Time-Bounded Large-Scale Mission Planning Under Uncertainty for UV Disinfection

Lara Brudermüller, Raunak Bhattacharyya, Bruno Lacerda and Nick Hawes

Oxford Robotics Institute, University of Oxford
{larab, raunakbh, bruno, nickh}@robots.ox.ac.uk

Abstract

The COVID-19 pandemic has motivated research on mobile robot-based disinfection methods to help contain the spread of the virus, including ultraviolet (UV) germicidal inactivation. Recent approaches have focused on formulating autonomous disinfection as a coverage problem. However, the focus so far has been on maximising coverage, rather than scaling solutions to large-scale environments or making solutions robust to environmental uncertainty. Since the intensity of UV light is strongly coupled with the distance to the target surface, localisation errors should be included in the decision making process to synthesise meaningful irradiation durations. Therefore, in this paper we solve a linked path and dosage planning problem, explicitly considering *localisation uncertainty* in the model. Our model is formulated as a Markov decision process (MDP) which maps localisation uncertainty to dose delivery distributions given radiation and localisation models. We solve this (MDP) over a finite horizon using prioritised value iteration to maximise dose delivery within specified time bounds. Simulation experiments performed on real-world data show successful disinfection, outperforming a rule-based baseline.

INTRODUCTION

Demand for large-scale disinfection solutions has increased drastically with the COVID-19 pandemic. Among them, Ultraviolet (UV) disinfection has received significant interest, due to its strong antimicrobial properties in the UVC (200-280nm) spectrum. Autonomous mobile disinfection robots, such as UVC lamp carriers, eliminate the need for human interference in the disinfection process and reduce the risk of infecting or irradiating cleaning staff. Moreover, as the effectiveness of UVC correlates with the distance and line-of-sight to the target surface, stationary disinfection fixtures are impractical and not suitable for large-scale environments. Current commercial disinfection robots either do not have autonomous navigation capabilities or require human supervision and intervention. Therefore, in this paper we develop solutions for fully autonomous disinfection robots which do not require human supervision.

Autonomous mobile robot UV disinfection is a planning problem that involves both navigation, i.e. a motion plan, and UV dose delivery. This problem has been tackled previously in environments such as a food bank (Pierson et al. 2021) and a hospital (Correia Marques et al. 2021). In general the robot is given a set of locations where it can perform a *disinfection action* to clean a surface. The robot should aim to reach a given percentage of inactivation of microbial concentrations on each surface (Chick 1908; Pierson et al. 2021). The nature of the disinfection action is specific to the hardware and method chosen, but could correspond to irradiating the surface from a stationary vantage point (Correia Marques et al. 2021), or the creation and execution of a coverage path (Pierson et al. 2021). In this paper we assume the former, but our methods generalise to any disinfection process that can be triggered as a discrete action.

Figure 1: The library environment with a topological map overlaid. The red arrows around the robot represent uncertainties from the particle cloud.

We are motivated by the task of disinfecting surfaces in a public library (see Fig. 1). In this setting the robot can only operate outside the library's opening hours, which enforces a natural time bound. We define the UV disinfection problem as a mission planning problem, rather than a coverage path planning problem, with the assumption that individual disinfection actions provide the necessary coverage of the tar-

get surface. Moreover, public environments such as a library naturally contain many sources of uncertainty, such as visitors moving books, chairs and shelves throughout the day. In addition, robot localisation uncertainty affects the UV doses delivered to the surfaces as its effectiveness is directly coupled to the distance to the target surface. Consequently, the duration of disinfection actions, and their associated rate of in microbial inactivation, cannot be known with certainty at planning time. Therefore, we formulate the time-bounded autonomous mobile robot UV disinfection task as a finite-horizon Markov decision process (MDP) (Puterman 1994). Commonly used for robot planning, an MDP is a mathematically principled framework for sequential decision making *under uncertainty* problems (Budd et al. 2021; Lacerda et al. 2019; Tomy et al. 2020).

Contributions The main contributions of this paper are: i) a formulation of the autonomous mobile robot disinfection problem as a finite-horizon MDP and ii) a model of the impact of localisation uncertainty on the disinfection performance which maps confidence levels in robot location to a distribution over durations to reach microbial inactivation thresholds. The MDP can be solved using standard solvers, such as value iteration (Bellman 1966). The uncertainty model is generated from real-world empirical robot data. The overall approach is demonstrated on a simulated library environment involving 70 locations (Fig. 1) and varying time bounds, proving its applicability to large-scale environments under uncertainty.

Related Work

UV disinfection applications Most approaches formulate the UV disinfection problem as a coverage path planning problem, where the task is to cover all target points with UV radiation. This problem has two steps: decomposing the free space into sub-components; and visiting each component using a path planning algorithm. For example, Pierson et al. (Pierson et al. 2021) decompose cells in a grid space into regions using a Voronoi tessellation and then identify a path connecting all regions using the A^* algorithm. Similarly, (Kurniawan and Adiprawita 2021) uses a Spanning Tree Coverage algorithm on a discretised grid and plans a collision-free path on top of that using a sampling-based path planner. Another approach, taken by Tiseni et al. (Tiseni et al. 2021) is to model the environment using 3D discrete surfaces which generate an attractive potential field (APF) modelling the radiance-distance correlation. These APFs are then used as constraints in a genetic algorithm motion planner. In contrast, Perminov et al. (Perminov et al. 2021) treat the task as a pure path planning problem. Most similar to our approach is the work of Marques et al. (Correia Marques et al. 2021) who discretise the disinfection task into a set of dwelling locations, build up a radiation model using tools from computer graphics and then, given a time bound, aim to minimise the dwell times in each location while reaching some minimum dosage thresholds. The sequence of points to be visited is obtained by solving a Traveling Salesperson Problem (TSP), whereas the allocation of dwell times to each location is solved as a linear program.

However, all of the above solutions assume a deterministic environment, where a disinfection action of a given duration achieves a deterministic change in cleanliness for the target surface. This is a strong assumption which restricts the system's ability to *guarantee* a minimum performance. UV radiation strongly depends on the distance to the target surface. In a robotic setting, any uncertainty in the robot pose will therefore affect dosages.

Planning under uncertainty In this subsection we summarise approaches which could be applied to the UV disinfection problem *under uncertainty*.

The work of (Nardi and Stachniss 2019) demonstrates how uncertainty in a robot's location can be included in the state of an MDP. It assumes *during planning* that location uncertainty can be approximated by a Gaussian distribution with isotropic variance. A distribution over the location variance is then discretised and used as a state factor in an MDP to represent different degrees of uncertainty. This creates an *Augmented MDP*, as initially formulated by Roy et al. (Roy and Thrun 2000), which approximates a Partially Observable MDP (POMDP) by modeling uncertainty as part of the state. We apply this approach to model the effect of localisation uncertainty on the UV planning problem. We also build on the work of (Lacerda, Parker, and Hawes 2017), using a *timed* MDP to explicitly model the distribution over discrete action durations.

In (Duckworth, Lacerda, and Hawes 2020) uncertainties in a time-bounded mission planning scenarios are approached by extending an MDP model with a Gaussian Process (GP) belief about the *a priori* uncertain dynamics. They solve this problem using a sampling-based approach, avoiding the need for the aforementioned discretisation. However it is unclear how well the assumptions of a GP would transfer to modelling the spatiotemporal variation of localisation uncertainty.

Finally, a problem from operations research literature related to our UV planning problem is the *Orienteering Problem (OP)*. Different from a TSP it associates a profit with each node, but does not required every node to be visited. Instead, the goal is to find a feasible tour which maximises profit, where the total cost of the tour satisfies a capacity constraint. Recently, OPs have been extended to include uncertainty (Gunawan, Lau, and Vansteenwegen 2016), e.g. with stochastic profits (Ilhan, Iravani, and Daskin 2008) or stochastic waiting times in the nodes, stochastic travel times (Angelelli et al. 2017) or stochastic weights on the profits (Evers et al. 2014). However, often these problem instances can only be solved with approximate algorithms. Moreover, their applicability is usually restricted to small environments, not scaling to larger ones.

PRELIMINARIES

Markov Models

We model the time-bounded mission planning problem under uncertainty using a finite-horizon *Markov decision process (MDP)*.

Definition 1. *A finite-horizon MDP (Puterman 1994) is a tuple $\mathcal{M} = \langle S, \bar{s}, A, \mathcal{T}, \mathcal{R}, H \rangle$, with S being a finite set of states; $\bar{s} \in S$ the initial state; A a finite set of actions; $R : S \times A \to \mathbb{R}$ being a reward structure; $\mathcal{T} : S \times A \times S \to [0, 1]$ a probabilistic transition function returning the probability of arriving at state s' after having taken action a in state s; and $H \in \mathbb{N}$ a time horizon.*

An MDP represents all possible evolutions of a system's state, depending on the choice of *actions* at each *state*, visiting successor state s' according to $T(s, a, s')$. A path through an MDP of length H is a sequence $x = s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} ... \xrightarrow{a_{H-1}} s_H$, with $T(s_i, a_i, s_{i+1}) > 0 \; \forall i < H$.

Finite Horizon Optimisation

The aim of this work is to find a deterministic Markovian policy, i.e. a mapping $\pi : \mathbf{S} \times \{1, ..., H\} \to A$. In the given setting, it should maximise the expected cumulative reward within the time horizon H, i.e. $\pi^* = \text{argmax}_\pi \mathbb{E}^\pi_{\mathcal{M},s} \big[\sum_{i=0}^{H} R(s_i, a_i) \big]$.

Topological Map

We represent the environment using a *topological map* with *locations* the robot can visit and *edges* along which it can navigate (Lacerda et al. 2019).

Definition 2. *A topological map is defined by a tuple $\mathcal{T} = \langle L, E, \xi \rangle$. Here, $L = \{l_1, \ldots, l_n\}$ is a set of relevant locations in the environment, represented by robot poses of the form (x, y, θ) on a global frame; $E \subseteq V \times V$ encodes a set of directed edges the robot can traverse and $\xi : E \to \mathbb{R}_{\geq 0}$ is a function which maps from edges to travel times.*

PROBLEM SETTING

We approach the UV disinfection problem as a mission planning problem on a topological map with a location placed at every point where the robot must start a cleaning action (e.g. irradiating a surface, or creating then executing a coverage plan). We refer to the length of time the robot spends at each location as the *dwell time*. We assume that the robot can observe its current topological location and can quantify its metric localisation uncertainty. The decision to be taken at each location is the choice of what *cleanliness level* should be achieved, where a cleanliness level is a guaranteed log-reduction in microbial activation. The corresponding planning problem is to synthesise a robot policy which maximises the cleanliness level across all locations given a fixed time bound.

For the remainder of this paper we assume the robot hardware in Fig. 2. This robot carries a panel of UVC LED strips, with each strip providing an array of point light sources. Our target environment is the library shown in Fig. 1, which includes a manually constructed topological map. Despite these specialisations, the method and MDP structure presented below can generalise to other robot and environment combinations. For example, with an appropriate model linking dwell time to change in cleanliness level, the same general approach could be used to build a policy for disinfecting a collection of hospital wards as in (Correia Marques et al. 2021), where each topological location would correspond to a ward.

METHOD

This section introduces the components needed for the final MDP formulation: a navigation component for *path planning*; an *irradiation model*; and a *model for localisation uncertainty* .

Path Planning

In our approach, we *decouple* navigation from the disinfection problem. Given the total time budget T and an initial location, path planning is performed by solving a TSP over the topological graph. This generates a tour visiting all locations in the topological map. Given the average speed of the robot and the total distance of the tour from the TSP, we compute its travel time T_{TSP}. The remaining budget for the allocation of dwell times in each location is then $T_{MDP} = T - T_{TSP}$. Note that the quality of the TSP solution influences the initial time budget. The transition between locations is the only *deterministic* component in this model. We define $\Lambda_{TSP} : S_l \to S_l$ as the mapping from a location to the next location in the TSP tour.

Irradiation Model

We next define an irradiation model to relate the time the robot spends in one location to the UV dose delivered to the associated surface.

Preliminaries on UV Disinfection The total UV dosage received at a point on a surface is defined by $D = I \cdot \Delta t$ (expressed in J/m^2 $= (\text{W} \cdot s)/\text{m}^2$), where I measures the intensity of the UVC light the point receives and Δt corresponds to the time duration of exposure. The intensity at a point is proportional to the inverse square distance to the robot (Pierson et al. 2021; Arguelles 2020). When treating the source of radiation as a point light source, I can be expressed as Eqn. 1, where P_{UVC} corresponds to the power rating of the UVC light, $\eta \leq 1$ defines an attenuation factor and r represents the distance from the light source to the sample point. For our model, we use a conservative estimate of $\eta \approx 0.1$, as proposed by (Arguelles 2020); and P_{UVC} is 8W per light source according to our hardware specifications.

$$I = \frac{P_{UVC}}{A_{exposed}} = \frac{\eta P}{4\pi r^2} \tag{1}$$

Model & Setup We assume that at each location in the topological map the robot only has to disinfect one surface, i.e. the shelf it is directly parked in front of, and approximate the surface of a shelving unit as a plane in 3D space. Using the physical model above, we therefore derive an irradiation model that can deal with any plane in 3D space. We calculate the dose received on this plane by discretising it into a grid of surface patches. Per grid cell of the shelf surface grid, we (i) iterate over all UV point light sources of the robot, (ii) compute whether the cell's centre point is in the field of view

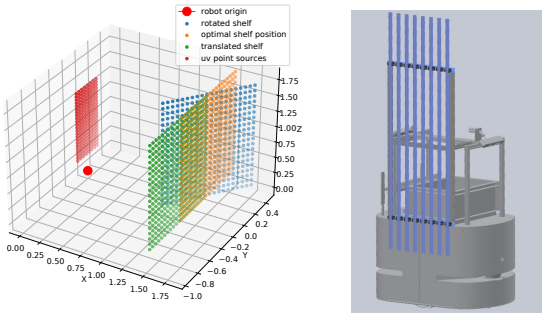

Figure 2: *Left:* Outline of the robot setup and how uncertainties in robot position and orientation are projected onto the shelf pose. *Right:* CAD model of the robot setup chosen after considering the given radiation model.

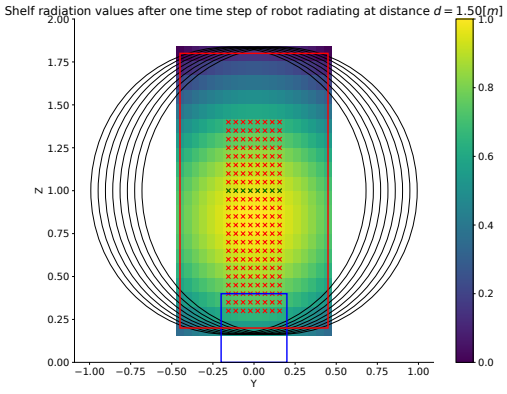

Figure 3: Example of normalized radiation values for each grid cell of a discretised shelf surface (red rectangle) radiated by the robot (blue) at a distance of 1.5 meters. The black circles correspond to the fields of view at the intersection of the shelf plane for the green array of point sources (crosses).

(corresponding to a 45 degree cone) and if so (iii) calculate the dosage D received from that source. Finally, the dosage value associated to a grid cell is the sum of dosage values received from all point light sources.

In order to analyse the effects of uncertainty in the robot's pose, we rotate, shift and translate the shelf relative to the robot, as shown in Fig. 2. In order to quantify dosages received on a shelf surface with just a single value, we use the first quartile of all dosage values along all grid cells. This gives a conservative estimate of the entire dosage. Fig. 3 shows a heatmap of the normalised dosage values of the discretised shelf surface when the robot radiates at a distance of 1.5m without any rotation or translation from the target pose. Analysing the effects of varying the robot pose in x, y and θ in more detail, we find that nodes in the topological map should ideally be placed at locations with a distance $d = 0.31$m from each shelf in order to achieve the highest possible radiation performance. However, this is hard to achieve in practice. Therefore, nodes in the current hand-constructed topological map were placed at an average distance of $d = 1.4$m to each shelf.

Cleanliness Levels

We model the dosage delivered in each location using discrete cleanliness levels corresponding to percentage inactivations of microbial concentrations. Since disinfection is usually modeled with log-linear models, cleanliness levels are linked to the required *log-reductions* (Pierson et al. 2021). A 1-log reduction corresponds to 90% inactivation and D_{90} represents the dosage needed to achieve this log-reduction. Since UVC disinfection is log-linear with respect to time, reaching a 99% inactivation rate takes twice the exposure time, i.e. $D_{99} = 2 \cdot D_{90}$. Recently, multiple studies have attempted to determine dosage thresholds for inactivation of SARS-CoV-2. However, since this is still ongoing research, values are not consistent across studies. Tab. 1 summarises the dosage values associated to log-reductions used in the field. As they are collected across different sources, they do not follow the log-linear property described above. Therefore, in our model, we only use the 1-log reduction threshold and compute values for the 2- and 3-log reduction (see Tab. 1).

Transition Functions

In our disinfection MDP the robot's actions correspond to a choice of which cleanliness level i to reach at the current location. The time remaining after this action is stochastic, with the duration distribution t_{clean_i} conditioned on the uncertainty in the robot's localisation. Below we describe how this probabilistic evolution of state is encoded in the MDP's transition function $T(s, a, s')$.

State space We model the MDP using a *factored* state space $S_{uv} = \{(l, c, t) \in S_l \times S_c \times \{0, ..., T_{MDP}\}\}$ where S_l is the set of topological locations, $S_c = \{0, ..., n\}$ a set of discrete localisation confidence levels, and t is the time remaining. Note, that with this formulation actions which may exceed the remaining time bound t in the state can be pruned from the MDP. In the remainder of this work, we will use s_l, s_c, s_t to refer to the values of the individual state factor l, t and c, respectively.

Transition function for c The duration required to disinfect a location depends on the confidence with which the robot is localised. Following the *Augmented MDP* approach (Nardi and Stachniss 2019; Roy and Thrun 2000) we represent the robot's confidence in its location using confidence levels S_c. We construct these levels by learning from a dataset $D = \{(\sigma_x, \sigma_y, \sigma_\theta)\}_{i=1}^{m}$ containing m samples of pose standard deviations (see *Experimental Evaluation* for more details about the dataset). We cluster the standard deviations using a Gaussian Mixture Model (GMM), defined as $P(\vec{\sigma}) = \sum_{i=1}^{K} \phi_i \mathcal{N}(\vec{\sigma}|\vec{\mu}_i, \Sigma_i)$, where K sets the number of multivariate Gaussian components and $\vec{\phi}$ associates the weights to each component in the mixture model. Hence, each component is a separate multi-variate Gaussian distribution modelling variations in x, y, θ. We map each component to a confidence level in the MDP, in order of decreasing norm of the component's mean. The number of components in the GMM1 (and hence $|S_c|$) was determined using the

Table 1: Rewards, dosages associated to cleanliness levels

Log-reduction	D_{90} (1-log)	D_{99} (2-log)	$D_{99.9}$ (3-log)
Cleanliness level	1	2	3
Literature dose thres. (J/m^2)	100 (Pierson et al. 2021)	134 (Kariwa, Fujii, and Takashima 2006)	280 (Correia Marques et al. 2021)
Model dose thres. (J/m^2)	100	200	400
Reward	R_1	$R_1/2$	$R_1/4$

Bayesian Information Criterion. We assume that the confidence level at a location is independent of all other state factors. We therefore compute $P(S_c|l)$ from the distribution across GMM components of the points from D that are from l, where we assign a point to a component using maximum likelihood.

Transition function for t The time remaining is updated using $t' = t - t_{clean_i}$ where t_{clean_i} is the duration needed to reach cleanliness level i. The probability of the duration being observed is modelled by $P(t_{clean_i}|c, a = clean_i)$, i.e. the duration required to clean to a given level defined by the action depends on both, the level chosen and the current confidence in the robot's localisation. For each confidence-cleanliness level pair we build $P(t_{clean_i}|c, a = clean_i)$ by sampling poses from the GMM component for c and using the UV irradiation model described above to determine the duration required to bring the lower dose quartile to the cleanliness level i. We then fit a categorical distribution to these samples using a discretisation of 5s. This pipeline is described in Alg. 1, and the resulting distributions are shown in Fig. 4 with the corresponding boxplots in Fig. 5. The time for the action of no cleaning is always zero (i.e. $t_{clean_0} = 0$). This allows the disinfection agent to always have an action available at every location even if no time remains.

Algorithm 1: Generation of $p(t_{clean_i}|c, clean)$

Given: RadiationModel
Parameters: `num_components`, `num_samples`, `cleanliness_levels`, optimal dwell pose $\boldsymbol{x}^*$
Init: $all\_dists \leftarrow \emptyset$
Fit GMM(`num_components`) to stdevs (AMCL data)
Order means of GMM components by their norm along each dimension
for `clean` in `cleanliness_levels` **do**
  for $\sigma_c = (\mu_x, \mu_y, \mu_\theta)$ in ordered GMM comps **do**
    $dist_{clean,c} \leftarrow \emptyset$
    while $i \leq$ `num_samples` **do**
      Sample displacement: $\Delta x \sim \mathcal{N}(0, \sigma_c)$
      $\boldsymbol{x'} = \boldsymbol{x}^* + \Delta x$
      $d_{\Delta t=1}$ = RadiationModel($\boldsymbol{x'}$)
      $t_{i,clean} = \texttt{clean}/d_{\Delta t=1}$
      $dist_{clean,c} \leftarrow dist_{clean,c} \cup \{t_{i,clean}\}$
    end while
    $all\_dists \leftarrow all\_dists \cup dist_{clean,c}$
  end for
end for
return all_dists

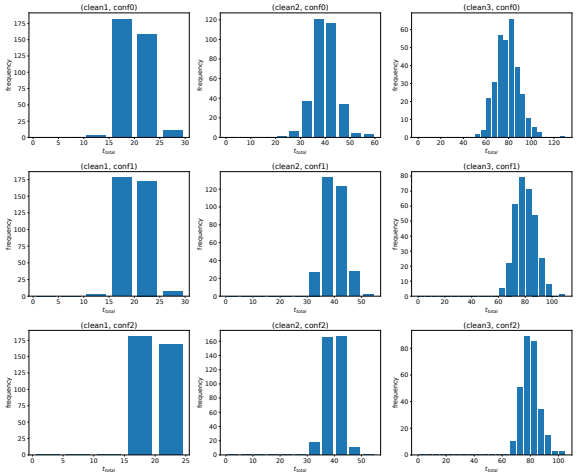

Figure 4: Categorical distributions for durations t_{clean_i} for each combination of confidence and cleanliness levels.

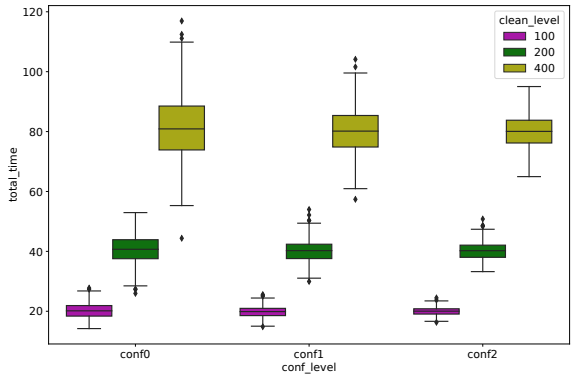

Figure 5: Boxplots showing spread of dwell times, given different levels of confidence for each cleanliness level.

Reward

Our model gives diminishing rewards for increasing levels of cleanliness, as shown in the last row of Tab. 1. R_1 corresponds to cleanliness level 1 and gives the highest reward. Per step increase in cleanliness level, the reward of the previous level is halved. This is to incentivise the disinfection agent to aim at reaching a minimum cleanliness level in every location, rather than to clean a subset of locations to a high level. Not cleaning at all, i.e. cleanliness level 0, gives 0 reward. R_1 is subject to tuning and has been set to 100 in

our implementation.

MDP Model of UV Disinfection Task

Given all the components above we construct a finite-horizon MDP $\mathcal{M}_{uv} = \langle S_{uv}, \bar{s}, A_{uv}, \mathcal{T}_{uv}, \mathcal{R}_{uv}, H \rangle$ over the fixed time-bound T_{MDP}. The horizon H of the MDP is governed by the number of locations $|S_l|$. In each location one action is taken: the decision on which cleanliness level to achieve. The other tuple elements in $\mathcal{M}_{uv}$ are summarised as follows:

(i) S_{uv} is the state space as introduced in the paragraph about transition functions in the Method section.

(ii) $A_{uv} = \{0, 1, 2, 3\}$ is a set possible actions in each location. It corresponds to the choice of cleanliness level to achieve in that location.

(iii) $\mathcal{T}_{uv}$ is the transition function modelling the stochastic dynamics in the robot localisation. It is defined as

$$T_{uv}(s, a, s') = \mathbb{I}[s'_l \neq \Lambda_{TSP}(s_l)] \cdot P(s'_c|s'_l) \cdot P(s'_t|s_t, a, t_{clean_i}) \quad (2)$$

where $\mathbb{I}[s'_l \neq \Lambda_{TSP}(s_l)]$ is an indicator function incorporating the deterministic transition of the locations, i.e. it will take the value 1 if the new location maps to the next location given by the TSP sequence; $P(s'_t|s_t, a, t_{clean_i})$ is the transition probability for s_t being conditioned on $t_{clean_i} \sim P(t_{clean_i}|s_c, a = clean_i)$ which is defined also defined in the transition function paragraph along with $P(s'_c|s_l)$.

(iv) $\mathcal{R}_{uv}$ is the reward function mapping from cleanliness levels a fixed reward value, as defined in the previous paragraph about rewards.

(v) $\bar{s} = (l_0, n, T)$ is the initial state, starting in the first location of a given TSP sequence.

EXPERIMENTAL EVALUATION

Data collection

In order to learn the radiation dynamics, we recorded a dataset from a deployment of a *MetraLabs SCITOS X3* robot[1] navigating in Oxfordshire County Library on a topological map (cf. Fig. 1) over a period of approximately 4 hours. The robot localised using Adaptive Monte Carlo Localisation (AMCL) (Fox 2001) and the dataset includes logs of the robot's localisation uncertainty from this method.

Baseline

In order to quantify the performance of our proposed model, we compare it against a rule-based *baseline model*. Similar to our model, the time budget for the baseline is the time remaining after subtracting the time needed for the TSP tour from the total budget T. Instead of setting a cleanliness level to be reached in each location, the baseline model allocates uniform dwell duration $\beta = T_{MDP}/|S_l|$ to all locations in the topological graph. In a deterministic world a constant dwell time per node should always generate the same cleanliness level i. However, in reality, some locations will not be able to reach a chosen cleanliness level, due to variations in the actual robot pose at each location.

For a meaningful comparison between the baseline and our model, the time bounds are set to values allowing the baseline to achieve optimal performance in a deterministic setting. Therefore, it takes values $T = T_{TSP} + |S_l| \cdot (D_j/\alpha^*)$, where α^* is the dose for one time step if the robot is in the exact location, computed via our radiation model; D_j the total dosage needed to reach different cleanliness levels; and T_{TSP} the time needed to complete the TSP tour at a predefined robot speed. Tab. 2 lists all values used for the experiments.

While value iteration generates optimal results by definition, the baseline results were computed over a sufficiently large number of 100 samples using the same deterministic reward structure as in our MDP formulation.

Results

Table 2 summarises the total rewards received by the baseline model as compared to our model for the different time bounds. Our UV MDP model outperforms the baseline in each of the experiments. We observe that with increasing time bounds, the difference between the two models decreases, since at some point there is enough slack in the time budget for the baseline to get all reward at every location, regardless of the uncertainty. This is the case in the limit of the highest time bound where the reward values are almost equal.

In addition, Fig. 6 provides more insight why the MDP model outperforms the baseline. The evolution of the cumulative path reward given a policy generated by our model versus the baseline model, shown in the top plot of Fig. 6, highlights the ability of our model to not only achieve higher rewards but also to achieve them more quickly, showing its ability to adapt to the action durations it observes in the environment. Moreover, from this we can also observe that the MDP does not use the entire time budget and yet generates higher rewards. This is due to the formulation of the MDP model, which does not attribute extra time to locations once the remaining time bound is not large enough to reach an entire increase in cleanliness level. This could be addressed by including a new action of assigning extra time to the action state space A_{uv} of the UV MDP model. Last, the histograms in the bottom of Fig. 6 show that the MDP always achieves a better distribution of higher average cleanliness levels across all nodes. For a time bound of 2745 (Exp. iii, i.e. in a deterministic setting the baseline should always reach cleanliness level 2) the middle histogram shows that in more than half of the locations it fails to reach sufficient dosages in order to reach that level, falling back to cleanliness level 1. In contrast, our model can guarantee a minimum cleanliness level of 2 in all locations and in some locations even achieves the maximum level.

[1] https://www.metralabs.com/en/mobile-robot-scitos-x3/

Table 2: Comparison of total rewards, given different time bounds

Exp.ID	Time bound	Baseline	Ours
i	1373 ($D_{90}/\alpha^* \cdot \lvert S_l \rvert$)	3334.0	7801.84
ii	2059 ($150/\alpha^* \cdot \lvert S_l \rvert$)	7003.0	9499.67
iii	2745 ($D_{99}/\alpha^* \cdot \lvert S_l \rvert$)	8611.0	10736.56
iv	4118 ($300/\alpha^* \cdot \lvert S_l \rvert$)	10502.0	11512.99
v	5490 ($D_{99.9}/\alpha^* \cdot \lvert S_l \rvert$)	11320.5	12249.89
vi	6863 ($500/\alpha^* \cdot \lvert S_l \rvert$)	12244.5	12249.99

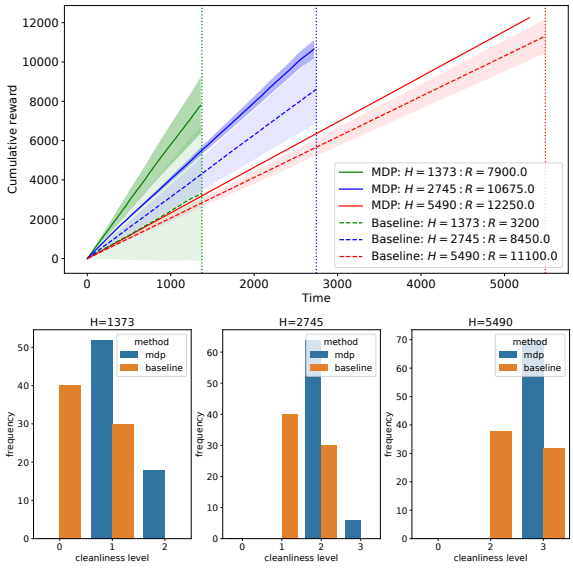

Figure 6: *Top:* Collected cumulative rewards over the given time horizons comparing the baseline (dashed lines) and our model (solid lines). Shaded areas show the corresponding variance of one standard deviation and dashed vertical lines correspond to the time bounds. *Bottom:* Average distributions of cleanliness levels across all locations with the baseline and our model for time bounds H corresponding to experiments 1,3, and 5 in Tab.2.

CONCLUSIONS

Our work successfully demonstrates a new approach of modelling and solving the UV disinfection task for large-scale environments under *uncertainty*. We quantify uncertainty by learning state transition probabilities from empirical robot data. These probabilities are used to parametrise a radiation model for the library application. Compared to a rule-based baseline, assigning dwell times uniformly across all locations, our model achieves significantly higher rewards. This is due to the fact that it is able to reason over the stochasticity in the system's dynamics, acting optimally with respect to its localisation uncertainty.

So far, our model has been only tested in simulations. However, the next step in this project is to deploy the model on the robot in a real-world environment. Moreover, we will investigate including more sources of uncertainty into our model, e.g. from changes in the environment. These changes could be tracked via real-time mapping algorithms and our model could incorporate such uncertainties from the sensor model into the state space and the corresponding transition function. Finally, the reward structure chosen in this work also allows us to configure the reward structure such that we are able to prioritise locations over others, given the fixed time budget. Such prioritisation could address high-touch surfaces, which could be based on models localising humans in different areas of the library during opening hours.

ACKNOWLEDGMENT

This work was supported by the EPSRC Programme Grant "From Sensing to Collaboration" (EP/V000748/1), and a gift from Amazon Web Services.

References

Angelelli, E.; Archetti, C.; Filippi, C.; and Vindigni, M. 2017. The probabilistic orienteering problem. *Computers and Operations Research*, 81: 269–281.

Arguelles, P. 2020. Estimation UV-C Sterilization Dosage for COVID-19 Pandemic Mitigation Efforts. Technical Report April.

Bellman, R. 1966. Dynamic programming. *Science*, 153(3731): 34–37.

Budd, M.; Lacerda, B.; Duckworth, P.; West, A.; Lennox, B.; and Hawes, N. 2021. Markov Decision Processes with Unknown State Feature Values for Safe Exploration using Gaussian Processes. 7344–7350.

Chick, H. 1908. An investigation of the laws of disinfection. *Journal of Hygiene*, 8(1): 92–158.

Correia Marques, J. M.; Ramalingam, R.; Pan, Z.; and Hauser, K. 2021. Optimized Coverage Planning for UV Surface Disinfection. In *IEEE International Conference on Robotics and Automation (ICRA)*, 9731–9737.

Duckworth, P.; Lacerda, B.; and Hawes, N. 2020. Time-Bounded Mission Planning in Time-Varying Domains with Semi-MDPs and Gaussian Processes. Technical Report CoRL.

Evers, L.; Glorie, K.; Van Der Ster, S.; Barros, A. I.; and Monsuur, H. 2014. A two-stage approach to the orienteering problem with stochastic weights. *Computers and Operations Research*, 43: 248–260.

Fox, D. 2001. KLD-sampling: Adaptive particle filters. *Advances in neural information processing systems*, 14.

Gunawan, A.; Lau, H. C.; and Vansteenwegen, P. 2016. Orienteering Problem: A survey of recent variants, solution approaches and applications.

Ilhan, T.; Iravani, S. M.; and Daskin, M. S. 2008. The orienteering problem with stochastic profits. *IIE Transactions (Institute of Industrial Engineers)*, 40(4): 406–421.

Kariwa, H.; Fujii, N.; and Takashima, I. 2006. Inactivation of SARS coronavirus by means of povidone-iodine, physical conditions and chemical reagents. In *Dermatology*, volume 212, 119–123. Karger Publishers.

Kurniawan, I. T.; and Adiprawita, W. 2021. Autonomy design and development for an ultraviolet-c healthcare surface disinfection robot. *Proceeding - 2021 International Symposium on Electronics and Smart Devices: Intelligent Systems for Present and Future Challenges, ISESD 2021*.

Lacerda, B.; Faruq, F.; Parker, D.; and Hawes, N. 2019. Probabilistic planning with formal performance guarantees for mobile service robots. *The International Journal of Robotics Research*, 38: 1098–1123.

Lacerda, B.; Parker, D.; and Hawes, N. 2017. Multi-objective policy generation for mobile robots under probabilistic time-bounded guarantees. In *Proceedings International Conference on Automated Planning and Scheduling, ICAPS*, 504–512.

Nardi, L.; and Stachniss, C. 2019. Uncertainty-aware path planning for navigation on road networks using augmented MDPs. In *Proceedings - IEEE International Conference on Robotics and Automation*, volume 2019-May, 5780–5786.

Perminov, S.; Mikhailovskiy, N.; Sedunin, A.; Okunevich, I.; Kalinov, I.; Kurenkov, M.; and Tsetserukou, D. 2021. UltraBot: Autonomous Mobile Robot for Indoor UV-C Disinfection. In *IEEE International Conference on Automation Science and Engineering*, volume 2021-Augus, 2147–2152.

Pierson, A.; Romanishin, J. W.; Hansen, H.; Ya, L. Z.; and Rus, D. 2021. Designing and Deploying a Mobile UVC Disinfection Robot. In *International Conference on Intelligent Robots and Systems (IROS)*, 2.

Puterman, M. L. 1994. *Markov decision processes: Discrete stochastic dynamic programming*. wiley.

Roy, N.; and Thrun, S. 2000. Coastal navigation with mobile robots. In *Advances in Neural Information Processing Systems*, figure 1, 1043–1049.

Tiseni, L.; Chiaradia, D.; Gabardi, M.; Solazzi, M.; Leonardis, D.; and Frisoli, A. 2021. UV-C mobile robots with optimized path planning. *IEEE Robotics and Automation Letters*, In press: 59–70.

Tomy, M.; Lacerda, B.; Hawes, N.; and Wyatt, J. L. 2020. Battery charge scheduling in long-life autonomous mobile robots via multi-objective decision making under uncertainty. *Robotics and Autonomous Systems*, 133.