The Aeronautical Journal (2025), **129**, pp. 3379–3401
doi:10.1017/aer.2025.10054

ROYAL AERONAUTICAL SOCIETY

RESEARCH ARTICLE

Low-Noise Path Planning for Urban Drone Missions: Acoustic Ray Tracing and DDPG Algorithm

M. Rinaldi, S. Primatesta and G. Guglieri

Department of Mechanical and Aerospace Engineering (DIMEAS), Politecnico di Torino, Torino, Italy
Corresponding author: M. Rinaldi; Email: marco_rinaldi@polito.it

Received: 30 April 2025; **Revised:** 1 July 2025; **Accepted:** 8 July 2025

Keywords: drone noise; path planning; aerial robotics; DDPG algorithm; quiet dones; acoustic ray tracing; UAM

Abstract

This paper presents a comprehensive approach for mitigating noise pollution from unmanned aerial vehicles (UAVs) in urban environment through path planning using reinforcement learning (RL). The study focuses on Turin, Italy, leveraging its diverse urban architecture to develop a comprehensive model. A detailed 3D occupancy grid map, based on OpenStreetMap data, was created to represent buildings' locations and heights while a population density map was developed to account for demographic variances. The research develops a dynamic noise source model that adjusts noise emission levels based on UAV velocity, ensuring realistic noise impact predictions. Acoustic ray tracing techniques are utilised to simulate noise propagation, accounting for atmospheric absorption and reflections from urban structures, providing a detailed analysis of noise distribution. The core of this work is the application of the deep deterministic policy gradient (DDPG) algorithm within the RL framework. The algorithm is tailored to optimise flight paths by minimising noise impact while balancing other factors like path length and energy efficiency. The RL agent learns to navigate complex urban landscapes, integrating penalties for idling, excessive path length and abrupt manoeuvers to refine its path planning strategy. Simulation results with several maps unseen during training reveal that the RL-based approach effectively reduces noise impact in urban settings, making it a viable solution for better integrating UAVs into urban air mobility (UAM) systems. The methodology is scalable and adaptable, with potential applications in various urban environments globally. This research contributes to the development of sustainable drone operations in UAM context by addressing the critical issue of noise pollution, enhancing public acceptance and regulatory compliance.

List of symbols

UAV	unmanned aerial vehicle
SPL	sound pressure level
RL	reinforcement learning
SDG	sustainable development goal
HPM	high-priority model
LPM	low-priority model
DDPG	deep deterministic policy gradient
v	drone velocity
L_0	noise source
L_{tot}	total SPL at ground level
L_i	attenuated SPL of acoustic ray i at ground level
A_{div}	geometric divergence attenuation
A_{atm}	atmospheric absorption attenuation
d	distance between noise source and receiver
d_0	reference distance
α	atmospheric absorption coefficient

© The Author(s), 2025. Published by Cambridge University Press on behalf of Royal Aeronautical Society.

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

$(x,y,z)_t$	vector of UAV coordinates at time instant t
$(x,y,z)_{dir}$	direction vector of acoustic ray
$(x,y,z)_{ref}$	direction vector of reflected acoustic ray
(dx,dy)	vector of UAV incremental movements
T	sample time
N	total number of acoustic rays
$\boldsymbol{I}$	vector of incident acoustic ray
$\boldsymbol{R}$	vector of reflected acoustic ray
$\boldsymbol{N}$	normal vector of surface
$R\,(t)$	reward function value of timestep t
λ	scaling factor
p_t	position of the drone at timestep t
Pd_t	population density of the cell where the drone is located at timestep t
p_{max}	maximum range of motion of the drone
N_{thr}	maximum noise peak
L_t	drone's noise at timestep t
N_t	cumulative noise up to timestep t
N_{MAX}	maximum noise at each timestep
θ_t	orientation of (dx,dy) at timestep t
$P_j\,(t)$	j^{th} penalty function value at timestep t
$v\,(t)$	actions' additive noise at timestep t
$\mathcal{N}\,(0, 1)$	normal random variable
$Mean_{ac}$	mean attraction constant
$Mean$	mean additive noise
std	standard deviation
DeR	decay rate of additive noise
τ	smoothing factor
π	actor policy
a	action
s	state
S	observation
Q	critic network
Q_t	target critic network
R	replay buffer
m_B	mini batch size
π_t	target actor policy
θ	actor parameter
ϕ	critic parameter
θ_t	target actor parameter
ϕ_t	target critic parameter
γ	discount factor
d_b	binary terminal state variable
α_R	learning rate

1.0 Introduction

Drone-based applications in urban air mobility (UAM) have the potential to revolutionise how cities operate, from healthcare to infrastructure management. Drones' applications in smart city context are many, such as logistics (including delivery of medical samples), infrastructure inspection, surveillance, traffic monitoring, emergency response and air taxi, among others [1–6]. All these uses of drones in the UAM context can significantly transform urban living, improving efficiency, lowering costs, and delivering faster, safer, and greener services. To fully realise these possibilities, they need to address a range of technological, regulatory, and societal issues. Surmounting those barriers will be critical to their

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

wide adoption and success in the UAM environment [7]. It is well-known that the social acceptance of UAVs by the public is hindered, mainly, by concerns over privacy, safety, and noise pollution [8]. UAVs produce distinct noise signatures characterised by high-frequency tones that can be more annoying and perceptible than other sources of urban noise [9]. Also, drones often produce pure tones, which are sounds at specific frequencies that can be particularly irritating. These tones are typically generated by the propeller blades and the motors. This issue is exacerbated in densely populated areas where the ambient noise levels are relatively low, making drone noise more prominent and bothersome. Therefore, addressing the noise impact of drones is crucial for the successful implementation and social acceptance of UAM. The noise generated by low-altitude UAS operations, especially in densely populated areas, can disrupt the tranquility of urban environments and potentially disturb residents and wildlife [9]. Urban areas are particularly sensitive to noise pollution, and UAVs contribute significantly to this problem due to their proximity to the ground and the density of human activity. Studies have shown that UAV noise can be more disturbing than ground vehicle noise because of its unique acoustic signature, characterised by higher frequency tones and irregular sound patterns. Research has indicated that prolonged exposure to UAV noise can lead to increased stress levels, sleep disturbances, and adverse effects on cardiovascular health [9]. These findings underscore the importance of developing UAV path planning strategies that minimise noise impact, particularly in densely populated areas [10, 11]. To address these concerns, it is essential to implement comprehensive and environmentally conscious low-noise flight strategies for UAVs.

This paper is a follow-up to our previous work on noise-aware UAV path planning presented at the 14th EASN International Conference [12]. Such work briefly introduced the noise-aware UAV urban path planning framework (along with the validation results), which is comprehensively presented in this work. Moreover, this paper gives comprehensive insights on the design and implementation of the path planning framework, including environment mapping, noise source modelling, noise propagation, reinforcement learning framework development, model training, model validation and demonstration of model flexibility by training two different models for high- and low-priority drone delivery scenarios. This paper aims therefore to inform future research on context-aware dynamic UAV navigation solutions based on reinforcement learning. By embracing sustainability practices in UAS operations management, we can pave the way for a future where smart cities thrive harmoniously with their surroundings, fostering a greener, quieter and more sustainable urban landscape. In this sense, this work enables context-aware intelligent path planning for reduced noise footprint of the UAV operation towards a seamless integration of the drone technology in society. The contributions of this paper are related to SDG 3,11, and 13 [13] as low-noise intelligent UAV path planning can foster sustainability, regulatory compliance and safety of UAM applications. The proposed approach represents a step towards sophisticated progress in terms of public well-being, safety, environmental and social sustainability related to the use of the drone technology in populated urban environments. The original contributions of this work are summarised as follows: (i) conceptualisation of a drone noise footprint evaluation framework for informing noise-aware UAV operations in urban air mobility context; and (ii) development and simulative validation of a low-noise UAV urban path planning solution based on environment mapping, acoustic ray tracing and reinforcement learning.

This paper is organised as follows. A literature review of path planning-based drone noise mitigation approaches is illustrated in Section 1.1. The proposed low-noise UAV path planning framework is presented in Section 2 along with the implementation details. This framework includes environment mapping, acoustic ray tracing and reinforcement learning. Section 3 presents the simulation results that validate the model. Our conclusions and future research directions are drawn in Section 4.

1.1 Related work

Advanced path planning algorithms can minimise UAV noise impact by avoiding densely populated areas and sensitive zones. Although urban locations create limits due to impediments and airspace

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

regulations, operational tactics including raising altitude help reduce ground-level noise through atmospheric absorption and sound dispersion [14]. Noise cost functions can be incorporated into path planning algorithms, particularly heuristic and optimisation-based ones, to steer clear of sensitive or crowded regions, such as hospitals and schools [11, 15]. For high-altitude aircraft, there are two primary approaches: reducing noise during crucial stages (such as takeoff and landing), and for low-altitude UAVs, optimising the entire route [14]. For the latter to avoid areas that are sensitive to noise, a thorough understanding of land use is necessary [15, 16]. However, in places like industrial parks or universities, silent zone sites could be changeable, making real-time planning difficult [16]. Numerous research use noise-map or grid-based methodologies. For example Ref. (11), applies the A* algorithm for noise-aware routing and overlays a grid to identify barriers and quiet zones, while Ref. (15) simulates noise over land use data using the 'iNoise' software. Similarly Ref. (16), diverts UAVs from crowded paths by using agent-based modeling. Other efforts incorporate sophisticated noise models; for example Ref. (17), embeds a noise objective in A* to balance barriers, distance and noise by combining Gaussian beam tracing (GBT) and heuristic planning. Additionally, machine learning techniques are used: Ref. (18) optimises departure paths by employing a soft actor-critic (SAC) reinforcement learning framework with topological and sociological data, exposing trade-offs between noise reduction and fuel efficiency. In Ref. (19), a vehicle routing problem (VRP)-like formulation is expoloited to reduce en-route drone noise impact. This paper introduced a novel hybrid cost function that combines the assessment of the accumulated noise exposure level and path length. Accurate modeling of sound sources and realistic sound propagation computation in urban environments is based on flight simulation. In Ref. (20), high-fidelity simulations are used to assess full-scale UAVs in urban settings by fusing aeroacoustics methods with rotor and fuselage noise models. Heuristic approaches adjust flight trajectories to account for far-field noise propagation, which is supported by a Gaussian beam method. Simpler models, such as the spherical point-source assumption in Ref. (21), permit effective noise estimation. Using simulated annealing, the study changes UAV trajectories to reduce exposure in residential zones while considering energy use. In Ref. (22), a power-aware method is introduced, where a mixed integer linear programming (MILP) model with dynamic programming combines energy and noise limitations by alternating between generator and battery modes in places that are sensitive to noise. Lastly, authors in Ref. (23) present a cost model that takes into account the dangers of property damage, noise and fatalities. The study uniquely considers the correlation between noise level and cost, setting a threshold at 40 dB to determine when noise impact becomes significant in path planning. The approach is based on the fusion of the estimation of distribution algorithm (EDA) and CostA* algorithm. Using a hybrid optimisation strategy over a noise grid, multi-objective path planning with 3D non-uniform rational B-splines (NURBS)-based trajectories in Ref. (24) takes obstacle avoidance, energy consumption and urban noise into account. The problem of minimising the UAV noise impact near the intersection points of multiple flight trajectories is addressed in Ref. (25) by means of a multi-agent Markov decision process framework solved by the Monte Carlo tree search algorithm. The simplified real-time noise prediction method improves scalability and makes it a valuable decentralised intelligent navigation framework. A contribution to assist low-noise flight planning and noise certification has been proposed in Ref. (26). From virtual flights, this work evaluates the noise footprint of unmanned aircraft system (UAS) operations using a GBT method implemented in an in-house solver Environmental Acoustic Ray Tracing Code.

2.0 Low-noise UAV path planning framework

2.1 Environment mapping

The city of Turin, Italy, is an ideal study area for evaluating UAV noise impact and path planning due to several compelling reasons. The city's dense urban core, surrounded by various residential, commercial and industrial zones, presents a realistic scenario for testing UAV operations in different contexts. Additionally, Turin's diverse population density, with both densely populated neighbourhoods

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

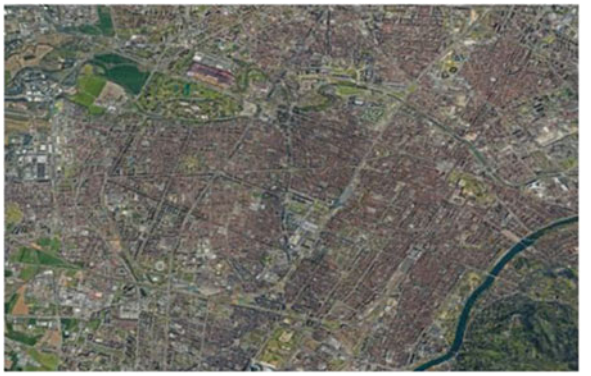


Figure 1. *Satellite photo of the operational area (Torino, Italy).*

and sparsely inhabited areas, allows for a comprehensive analysis of UAV noise impact across different demographic settings. The architectural landscape of Turin, characterised by a mix of historical buildings, modern skyscrapers and extensive residential complexes, provides a varied environment for testing the noise model, as noise propagation and reflection characteristics differ with building types and materials. Furthermore, Turin's involvement in smart city initiatives and sustainable urban development projects, coupled with its regulatory support for innovative transportation solutions, makes it a forward-thinking city likely to adopt advanced UAV technologies. These unique characteristics of Turin, combined with its rich data resources and proactive regulatory climate, make it a highly suitable and relevant study area for developing and validating a novel UAV path planning approach aimed at minimising noise impact. The insights gained from this research in Turin can be generalised to other urban areas with similar characteristics, contributing to broader advancements in UAV noise mitigation and urban air mobility. The satellite photo of the study area is depicted in Fig. 1.

To construct the 3D occupancy grid for the study area, building location data was sourced from OpenStreetMap (OSM). The area of interest was defined by the longitude range from 7.6039E to 7.71E and the latitude range from 45.0386N to 45.1094N. Using the Overpass API, building data within these coordinates was downloaded. Once the data was acquired, the geographic coordinates (longitude and latitude) of the buildings were converted into Cartesian coordinates. This step was essential to preserve the relative positions of the buildings accurately on the map, ensuring that the spatial relationships among buildings were maintained. A random height between 12 m and 18 m was assigned to each building, reflecting typical urban building heights in the region. The map was then trimmed to include only a specific portion with the remaining section discarded. This selection ensured that the resulting maps would be uniform in size, specifically 200 $\times$ 200 cells, which is crucial for consistency during the training phase. Each cell has a length of 2 m. The entire study area was then represented by a three-dimensional matrix, denoted as R. Each grid point $R(i, j, k)$ in this matrix represents an occupancy value, indicating whether the corresponding geographical volume is occupied (value 1) or unoccupied (value 0). Each building's longitude and latitude were converted to (x, y) coordinates which were then mapped to the corresponding (i, j) indices in the matrix R. The height of each building was used to determine the k index, representing the vertical dimension. This ensured that the 3D representation accurately reflected the spatial structure of the urban environment. The resulting map is shown in Fig. 2(a). To facilitate efficient path planning and simulation, the large occupancy grid was subdivided into smaller chunks of 200 $\times$ 200 $\times$ 20 cells, resulting in 456 smaller maps. The uniform size of these chunks ensured consistency in the path planning process.

To accurately represent population distribution within the study area, population density data for Italy was sourced from Meta. This data was filtered to include only the regions within the longitude range

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

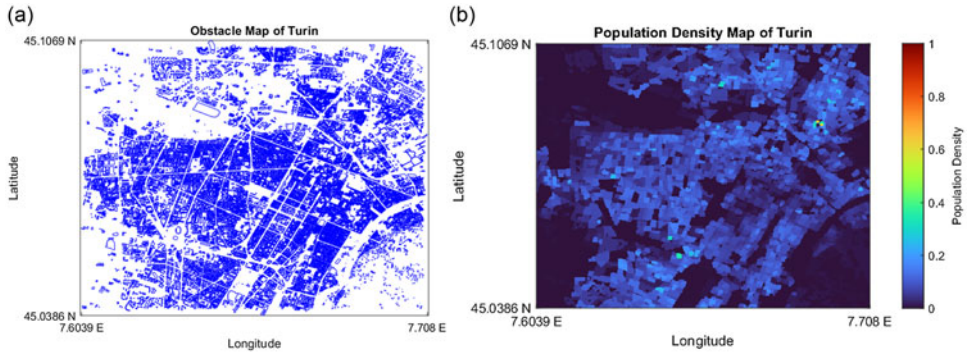


Figure 2. *(a) Obstacle map of the operational area. (b) Population density map of the operational area.*

from 7.6039 E to 7.71 E and the latitude range from 45.0386 N to 45.1094 N, ensuring alignment with the area covered by the obstacles map. The dataset provides the number of individuals in each cell, where each square cell measures 1 arc second, corresponding to approximately 30.87 m at the equator. As latitude increases, the cell width decreases, necessitating adjustments in cell size representation. The dataset includes the longitude and latitude coordinates of each cell's centre along with the corresponding population count. To process this data, the vertices of each cell were calculated from the centre coordinates based on their latitude, determining the exact boundaries of each cell. The population count assigned to each cell was uniformly distributed across all points within the cell. These geographic coordinates were then converted to Cartesian coordinates to match the coordinate system used for the obstacles map. A 2D grid, mirroring the obstacles map was constructed to represent the study area. Population density values were assigned to each point within this grid based on the processed population data. The resulting population density matrix was then normalised to ensure data consistency across the grid. The map is illustrated in Fig. 2(b). This subdivision ensured uniformity in map sizes and accurate overlap with the obstacles data which is necessary for training the model.

2.2 Noise source and acoustic ray tracing

For this study, a point-source spherical propagation model for the sound generated by the UAV was utilised. This model simplifies the calculations for computational efficiency within a simulation-based optimisation framework while remaining reasonably representative of established findings [21]. Although multi-rotor UAVs produce highly directional noise in their immediate vicinity, it's assumed that the noise propagates isotropically at the distances considered in this study. Consequently, as the UAVs traverse their routes, the noise pattern generated on the ground does not distort when turning.

Typically, a constant noise source is assumed in such models; however, the velocity of the drone is correlated with its noise level, and since the RL agent can adapt its velocity, it is beneficial to incorporate this variable into the noise model. To achieve this, data from the study in Ref. (27) was utilised that recorded noise levels at a fixed distance of 1 m from the 'DJI Inspire 2' quadrotor at speeds of 5, 10 and 20 km/h. Measurements were taken across various frequencies, ranging from 20 Hz to 20 kHz, using third-octave bands. For each velocity, the mean SPL across all frequencies was computed to obtain the overall SPL.

Once these three data points were established, an exponential curve was fit to the data, resulting in the SPL-velocity relationship. This results in $L_0 = 60.24 \cdot \exp(0.003379 \cdot v)$, which allows dynamic noise level adjustment based on UAS velocity during flight. L_0 denotes the source SPL in dB, and v denotes the velocity in km/h.

The apparently overly simplified drone's noise model introduced by the relation between L_0 and v can be re-addressed easily in the future by interpolating data resulting from extensive experimental results,

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

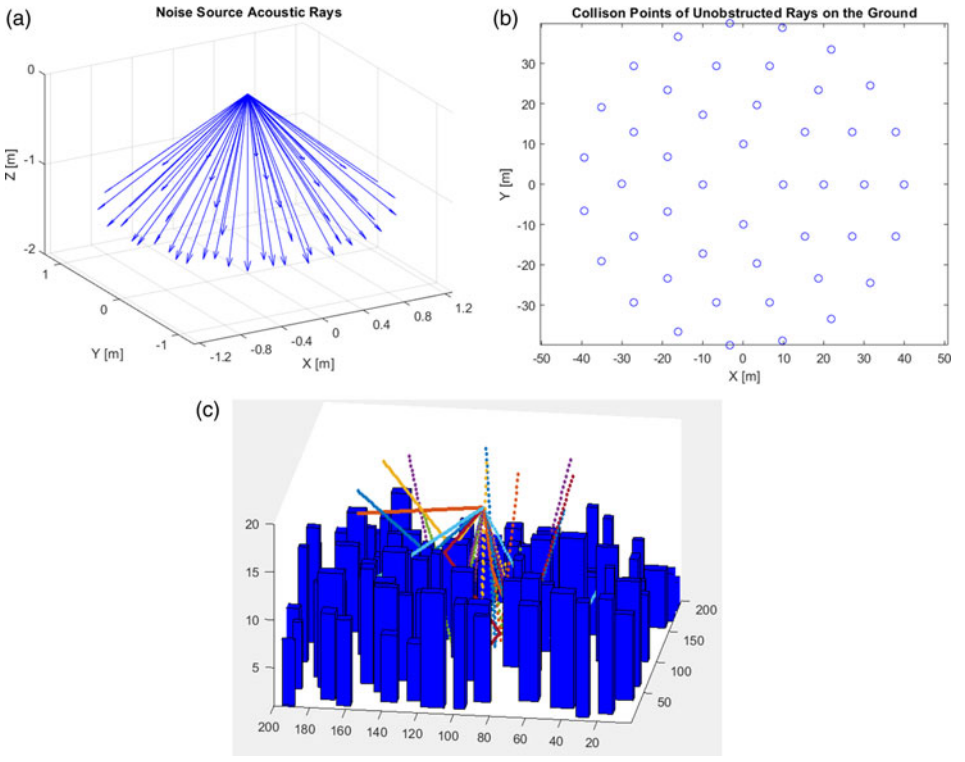


Figure 3. *The noise source: (a) the acoustic rays emitted from the source at an altitude of 40 m; (b) collision points of the unobstructed rays on the ground; (c) synthetic representation of the scenario with buildings and the interaction of the acoustic rays with the environment.*

and/or introducing new parameters and functions for assessing the impact of drone noise on people based on the drone's characteristics as well as the operating conditions [28]. The latter is a cutting-edge emerging research topic that is currently being studied by the psychoacoustic UAS scientific community. Also, a monotonic increase of the emitted noise for increasing values of drone's velocity may not be a high-fidelity noise emission model, and research is currently being carried out about it. On the other hand, it can be seen more generally as a way of accounting for the increase of parasite power consumption with respect to the drone's operational speed.

In the context of UAV path planning aimed at minimising noise impact, accurately modelling noise propagation is critical. This process involves generating direction vectors from a noise source, which are utilised in ray tracing simulations. The following method was employed to generate these rays using. First, the altitude of the noise source was defined at 40 m. A radius of 10 m was set for generating points within concentric circles around the source at ground level. Four concentric circles were set around the source, each with a radius incrementally larger by 10 m, and each populated with an increasing number of points: 4, 9, 14 and 19 points, respectively, distributed equally on the circle resulting in a total of 46 rays. Subsequently, direction vectors from the source point in the sky to each ground point were computed and normalised to unit vectors. This approach ensured a consistent direction magnitude for the vectors. By employing this method, the noise source is effectively modelled. The generated vectors were visualised in a 3D space to verify accuracy as illustrated in Fig. 3.

Accurately modelling the propagation of UAV noise is essential for minimising its impact on the environment. Acoustic ray tracing is a computational technique used to model the propagation of waves through a medium, interaction with obstacles and attenuation over distance. This section details the ray tracing algorithm employed to evaluate the SPL at ground level due to emitted UAV noise. This method

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

involves tracing the path of sound rays emitted from the noise source and calculating their interactions with the environment, considering reflections and atmospheric absorption. The following assumptions are made in the ray tracing model: medium homogeneity, linear propagation, specular reflection, negligible diffraction. The ray tracing algorithm used to evaluate the SPL at ground level involves several key steps: initialisation, propagation, interaction, attenuation and SPL computation.

The algorithm initialises several arrays to store data for each ray, including travelled distance, coordinates along the ray path, attenuation due to geometric divergence and atmospheric absorption and SPL values along the ray path. The source noise magnitude emitted by the UAV is calculated based on its velocity. At the first iteration of a new episode, the minimum noise corresponding to the UAV's minimum movement ([0,0]) is considered. The velocity, v, is computed as $v = ||dx, dy|| \,/T$, where $||dx, dy||$ represent the Euclidean norm of the movement vector, and T is the sample time.

For each ray, the initial position is set at the UAV's current coordinates (x, y, z). The direction of each ray is initialised using pre-calculated direction vectors. The algorithm propagates the ray through the medium by iteratively updating the ray's position, x_t, y, z_t, based on the direction, x_{dir}, y_{dir}, z_{dir}, and step length, α: $x_{t+1} = x_t + x_{dir} \cdot \alpha$, $y_{t+1} = y_t + y_{dir} \cdot \alpha$, $z_{t+1} = z_t + z_{dir} \cdot \alpha$. The updated coordinates are constrained within the map boundaries to ensure valid indices. The step length was selected to be 0.3 units. This value was determined through iterative ray tracing simulations and visualisation of sound propagation to ensure model fidelity. A larger step length resulted in rays failing to detect obstacles, while a smaller step length significantly increased computational load. This step length struck a balance between accuracy and computational time.

It's necessary to check for interactions with environmental elements (e.g., ground, buildings). If the ray reaches the map boundary (excluding the ground), the ray tracing for that ray is terminated as it is leaving the area of interest and will no longer have any effect on it. If the ray hits an obstacle, the reflection is handled using the law of reflection, as follows: $\boldsymbol{R} = \boldsymbol{I} - \mathbf{2}(\boldsymbol{I} \cdot \boldsymbol{N}) \cdot \boldsymbol{N}$. The updated direction vector of the reflected acoustic ray is expressed as $(x_{ref} = x_{dir} - 2\,(x_{dir} \cdot N) \cdot N, y_{ref} = y_{dir} - 2\,(y_{dir} \cdot N) \cdot N$, $z_{ref} = z_{dir} - 2\,(z_{dir} \cdot N) \cdot N)$.

The SPL loss calculation as described here aligns with the methods prescribed by the ISO 9613-2 standard, which provide guidelines for predicting the attenuation of sound during propagation outdoors. It considers various factors such as geometric divergence, atmospheric absorption, ground effects and barriers. The SPL loss is calculated for each segment of the ray path due to atmospheric absorption and the geometric divergence.

The attenuation from geometric divergence, A_{div}, accounts for the reduced sound level as the sound propagates over a distance. For a point sound source in a free field, the attenuation in decibels (dB) is given by $A_{div} = 20 \log (d/d_0) + 11$. The attenuation from atmospheric absorption A_{atm} during propagation over a distance d is given in (dB) by $A_{atm} = \alpha d/1000$. d is the distance travelled from source to receiver in (m), $d0$ is the reference distance (= 1 m); and α is the coefficient of atmospheric attenuation in (dB/km) at the nominal median frequency of each third-octave band. Given a temperature of 10 degrees Celsius, relative humidity (RH) of 70%, and assuming a nominal median frequency of 200 Hz (considering drones typically operate between 100 and 300 Hz with an average of 200 Hz), the atmospheric absorption coefficient (α) is set to be 0.76 dB/km as per ISO 9613-2 [29].

The walls are assumed to be solid with no windows, and as per ISO 9,613, they are characterised by a coefficient of 1. This indicates that these walls act as perfect reflectors, meaning they do not allow sound to transmit through them. Consequently, there is no additional attenuation from these walls beyond the reflections they cause. The SPL for each ray is recorded when the ray reaches within a height 2 m above the ground. This is done based on the equation: $L_i = L_0 - A_{div} - A_{atm}$, where L_i is the SPL for each ray in (dB). Once the SPL for each ray is known, the contributions from all N rays are summed to determine the overall SPL using the following equation: $L_{tot} = 10 \log \left(\sum_i^N 10^{\mathrm{Li}/10}\right)$, with L_{tot} denoting the noise impact at the ground level in (dB).

Figure 4 shows the snapshot of the pseudo-code of the ray tracing algorithm proposed in this work.

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

Input: UAV current coordinates (x, y, z); noise rays direction vectors x_{dir}, y_{dir}, z_{dir}; action taken (dx, dy)
Initialize arrays to store distance travelled and attenuations
If first iteration then
  $v = 0$
else
  $v = \frac{||dx,dy||}{T}$
end if
Evaluate the source SPL: $L_0 = 60.24 \times \exp(0.003379 \times v)$
for *ray* = *1, N* do
  while $d < 200$ do
    $x_{t+1} = x_t + x_{dir} \times \alpha$
    $y_{t+1} = y_t + y_{dir} \times \alpha$
    $z_{t+1} = z_t + z_{dir} \times \alpha$
    If *ray leaves map boundary*, then
      Break
    end if
    If *ray collides with obstacles*, then
    $x_{ref} = x_{dir} - 2 \cdot (x_{dir} \cdot N) \cdot N$
    $y_{ref} = y_{dir} - 2 \cdot (y_{dir} \cdot N) \cdot N$
    $z_{ref} = z_{dir} - 2 \cdot (z_{dir} \cdot N) \cdot N$
    Update the direction vectors: $x_{dir} \leftarrow x_{ref}$, $y_{dir} \leftarrow y_{ref}$, $z_{dir} \leftarrow z_{ref}$
    end if
    Evaluate distance travelled: $distance = ||[x_t, y_t, z_t] - [x_{t+1}, y_{t+1}, z_{t+1}]||$
    Update the distance travelled: $d \leftarrow d + distance$
    If $z_{t+1} <= 2$, then
    Evaluate geometric divergence attenuation: $A_{div} = 20 \log\left(\frac{d}{d_0}\right) + 11$
    Evaluate atmospheric absorption attenuation: $A_{atm} = \frac{\alpha d}{1000}$
    Evaluate the SPL: $L_i = L_0 - A_{div} - A_{atm}$
    end if
  end while
end for
Evaluate the overall SPL: $L_{tot} = 10\, log(\sum_i^N 10^{(\frac{L_i}{10})})$

Figure 4. *Snapshot of the pseudo-code of the Ray Tracing algorithm proposed in this work.*

2.3 Deep deterministic policy gradient algorithm

The deep deterministic policy gradient (DDPG) algorithm [30], is a model-free, off-policy RL method. It is an actor-critic RL agent designed to concurrently learn a Q-function and a policy, leveraging off-policy data and the Bellman equation to optimise both. In RL, policy gradient methods are used to find the optimal policy by optimising a parametric policy π_θ. Traditional policy gradient methods use stochastic policies, where the action is sampled from a probability distribution. Deterministic policy gradient (DPG) methods, however, use a deterministic policy $\pi_\theta(s)$ that directly maps states to actions. DDPG agents are suitable for environments with continuous or discrete observation spaces and continuous action spaces. DDPG agents utilise two main components:

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

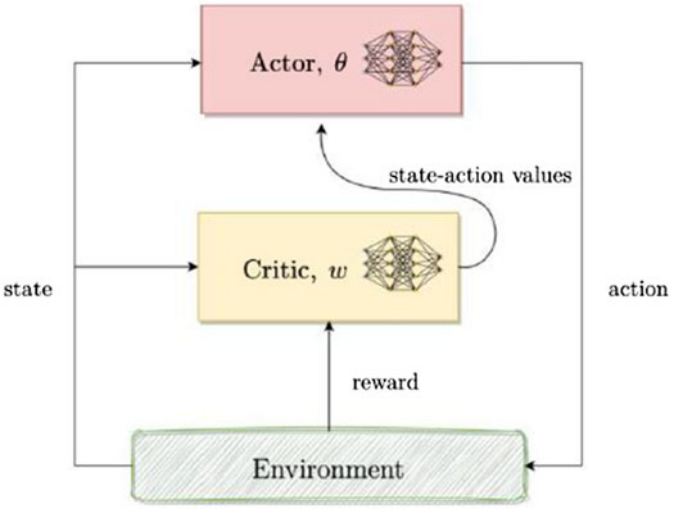


Figure 5. *Actor-critic agent: representation taken from Ref. (18).*

- **Actor**: The actor represents a deterministic policy $\pi(s)$. A deterministic actor takes an environment observation as input and returns as output an action that is a parametrised deterministic function of the observation, thereby implementing a parametrised deterministic policy.
- **Critic**: The critic evaluates the action-value function $Q(S, A)$ which is a mapping from an environment observation-action pair to the value of a policy.

During training, DDPG agents update both the actor and critic properties at each time step. They store past experiences in a circular experience buffer and update the actor and critic using mini-batches of experiences randomly sampled from this buffer. To ensure stable behaviour, the replay buffer must be sufficiently large to encompass a diverse range of experiences. However, it is not always beneficial to retain all past experiences. Using only the most recent data can lead to overfitting, while using an excessive amount of data can slow down the learning process. Therefore, careful tuning is required to strike the right balance. DDPG trains a deterministic policy using an off-policy approach. DDPG trains a deterministic policy using an off-policy approach (Fig. 5).

Since the policy is deterministic, on-policy exploration at the start may not cover a sufficiently diverse range of actions to obtain useful learning signals. To enhance exploration, noise is added to actions during training using the Ornstein-Uhlenbeck noise model. The noise scale can be reduced over the course of training using a decaying factor to balance the exploration-exploitation tradeoff. The noise value $v(t)$ at each time step t is updated as $v\,(t+1) = v\,(t) + Mean_{ac} \cdot (Mean - v\,(t)) \cdot T + std\,(t) \cdot \mathcal{N}\,(0,1) \cdot \sqrt{T}$, the standard deviation decays at each time step as $std\,(t+1) = std\,(t) \cdot (1 - DecR)$. It is crucial to set the noise standard deviation appropriately to promote exploration, typically setting $std \cdot \sqrt{T}$ between 1% and 10% of the action range. DDPG agents maintain four function approximators to estimate the policy and value functions:

- **Actor** $\boldsymbol{\pi}\,(\boldsymbol{S};\boldsymbol{\theta})$: Takes observation S as input and returns the corresponding action that maximises the long-term reward.
- **Target Actor** $\boldsymbol{\pi}_t\,(\boldsymbol{S};\boldsymbol{\theta t})$: Periodically updated using the latest actor parameters θ to improve optimisation stability. The actor policy is updated using the sampled policy as follows:

$$\frac{1}{M}\sum_{i=1}^{M} \nabla_a Q\,(s,a;\phi)\,\big|_{s=s_1,\,a=\pi(s_i)} \nabla_b \pi\,(s;\theta)\big|_{s_i} \tag{1}$$

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

Randomly initialize critic network $Q(s,a;\phi)$ and actor network $\pi(s;\theta)$ with weights ϕ and θ respectively
Initialize critic and actor target networks $Q_t(s,a;\phi_t)$ and $\pi_t(s;\theta_t)$ with weights ϕ and θ respectively
Initialize replay buffer R
for *episode =1, K* **do**
  Initialize a random noise N for exploration
  Receive initial observation state s_1
  for t =1, T **do**
    Select action $a_t = \pi(S_t;\theta) + N_t$ based on the current policy and exploration noise
    Execute action a_t and observe reward r_t and new state s_{t+1}
    Store transition (s_t, a_t, r_t, s_{t+1}) in R
    Sample a mini batch of M transitions (s_i, a_i, r_i, s_{i+1}) from R
    Set $y_i = r_i + \gamma\, Q_t(s_{i+1},\ \pi_t(s_{i+1};\ \theta_t);\ \phi_t)$
    Update critic by minimizing the loss: $L = \frac{1}{M}\sum_{i=1}^{M}(y_i - Q(s_i,a_i;\phi))^2$
    Update the actor policy using the sampled policy gradient:

$$\nabla_\theta J \approx \frac{1}{M}\sum_{i=1}^{M} \nabla_a Q(s,a;\phi)|_{s=s_i,a=\pi(s_i)}\ \nabla_\theta\, \pi(s;\theta)|s_i$$

    Update the target networks:

$$\phi_t \leftarrow \tau\phi + (1-\tau)\phi_t$$
$$\theta_t \leftarrow \tau\theta + (1-\tau)\theta_t$$

  end for
end for

Figure 6. *Snapshot of the pseudo-code of the DDPG strategy adopted in this work.*

- **Critic** $\boldsymbol{Q}$ $(\boldsymbol{S},\boldsymbol{a};\boldsymbol{\phi})$: Takes observation S and action a as inputs and returns the expected long-term reward.
- **Target Critic** $\boldsymbol{Q}_t$ $(\boldsymbol{S},\boldsymbol{a};\boldsymbol{\phi t})$: Periodically updated using the latest critic parameters θ to maintain stability. The update is defined as follows:

$$\frac{1}{M}\sum_{i=1}^{M}\left[y_i - Q\,(s_i, a_i;\phi)\right]^2 \tag{2}$$

Both the target and main networks (actors and critics) share the same structure and parameterisation. Q-learning algorithms utilise target networks to stabilise training. The expression $r + \gamma\,(1-d)\max Q_t\,(s',a')$ is referred to as the target, as minimising the mean squared Bellman error (MSBE) loss aims to align the Q-function with this target. d represents a binary indicator variable that signifies whether the next state s' is terminal or not. Specifically, d=1 if s' is a terminal state and d=0 otherwise. However, the target relies on the same parameters ϕ that are being optimised, leading to instability in MSBE minimisation. To address this, a secondary network called the target network, with parameters ϕ_t is introduced. This target network lags behind the main network, providing more stable target values. DDPG algorithms update the target network more smoothly using Polyak averaging: $\phi\tau + \phi_t\,(1-\tau)$, where τ is a hyperparameter between 0 and 1, referred to as smoothing factor. This continuous update method helps maintain stability during training.

A snapshot of the pseudo-code of the proposed DDPG algorithm is illustrated in Fig. 6.

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

2.4 Implementation details

The RL-based low-noise path planning framework proposed in this work requires the definition of UAV state and UAV action space as well as the agent's reward function. The implementation details are given as follows. The simulation settings are also presented.

2.4.1 State and action space

The input layer's neuron count is determined by the dimensionality of the state representation. This state representation encapsulates the current condition of the environment, providing a comprehensive snapshot that the RL uses to make informed decisions. The output layer, on the other hand, corresponds to the action space. The number of neurons in this layer should align with the range of possible actions that the agent can take in the environment. The construction and significance of the state representation, which is a critical component of the RL approach is described here. The state representation is a 24-dimensional vector, with the features corresponding to the input layer of the neural network. The vector encapsulates a comprehensive set of features that characterise the current state of the drone and its environment:

- **Drone Position**: The cartesian coordinates of the drone in XY plane.
- **Target Position**: The cartesian coordinates of the target in XY plane.
- **Distance to Target**: The Euclidean distance from the drone to the target.
- **Target Orientation**: The orientation of the target with respect to the drone, represented as an angle in the XY plane.
- **Local Population Density**: The population density at the current drone location.
- **Mean Population Density**: The mean population density across two sub grids of sizes 15×15 and 51×51, with the drone at the centre of each.
- **Orientation of Maximum Population Density**: The orientation of the point of maximum population density in the sub grid with respect to the drone.
- **Noise Level**: The overall SPL value on the ground.
- **Obstacle Distance**: The distances from the drone to the three closest and three furthest obstacles.
- **Obstacle Orientation**: The orientations of the six obstacles with respect to the drone.

To ensure stability and prevent divergence during the training process, each of these values is normalised to fall within the range [0, 1]. When training a RL agent using techniques like policy gradients or Q-learning, the scale of the input features can affect the scale of the gradients. If the input features have vastly different scales, it can lead to imbalanced gradients and slower convergence. Normalising the input features ensures that the gradients are scaled appropriately, allowing for more stable and efficient learning. Normalising input features helps in improving the generalisation capability of the RL agent. By scaling the features to a similar range, the agent can learn a more generalised policy that performs well across different states or observations.

The agent operates within a fixed height airspace and possesses the capability to navigate in any direction across the continuous space of the XY plane. Such an environment presents unique challenges in terms of decision-making and path planning. To effectively train the agent in this dynamic setting, a critical consideration is the selection of an appropriate sampling time, T, which dictates the frequency at which the agent takes actions. The sampling time chosen for the training phase is 0.3 units. This selection is deliberate, as a smaller T implies that the agent must make decisions more frequently. This characteristic proves advantageous in environments characterised by high dynamism or instances where the drone's situational awareness is of paramount importance. However, it's essential to acknowledge that reducing T comes with a computational cost, as the agent must process information and make decisions more rapidly. Thus, there exists a trade-off between the frequency of actions and computational efficiency. The value of T influence both responsiveness and computational feasibility. To further enhance the realism

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

of the simulation, it's imperative to appropriately define the range of movement available to the agent at each time step. This range directly influences the agent's velocity and, by extension, its navigational capabilities. To align the simulation with real-world parameters, the range of movement is calibrated to match the velocity of commercial drones, which typically have a maximum speed of 100 km/h. By setting the action range $[dx, dy]$ for both the x and y directions to [3,3], corresponding to a movement of 6 meters per time step, the simulated drone achieves a maximum velocity of approximately 101 km/h. This adjustment ensures that the agent's movement closely mirrors the operational constraints and capabilities of real-world drones, thereby enhancing the fidelity and applicability of the simulation outcomes.

2.4.2 Reward function

The reward function plays a pivotal role in achieving the desired behaviour from the model. At any timestep, if the episode terminates (the agent reaches the target), a reward of 2 is given. Otherwise, the reward is calculated by the weighted sum of the following terms, which are all individually normalised in the range of [0,−1]:

1. **Idle penalty $P_1(t)$**: penalises the drone for remaining stationary. A penalty of −1 is applied if the drone's position has not changed in between timesteps.

$$P_1(t) = \begin{cases} -1, & if\ p_t = p_{t-1} \\ 0, & otherwise \end{cases} \tag{3}$$

2. **Distance penalty $P_2(t)$**: encourages the drone to reduce the distance to the target. The penalty is based on the change in Euclidean distance to the target.

$$P_2(t) = -\frac{\|p_t - p_{target}\| - \|p_{t-1} - p_{target}\|}{\|p_{max}\|} \tag{4}$$

3. **Population density penalty $P_3(t)$**: discourages navigation through high-density areas. A penalty of −1 is applied if the population density increases from the previous timestep; otherwise, a smaller penalty of −0.1 is applied.

$$P_3(t) = \begin{cases} -1, & if\ Pd_t > Pd_{t-1} \\ -0.1, & otherwise \end{cases} \tag{5}$$

4. **Noise penalty $P_4(t)$**: penalises high noise levels. If the SPL exceeds the threshold (maximum possible SPL on the ground), a penalty of −1 is applied. Below the threshold, the penalty decreases exponentially to 0 as the SPL decreases to 0.

$$P_4(t) = \begin{cases} -1, & if\ L_t > N_{thr} \\ -e^{-3 \cdot \frac{N_{thr} - L_t}{N_{thr}}}, & otherwise \end{cases} \tag{6}$$

5. **Cumulative noise penalty $P_5(t)$**: encourages lower cumulative noise and, indirectly, shorter flight time. The penalty increases based on the cumulative noise up to the current timestep.

$$P_5(t) = -\frac{N_t}{t \cdot N_{MAX}} \tag{7}$$

6. **Smoothness penalty $P_6(t)$**: prevents erratic behaviour and encourages smooth navigation. The penalty is based on the angular difference in the movement vector between consecutive timesteps.

$$P_6(t) = -\frac{|\theta_t - \theta_{t-1}|}{\pi} \tag{8}$$

The total reward at each timestep, $R(t)$, is calculated as $R(t) = \lambda_1 P_1(t) + \lambda_2 P_2(t) + \lambda_3 P_3(t) + \lambda_4 P_4(t) + \lambda_5 P_5(t) + \lambda_6 P_6(t)$, where λ is the scaling factor.

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

2.4.3 Simulation settings

To demonstrate the potential of RL-based navigation, two distinct drone delivery scenarios are considered. Therefore, two versions of the RL model are developed:

- **High-Priority Model (HPM)**: In scenarios where timely delivery is critical, such as delivering medical supplies to a hospital, the HPM emphasises the reduction of the delivery time over noise minimisation, and optimises the flight path primarily for speed and efficiency, accepting higher noise levels as a trade-off.
- **Low-Priority Model (LPM)**: For deliveries that do not require urgency, such as regular postal services, the model focuses on minimising noise impact as much as possible while still ensuring reasonable delivery times. The LPM focuses on optimising flight paths to reduce noise pollution, even if this results in longer delivery times.

For optimisation, the Adam algorithm (adaptive moment estimation) is employed, leveraging adaptive learning rates along with first and second moment estimates to enhance convergence speed and manage sparse gradients effectively. Both networks incorporate a gradient threshold of 1 for gradient clipping, thereby mitigating the exploding gradient problem and ensuring stable updates. Gradient clipping stabilises the training process, particularly at higher learning rates and in the presence of outliers, without typically affecting the accuracy of the learned task. The learning rate, determines the step size for each iteration of gradient descent towards minimising the loss function. The tuned value favours training stability, where higher learning rates might initially accelerate convergence but could result in oscillations or divergence later, and lower learning rates may converge more slowly but with greater robustness. The DDPG agent is then configured, with the first and last layers matching the observation and action space. A smoothing target update method is chosen, and the network is updated at each step. The update rate of the target networks determines how swiftly they track the main networks' parameters. A higher update rate could lead to faster convergence but might introduce instability or oscillations. Conversely, a lower update rate necessitates more iterations for convergence but offers more stability. The discount factor prioritises long-term rewards over immediate ones, encouraging the agent to develop more farsighted policies. The replay buffer size affects the diversity of experiences the agent encounters during training. A larger buffer stores more experiences, potentially leading to better policies through diverse training data. However, excessively large buffers may slow down training and increase memory requirements. The batch size determines the number of experiences sampled from the replay buffer for each update step. The main hyperparameters meticulously tuned to optimise the training process are defined as follows: mini batch size: 128, discount factor γ: 0.99, smoothing factor τ: 0.001, Learning rate α: 0.00001, 3 hidden layers, 256 nodes in each hidden layer. Also, $Mean_{ac} = 0.15$, $Mean = 0$, $std = 0.85$, and $Decay\,Rate = 10^{-4}$. The scaling factors of the reward function are set as $\lambda_1 = 0.02$, $\lambda_2 = 0.07$, $\lambda_3 = 0.43$, $\lambda_4 = 0.01$, $\lambda_5 = 0.47$, $\lambda_6 = 1$. This configuration refers to the HPM. The LPM is defined with a different set of scaling factors, as follows: $\lambda_1 = 0.02$, $\lambda_2 = 0.06$, $\lambda_3 = 0.48$, $\lambda_4 = 0$, $\lambda_5 = 0.44$, $\lambda_6 = 1$.

2.4.4 Model training

To initiate the training process, it is essential to set up the environment and configure both the actor and critic neural networks. The environment encompasses the algorithms previously described and any auxiliary functions necessary for the seamless execution of its operations. The actor and critic networks are each initialised with three hidden layers, each containing 256 nodes. This symmetry ensures consistent parameterisation across both networks. The overall dataset, consisting of 456 obstacle and population density maps, is divided into a training set of 410 maps and a testing set of 46 maps. During the training phase, obstacle and density maps and noise ray direction vectors are imported. At the beginning of each episode, a random obstacle map and its corresponding density map are selected, and the agent's starting position and target coordinates are specified randomly. Training proceeds by executing actions in the environment, observing the resultant state and reward, and continuing until the episode's termination condition is met: either the agent reaches within 2 m of the target or the maximum of 280

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

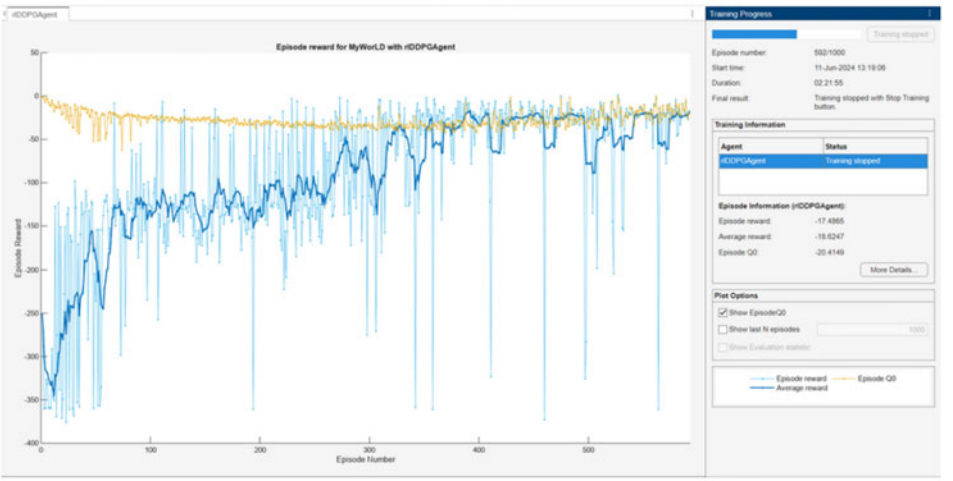


Figure 7. *Snapshot of the learning curve throughout the training process of the HPM.*

steps is reached. A new episode begins with another random map and initial coordinates. Throughout training, learning curves such as cumulative reward per episode, moving average of cumulative reward, and initial Q-value of each episode are monitored. These metrics, along with the agent's success rate, guide the continuation of training until convergence to the optimal policy is achieved. This kind of training process is repeated for both models: HPM and LPM. The evolution of the learning curve throughout the training process of the HPM is illustrated in Fig. 7.

3.0 Simulation results

3.1 Model validation

In order to validate the accuracy of the agent's learned policy, the trained agent is tested on a map different from the testing dataset. To compare its performance, this trajectory is compared with the direct path to the target and the optimal path found from the A* algorithm. The heuristic function in the A* algorithm is configured to account for both distance (40%) and population density (60%). Since the noise model is velocity-dependent, defining a velocity for the alternative methods is essential to evaluate their noise impact accurately. For a fair comparison, the average velocity of the RL path is used as a constant velocity along the entire path computed by the other methods. The trajectories for the test are depicted in Fig. 8. The heatmap of the noise along the trajectory, the total ground noise impact of the path, and the distribution of the SPL along the trajectory are depicted in Figs 9–11, respectively. The performance metrics of the paths are illustrated in Fig. 12.

The comparative analysis reveals that the RL model, despite generating a longer path and higher overall noise levels, effectively minimises SPL per unit length and achieves lower minimum SPL values. The increased path length and flight time associated with the RL model suggest a deliberate trade-off, where the model prioritises noise reduction over directness of the path. The higher maximum SPL observed in the RL model's path indicates occasional peaks in noise, which may be a result of the model's dynamic adjustments to minimise overall noise exposure. The reductions in SPL per unit length achieved by the RL model highlight its potential for applications where minimising noise pollution is critical. These findings suggest that while the RL model may not always produce the shortest or fastest path, it is more effective in reducing noise impact on the ground, which is a key objective for noise-sensitive environments.

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

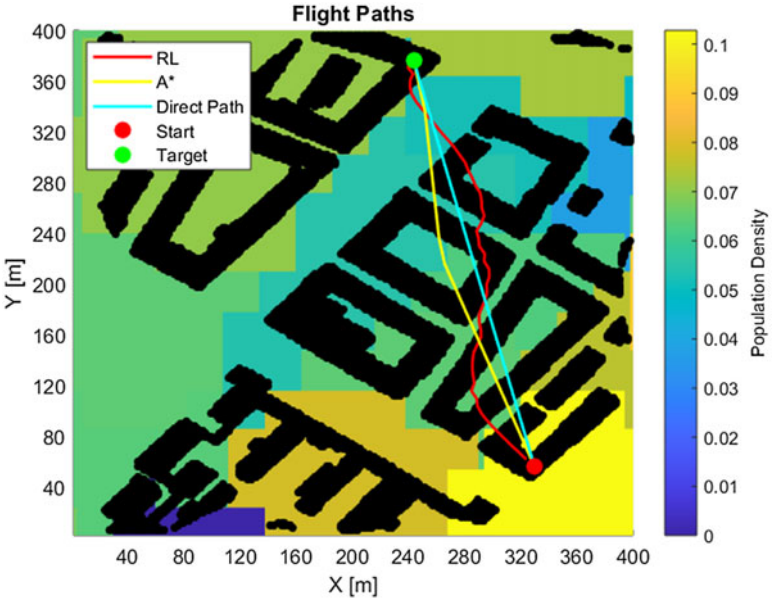


Figure 8. *UAV flight paths: RL path (red line), A* path (yellow line), and direct path (blue line).*

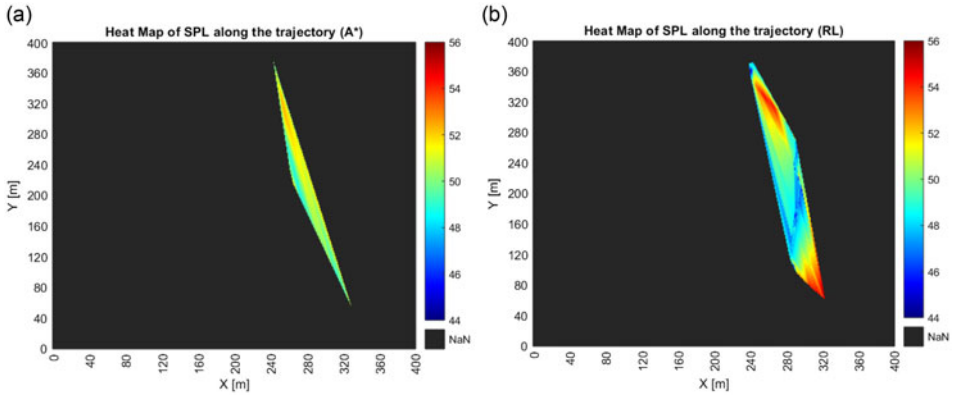


Figure 9. *(a) Noise heatmap along the A* path. (b) Noise heatmap along the RL path.*

To rigorously evaluate the generalisation capability and accuracy of the RL model, extensive testing was conducted on 10 different maps, each featuring random start and target locations. For each map, two separate tests were performed, and the key performance metrics were recorded for the three methods. The statistical analysis, presented as the median and standard deviation (std) for each performance metric, provides a comprehensive view of the model's efficacy across varied scenarios. The statistical results are reported in Fig. 13. Figure 14 shows the complete set of performance metrics related to the validation results.

The RL agent exhibited a median path length of 333.18 m with a standard deviation of 138.8 m. This performance closely mirrors the A* path. In contrast, the Direct path demonstrated a shorter median path length of 295.44 m. The relatively longer path lengths of the RL and A* trajectories indicate a prioritisation of navigational strategies beyond mere distance minimisation, aligning with the goal of noise reduction. The higher variability in the RL path length (std = 138.8 m) suggests an adaptive strategy, capable of tailoring paths to specific environmental conditions and noise constraints. The RL agent

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

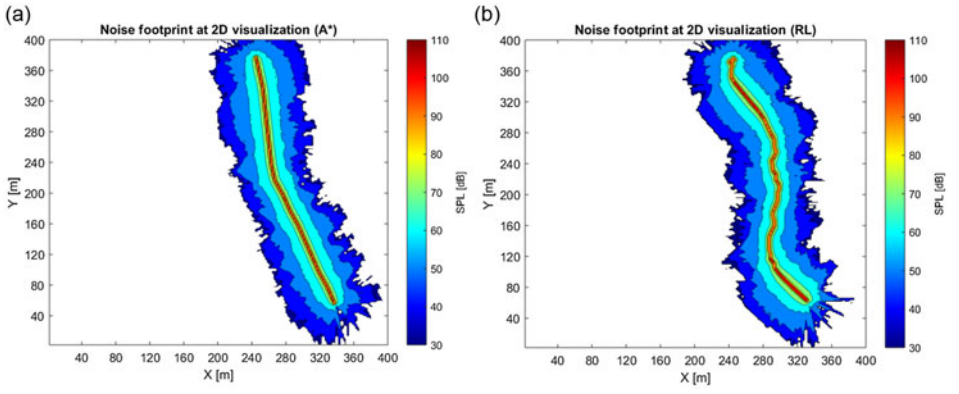


Figure 10. *(a) Noise footprint of the A* path. (b) Noise footprint of the RL path.*

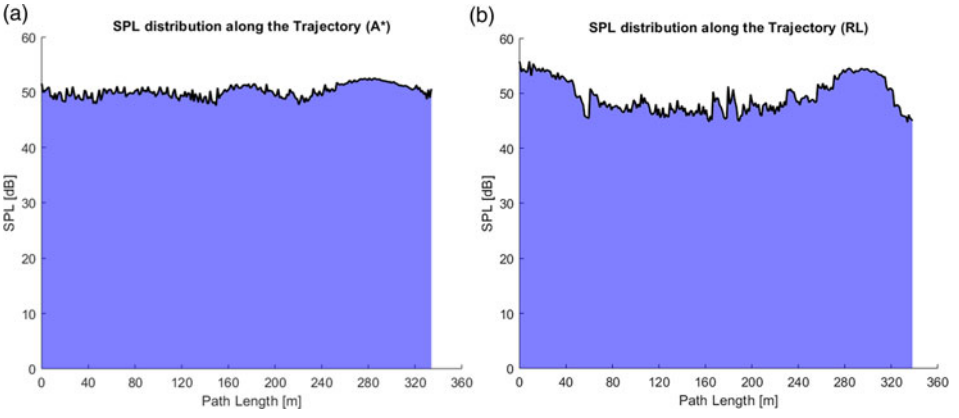


Figure 11. *(a) SPL distribution along the A* path. (b) SPL distribution along the RL path.*

	Path Length [m]	Average Velocity [km/h]	Flight Time [s]	Minimum SPL [dB]	Maximum SPL [dB]	Total SPL [dB]	SPL per unit length [dB]
Direct	331.3548	82.8	14.41	47.4021	52.5382	8288.34	50.027
A*	334.0122	82.8	14.53	47.6016	52.5382	8384.79	50.2065
RL	339.33	82.8	14.76	44.7236	55.819	8364.66	49.301
RL vs. Direct	2.41%	0%	2.41%	– 5.65%	6.24%	0.92%	– 1.45%
RL vs. A*	1.59%	0%	1.59%	– 6.05%	6.24%	– 0.24%	– 1.8%

Figure 12. *Comparison of performance metrics for A*, RL, and direct paths.*

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

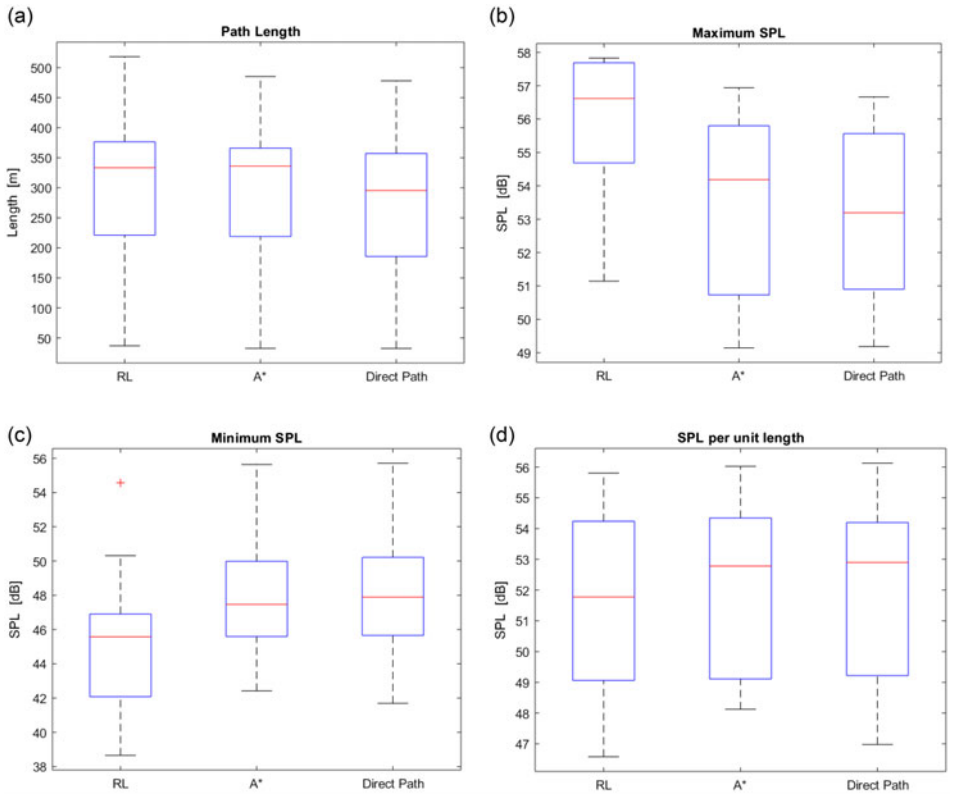


Figure 13. *Box plot for (a) path length, (b) maximum SPL, (c) minimum SPL, (d) SPL per unit length of the three paths (A*, direct, RL): model validation results.*

	Path Length [m]		Minimum SPL [dB]		Maximum SPL [dB]		Total SPL [dB]		SPL per unit length [dB]	
	Median	*Std*	*Median*	*Std*	*Median*	*Std*	*Median*	*Std*	*Median*	*Std*
Direct	295.44	122.1	47.89	3.82	53.19	2.58	14925.4	6446	52.9	2.91
A*	336	128.28	47.47	3.65	54.19	2.7	16588.8	6794.9	52.78	2.73
RL	333.18	138.8	45.57	3.88	56.61	1.91	17625.4	7363.5	51.77	2.97
RL vs. Direct	12.78%	13.68%	− 4.8%	1.68%	6.44%	− 26%	18.09%	14.2%	− 2.1%	2.13%
RL vs. A*	− 0.8%	8.2%	− 4%	6.24%	4.49%	− 29%	6.25%	8.37%	− 1.9%	8.84%

Figure 14. *Comparison of performance metrics for A*, RL, and direct paths: model validation results.*

achieved the lowest median minimum SPL at 45.57 dB, significantly lower than the Direct path and the A* path. The standard deviation for the RL path's minimum SPL was 3.88 dB, slightly higher than the direct (3.82 dB) and A* (3.65 dB) paths. This notable reduction in minimum SPL highlights the RL agent's effectiveness in identifying possibilities to minimise the noise impact in sensitive regions. The RL path's median maximum SPL was 56.61 dB with a standard deviation of 1.91 dB. While this value

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

is slightly higher than the direct and A* paths, the low standard deviation for the RL path suggests a consistent upper bound on noise levels. This consistency is crucial for applications requiring predictable noise management. The slightly higher maximum SPL encountered by the RL path can be attributed to its dynamic velocity adjustments, which allow for brief increases in noise to optimise overall trajectory efficiency. The total SPL for the RL path had a median value of 17625.4 dB with a standard deviation of 7363.5 dB, which is higher than both the A* and the direct paths. The higher total SPL is a consequence of the longer path lengths and the emphasis on reducing noise per unit length rather than total noise. The larger variability in the RL path's total SPL reflects its adaptive nature, adjusting paths to balance noise distribution effectively. The RL path excelled in minimising noise impact per unit distance, achieving a median SPL per unit length of 51.77 dB/m with a standard deviation of 2.97 dB. This performance surpasses both the direct and A* paths. The lower SPL per unit length indicates a more efficient distribution of noise along the trajectory, reducing the noise impact in noise-sensitive areas despite longer overall paths.

The validation results confirm that the RL model demonstrates strong generalisation capability and accuracy across diverse map layouts. The RL agent consistently achieves lower minimum SPL and SPL per unit length, highlighting its proficiency in managing noise effectively. While the RL paths are generally longer and exhibit higher total SPL, the significant reduction in noise per unit length underscores the agent's strategic trade-off between path length and noise distribution. For the sake of clarity, it is hereby specified that these results refer to the HPM.

The ray tracing algorithm provides a comprehensive method for evaluating the SPL at ground level due to UAV noise. By accounting for reflections, atmospheric absorption and geometric divergence, the algorithm accurately models noise propagation. This detailed analysis is crucial for optimising UAV flight paths to minimise noise impact on the environment, ensuring compliance with noise regulations and enhancing community acceptance of UAV operations. Given the validation results presented in the manuscript, it is possible to state that future re-implementation of the SPL noise source generation model can be easily addressed; provided that extensive experimental results have been collected. Currently, such results are not present in the literature as this is an emerging research topic with just a few years of publication record. Also, comprehensive experimental results about the perception of drone noise (which may be added as an additional factor in the optimisation process) based on environmental and operating conditions is not present in the literature. This kind of experimental validation can only increase the fidelity and usefulness of the proposed approach.

3.2 Logistics drone mission

To test the effectiveness of the two models (HPM and LPM), a map was selected, and both versions of the RL model were tested under identical conditions, and the performance metrics were recorded for comparison. The high priority model, with its higher average velocity, is more effective in reducing flight time. However, this comes at the cost of a marginally higher total SPL and SPL per unit length. Conversely, the low priority model, with its longer path and flight time, achieves lower total SPL and SPL per unit length, indicating a more efficient noise management relative to the distance traveled. This approach highlights the adaptability of the RL models in handling different delivery scenarios. Integrating RL models into UAV delivery systems offers a robust solution for achieving flexible and optimised drone operations, tailored to the specific needs of different delivery scenarios. By adapting the models for different delivery priorities, we can ensure that drones can operate efficiently in diverse environments, meeting specific operational goals while adhering to noise regulations and reducing environmental impact. The results from these tests are shown in Figs. 15–19.

4.0 Conclusion

The results of this study underscore the significant potential of RL for noise-aware navigation of UAVs. Through comprehensive testing on both specific and generalised scenarios, the RL agent demonstrated

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

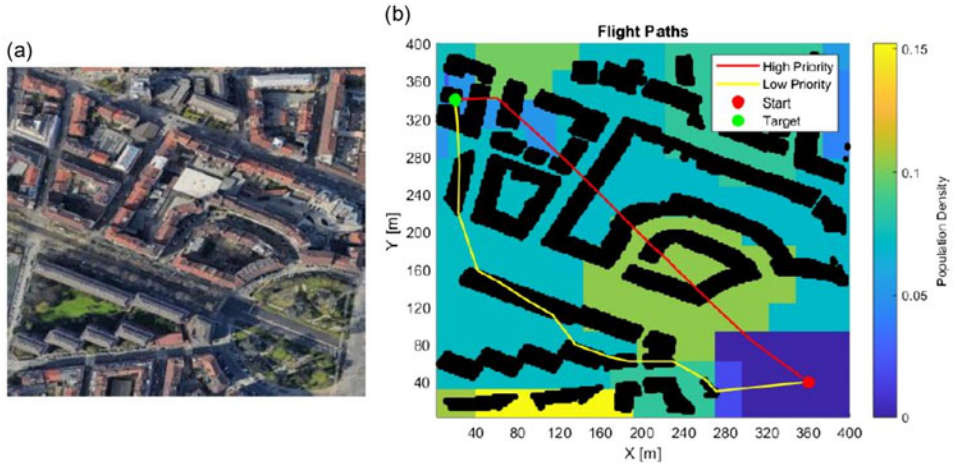


Figure 15. *(a) Satellite photo of the area for testing HPM and LPM. (b) UAV flight paths from HPM and LPM.*

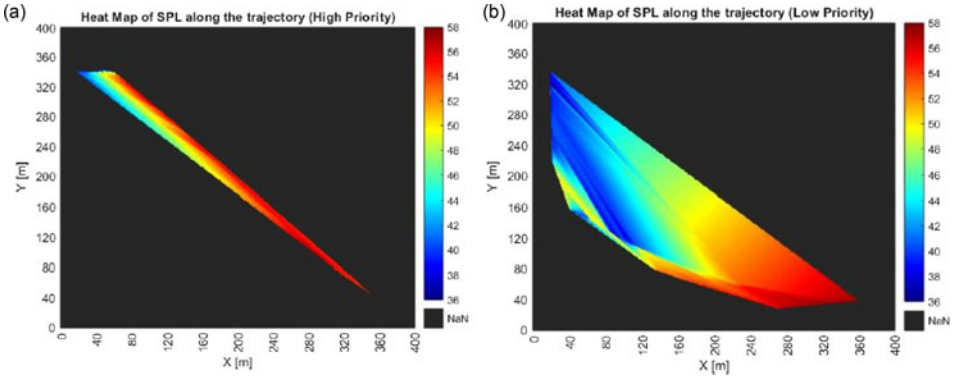


Figure 16. *(a) Noise heat map of the HPM UAV path. (b) Noise heat map of the LPM UAV path.*

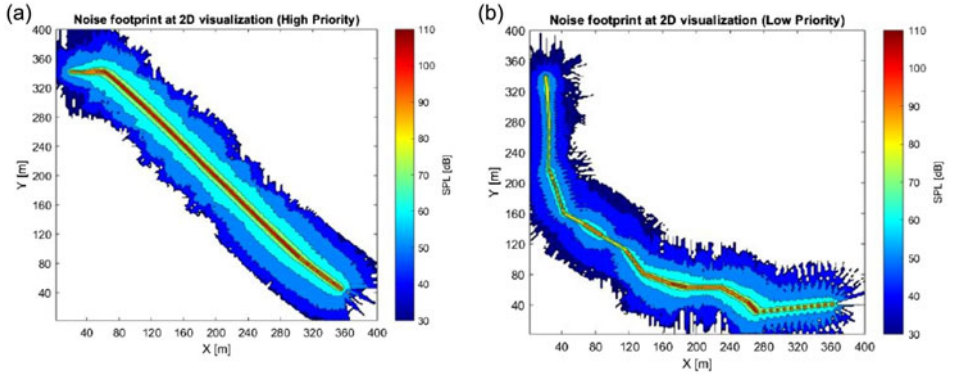


Figure 17. *(a) Noise footprint of the HPM UAV path. (b) Noise footprint of the LPM UAV path.*

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

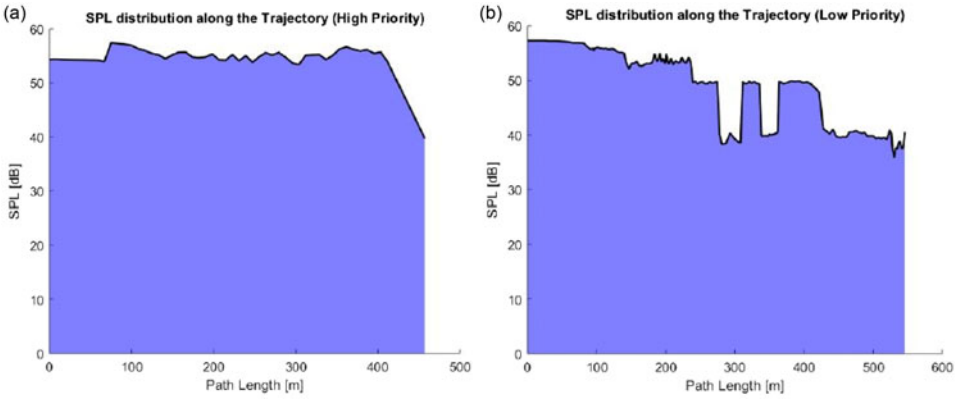


Figure 18. *(a) SPL distribution along the HPM UAV path. (b) SPL distribution along the LPM UAV path.*

	Path Length [m]	Average Velocity [km/h]	Flight Time [s]	Minimum SPL [dB]	Maximum SPL [dB]	Total SPL [dB]	SPL per unit length [dB]
High Priority	457.51	96.8	17.02	38.61	57.43	12415.3	54.31
Low Priority	548.01	71.32	27.66	35.79	57.27	13322.2	48.62
Low vs. High	19.78%	– 26.32%	62.51%	– 7.31%	– 0.28%	7.3%	– 10.48%

Figure 19. *Comparison of performance metrics for HPM and LPM paths.*

a sophisticated balance between path length and noise management, outperforming traditional methods such as direct paths and A* algorithms in key aspects of noise control. The RL agent's ability to follow a longer trajectory, while reducing noise per unit length and achieving lower minimum noise levels, highlights its effectiveness in prioritising noise-sensitive navigation. Despite the longer paths resulting in higher cumulative noise, the RL agent's capacity to minimise noise impact per distance traveled showcases a nuanced approach to noise distribution. This capability is particularly crucial in environments where minimising noise pollution is essential, such as urban areas or noise-sensitive zones. The comparative analysis revealed that the RL agent's trajectory, although the longest, effectively managed noise through dynamic adaptation of velocity based on environmental noise sensitivity. This adaptive control allowed the RL agent to reduce noise in critical areas while balancing overall travel efficiency. The lower minimum noise levels and efficient noise per unit length metrics indicate the RL agent's proficiency in navigating through noise-sensitive environments with minimal disruption. Furthermore, the validation on multiple maps with varying start and target locations affirmed the RL agent's generalisation capability. The consistent performance in minimising noise and managing maximum noise levels across diverse scenarios underscores the robustness and reliability of the RL model. The adaptability in path length and cumulative noise, driven by context-aware adjustments, demonstrates the RL agent's potential to

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

operate effectively in dynamic and varied environments. The study also highlighted the inherent trade-off between path length and noise management, a critical consideration for optimising UAV navigation. While the RL agent currently prioritises noise reduction over the shortest distance, ongoing refinements could enhance its overall performance, balancing both path efficiency and noise management. In conclusion, this research validates the efficacy of RL-based strategies for UAV navigation in noise-sensitive applications. The RL agent's advanced noise management capabilities, combined with its adaptability and generalisation across different scenarios, make it a promising approach for achieving sophisticated environment-aware navigation solutions. Future research directions may include high-fidelity environment mapping including vegetation distribution as well as noise assessment metrics from both acoustics and psycho-acoustics domains.

Acknowledgements. The authors wish to thank the Piedmont Aerospace Cluster (https://www.distrettoaerospazialepiemonte.com/en/) for supporting the PhD activity of Marco Rinaldi.

References

[1] Straubinger, A., Rothfeld, R., Shamiyeh, M., Büchter, K.D., Kaiser, J. and Plötner, K.O. An overview of current research and developments in urban air mobility–setting the scene for UAM introduction, *J. Air Transport Manag.*, 2020, **87**, p 101852.
[2] Rinaldi, M., Wang, S., Geronel, R.S. and Primatesta, S. Application of task allocation algorithms in multi-UAV intelligent transportation systems: a critical review, *Big Data Cognit. Comput.*, 2024, **8**, (12), p 177.
[3] Rinaldi, M. and Primatesta, S. Comprehensive task optimization architecture for urban UAV-based intelligent transportation system, *Drones*, 2024, **8**, (9), p 473.
[4] Rinaldi, M., Primatesta, S., Bugaj, M., Rostáš, J. and Guglieri, G. Urban air logistics with unmanned aerial vehicles (UAVs): double-chromosome genetic task scheduling with safe route planning, *Smart Cities*, 2024, **7**, (5), pp 2842–2860.
[5] Rinaldi, M., Primatesta, S., Bugaj, M., Rostáš, J. and Guglieri, G. Development of heuristic approaches for last-mile delivery TSP with a truck and multiple drones, *Drones*, 2023, **7**, (7), p 407.
[6] Geronel, R.S., Rinaldi, M., da Silva, M.M. and Primatesta, S. Quadrotor UAV elastically attached to uncertain payload: trajectory control and vibrations suppression, *Unmanned Syst.*, 2025. doi: 10.1142/s2301385026500408
[7] Wang, S., Liu, H., Rinaldi, M. and Tsang, Y.P. Evaluating the impact of air corridors on the environment and public interests, *Transport. Res. Part D: Transport Environ.*, 2025, **143**, p 104732.
[8] Eißfeldt, H. and Biella, M. The public acceptance of drones–challenges for advanced aerial mobility (AAM), *Transport. Res. Proc.*, 2022, **66**, pp 80–88.
[9] Schäffer, B., Pieren, R., Heutschi, K., Wunderli, J.M. and Becker, S. Drone noise emission characteristics and noise effects on humans—a systematic review, *Int. J. Environ. Res. Public Health*, 2021, **18**, (11), p 5940.
[10] Çetin, E., Cano, A., Deransy, R., Tres, S. and Barrado, C. Implementing mitigations for improving societal acceptance of urban air mobility, *Drones*, 2022, **6**, (2), p 28.
[11] Adlakha, R., Liu, W., Chowdhury, S., Zheng, M. and Nouh, M. Integration of acoustic compliance and noise mitigation in path planning for drones in human–robot collaborative environments, *J. Vib. Control*, 2023, **29**, pp 4757–4771.
[12] Sarhan, S., Rinaldi, M., Primatesta, S. and Guglieri, G. Noise-aware UAV path planning in urban environment with reinforcement learning. *Eng. Proc.*, 2025, **90**, (1), p 3.
[13] Varotsos, C.A. and Cracknell, A.P. Remote sensing letters contribution to the success of the sustainable development goals – UN 2030 agenda, *Remote Sens. Lett.*, 2020, **11**, pp 715–719.
[14] Kapoor, R., Kloet, N., Gardi, A., Mohamed, A. and Sabatini, R. Sound propagation modelling for manned and unmanned aircraft noise assessment and mitigation: a review, *Atmosphere (Basel)*, 2021, **12**, p 1424.
[15] Kennedy, J., Garruccio, S. and Cussen, K. Modelling and mitigation of drone noise, *Vibroeng. Proc.*, 2021, **37**. doi: 10.21595/vp.2021.21988
[16] Šiljak, H., Einicke, K., Kennedy, J. and Byrne, S. Noise mitigation of UAV operations through a complex networks approach, *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, InterNoise22, Glasgow, Scotland, Pages 2000–2998, 2022, pp 2208–2214. doi: 10.3397/IN_2022_0316
[17] Tan, Q. et al. Enhancing sustainable urban air transportation: low-noise UAS flight planning using noise assessment simulator, *Aerosp. Sci. Technol.*, 2024, **147**, p 109071.
[18] Nguyen, C., Shihua, J.M., and Liem, R.P. Fuel- and noise-minimal departure trajectory using deep reinforcement learning with aircraft dynamics and topography constraints, *Commun. Transp. Res.*, December 2025, **5**, p 100165. doi: 10.1016/j.commtr.2025.100165
[19] Tan, Q. et al. Low-noise flight path planning of drones based on a virtual flight noise simulator: a vehicle routing problem, *IEEE Intell. Transport. Syst. Mag.*, 2024, pp 2–17.
[20] Tan, Q. et al. Noise assessment and low-noise flight path planning platform for urban air mobility, *J. Acoust. Soc. Am.*, 2023, **154**, p A293.
[21] Scozzaro, G., Delahaye, D. and Ernesto Vela, A. Noise abatement trajectories for a UAV delivery fleet, *9th SESAR Innovat. Days*, December 2019, Athenes, Greece. hal-02388280v3.

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press

[22] Scott, D., Manyam, S.G., Casbeer, D.W., Kumar, M., Rothenberger, M.J. and Weintraub, I.E. Power management for noise aware path planning of hybrid UAVs, *2022 American Control Conference (ACC)*, Atlanta, GA, USA, 2022, pp 4280–4285, doi: 10.23919/ACC53348.2022.9867385
[23] Pang, B., Hu, X., Dai, W. and Low, K.H. UAV path optimization with an integrated cost assessment model considering third-party risks in metropolitan environments, *Reliab. Eng. Syst. Saf.*, 2022, **222**, p 108399.
[24] Hohmann, N., Bujny, M., Adamy, J. and Olhofer, M. Multi-objective 3D path planning for UAVs in large-scale urban scenarios, *2022 IEEE Congress on Evolutionary Computation (CEC)*, Padua, Italy, 2022, pp 1–8. doi: 10.1109/CEC55065.2022.9870265
[25] Tan, Q., Li, Y., Bao, H., Zhou, P., Lo, H.K., Zhong, S. and Zhang, X. Low-noise multi-agent intelligent navigation for unmanned aircraft systems, in *30th AIAA/CEAS Aeroacoustics Conference (2024)*, 2024. doi: 10.2514/6.2024-3234
[26] Bian, H., Tan, Q., Zhong, S. and Zhang, X. Reprint of: assessment of UAM and drone noise impact on the environment based on virtual flights, *Aerosp. Sci. Technol.*, 2022, **125**, p 107547.
[27] Škultéty, F., Bujna, E., Janovec, M. and Kandera, B Noise impact assessment of UAS operation in urbanised areas: field measurements and a simulation, *Drones*, 2023, **7**, p 314.
[28] Kawai, C., Jäggi, J., Georgiou, F., Meister, J., Pieren, R. and Schäffer, B How annoying are drones? A laboratory study with different drone sizes, maneuvers, and speeds, in *Proceedings of the 2024 Quiet Drones Conference*, Manchester, UK, 8–11 September 2024, 2024.
[29] ISO 9613-2:2024; Acoustics—attenuation of sound during propagation outdoors—Part 2: engineering method for the prediction of sound pressure levels outdoors. International Standards Organisation: Geneva, Switzerland, 2024.
[30] Silver, D., Lever, G., Heess, N.M., Degris, T., Wierstra, D. and Riedmiller, M.A Deterministic policy gradient algorithms, in *Proceedings of the 31st International Conference on Machine Learning (ICML 2014)*, Beijing, China, 21–26 June 2014, 2014.

Cite this article: Rinaldi M., Primatesta S. and Guglieri G. (2025). Low-Noise Path Planning for Urban Drone Missions: Acoustic Ray Tracing and DDPG Algorithm. *The Aeronautical Journal*, **129**, 3379–3401. https://doi.org/10.1017/aer.2025.10054

https://doi.org/10.1017/aer.2025.10054 Published online by Cambridge University Press