficulty in achieving an optimal trade-off between these factors highlights the inherent complexities in developing robust CPP strategies for nuclear emergency scenarios.

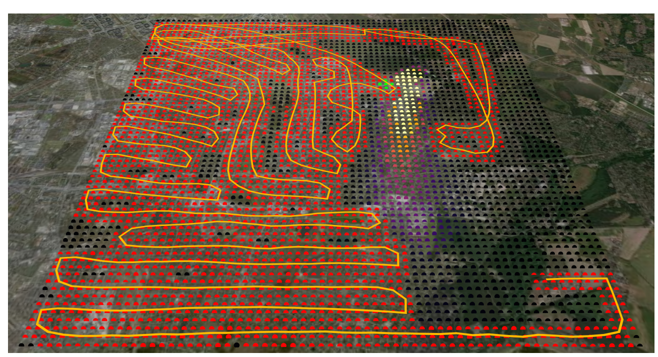

Figure 3: A policy in flat terrain with too many turns but a path avoiding radiation most of the time.

References

[Fujimoto *et al.*, 2018] Scott Fujimoto, Herke van Hoof, and David Meger. Addressing function approximation error in actor-critic methods, 2018.

[Grasty *et al.*, 1979] R. L. Grasty, K. L. Kosanke, and R. S. Foote. Fields of view of airborne gamma-ray detectors. *Geophysics*, 44(8):1447–1457, 1979.

[Japan Aerospace Exploration Agency (JAXA), 2024] Japan Aerospace Exploration Agency (JAXA). Alos world 3d-30m (aw3d30) [digital surface model], 2024.

[Jonnarth *et al.*, 2024] Arvi Jonnarth, Jie Zhao, and Michael Felsberg. Learning coverage paths in unknown environments with deep reinforcement learning. In *Proceedings of the 41st International Conference on Machine Learning*, 2024.

[Raffin *et al.*, 2021] Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268):1–8, 2021.

[Towers *et al.*, 2024] Mark Towers, Ariel Kwiatkowski, Jordan Terry, John U. Balis, Gianluca De Cola, Tristan Deleu, Manuel Goulão, Andreas Kallinteris, Markus Krimmel, Arjun KG, Rodrigo Perez-Vicente, Andrea Pierré, Sander Schulhoff, Jun Jet Tai, Hannah Tan, and Omar G. Younis. Gymnasium: A standard interface for reinforcement learning environments, 2024.