
This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Dual Control for Exploitation and Exploration (DCEE) in autonomous search

PLEASE CITE THE PUBLISHED VERSION

<https://doi.org/10.1016/j.automatica.2021.109851>

PUBLISHER

Elsevier

VERSION

AM (Accepted Manuscript)

PUBLISHER STATEMENT

This paper was accepted for publication in the journal *Automatica* and the definitive published version is available at <https://doi.org/10.1016/j.automatica.2021.109851>.

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Chen, Wen-Hua, Callum Rhodes, and Cunjia Liu. 2021. "Dual Control for Exploitation and Exploration (DCEE) in Autonomous Search". Loughborough University. <https://hdl.handle.net/2134/16577825.v1>.

Dual Control for Exploitation and Exploration (DCEE) in Autonomous Search [★]

Wen-Hua Chen ^a Callum Rhodes ^a Cunjia Liu ^a

^a*Department of Aeronautical and Automotive Engineering
Loughborough University
Leicestershire, LE11 3TU
UK*

Abstract

This paper proposes an optimal autonomous search framework, namely Dual Control for Exploration and Exploitation (DCEE), for a target at unknown location in an unknown environment. Source localisation is to find sources of atmospheric hazardous material release in a partially unknown environment. This paper proposes a control theoretic approach to this autonomous search problem. To cope with an unknown target location, at each step, the target location is estimated by Bayesian inference. Then a control action is taken to minimise the error between future robot position and the hypothesised future estimation of the target location. The latter is generated by hypothesised measurements at the corresponding future robot positions (due to the control action) with the current estimation of the target location as a prior. It shows that this approach can take into account both the error between the next robot position and the estimate of the target location, and the uncertainty of the estimate. This approach is further extended to the case with not only an unknown source location, but also an unknown local environment (e.g. wind speed and direction). Different from current information theoretic approaches, this new control theoretic approach achieves the optimal trade-off between exploitation and exploration in a unknown environment with an unknown target by driving the robot moving towards estimated target location while reducing its estimation uncertainty. This scheme is implemented using particle filtering on a mobile robot. Simulation and experimental studies demonstrate promising performance of the proposed approach. The relationships between the proposed approach, informative path planning, dual control, and classic model predictive control are discussed and compared. This work opens a door for further developing control systems operating in unknown environments, or performing tasks with unknown parameters.

Key words: Autonomous search, informative path planning, dual control, goal oriented control systems, exploration and exploitation

1 Introduction

Searching sources of airborne substance release could find a wide range of applications from disaster management and environment protection, to gas leakage detection in oil and gas industry Singh et al. (2015), Hutchinson et al. (2017). Chemical and biological materials could be released into the atmosphere deliberately (e.g. discharged by a plant, terrorism), naturally (e.g. methane emission, volcanic eruption), or accidentally (e.g. accidents in a chemical plant, Fukushima nuclear disaster).

It can also be seen in natural behaviours from animals searching for food sources based on odours, to insects seeking mating in a large field or forest Vergassola et al. (2007). Quite often source localisation and quantification is referred to as a source term estimation (STE) problem in literature Bieringer et al. (2015), Platt & Deriggi (2010), Hutchinson et al. (2017). In the presence of hazardous material release events, there are two main approaches for existing emergency response practices: a static network of pre-deployed sensors and the manual collection of sensor measurements, e.g. using hand held devices and dedicated manned vehicles. The former is costly and requires substantial pre-planning while the latter puts people in harms way. After data are collected, estimation algorithms can be utilised to estimate the location of sources and release rates.

[★] This paper has not been presented on any IFAC conference. Corresponding author: Wen-Hua Chen. w.chen@lboro.ac.uk

Email addresses: W.Chen@lboro.ac.uk (Wen-Hua Chen), C.Rhodes@lboro.ac.uk (Callum Rhodes), C.Liu5@lboro.ac.uk (Cunjia Liu).

With the advances of robotics and autonomous system

technologies, there is a strong interest in developing a new STE solution using recently available mobile sensor platforms, where chemical or biological sensors are installed on a mobile ground robot or an unmanned aerial vehicle (UAV) Hutchinson et al. (2017). Essentially, this is to search a source in an unknown environment. It could be conducted through exhaustive search, so it can be considered as a coverage search problem. However, due to the time critical nature of this type of mission, there is a strong need in speeding up the search process, particularly for searching in a large area. To this end, various tools have been introduced including optimisation and Bayesian approaches Hutchinson et al. (2017). A key challenge arising in autonomous search using a mobile sensor platform is the *deep interaction between estimation and planning* – based on the current belief, a decision shall be made as to where to take the next measurement in order to maximise the chance of finding the source which, in turn, changes the belief since the new measurement is obtained at a new location. Driven by information-theoretic approaches, informative path planning (IPP) offers a promising solution to the STE problem Boström-Rost (2019). In an IPP approach, a mobile sensor platform plans a path that maximises the information gain about the source and local environment, based on its current belief, and updates its belief through incorporating new measurements of the dispersion environment using on-board sensors Ristic et al. (2010) Hutchinson (2019). A number of cognitive search strategies have been developed based on different reward functions of the information gain; for example, Ristic et al. (2017) Hutchinson, Oh & Chen (2018). A complete autonomous search system has been developed and tested in both indoor and outdoor environments with ground robots and unmanned aerial vehicles, respectively Hutchinson et al. (2019) Hutchinson, Liu & Chen (2018). In all the above studies, source localisation and search is considered as an estimation problem, or an exploration problem of unknown environment, so information theoretic approaches were advocated.

In this paper, instead of taking the information theoretic approach, we reconsider this autonomous search problem from a control theoretic perspective, and propose a new framework that is significantly different from the existing ones. Finding an unknown source is interpreted as *controlling* a mobile sensor platform approaching a target. That is, our goal is to design a control system that autonomously drives a robot equipped with chemical sensors to approach a target at an *unknown location* in an *unknown environment*. Different from a classic control setting, there is no reference trajectory or prescribed setpoint for the robot to follow. Indeed, the target location is unknown and the path to the target is to be defined in this autonomous search problem. To solve this problem, we formulate it as a stochastic model predictive control (MPC) problem with an undefined target location. It is shown that our approach is not only intuitive and promising, but also closely links with the dual

control concept Feldbaum (1960a) Feldbaum (1960b), Heirung et al. (2015), Mesbah (2018).

Different from a classic control setting, dual control considers that a control action affects not only future states of a system, but also uncertainty of its estimation Bar-Shalom & Tse (1974). In general, dual control is intractable and computationally expensive, therefore it has found very few practical applications (Wittenmark (2008)). It shall be noted that there is a significant difference between the existing dual control formulation and our approach. In the current dual control setting, dual effects consider the effect of control on systems (e.g. state estimation in stochastic control, e.g. Mesbah (2018), Simpkins et al. (2008), or parameter estimation in adaptive control, e.g. Wittenmark (2008), Filatov & Unbehauen (2000), Heirung et al. (2015)). The dual control effect we consider in this paper is about the probing effect on the operational environment or the target, rather than on the uncertainty of the dynamic system itself (e.g. state or parameters). We aim to drive the robot to a believed target position (to be defined later) but at the same time reduce the uncertainty associated with this believed target position, and build up a better understanding of the operational environment. This very difference gives us the ability to deal with the problem of controlling a robot approaching a target at an *unknown* location and in an *unknown* environment.

The problem we are trying to solve also belongs to a much broader class of problems arising in machine learning and artificial intelligence - the trade-off between *exploitation* and *exploration* in an unknown environment. In an unknown environment, *exploitation* makes best choice based on current information which is biased to what you have known and may lead to local results. In contrast, *exploration* aims to gather more information and make the best overall choice, but may lead to wasted efforts. In our view, IPP in robotics and autonomous systems focuses on exploration to increase information gain so to reduce uncertainty, while traditionally control engineering mainly focuses exploitation, i.e. making use of information to derive a suitable control action/plan. It is shown that for an autonomous search of the source of airborne substance release, our approach provides an optimal trade-off between exploitation and exploration in the sense of Bellman’s principle of optimality (1957) Bellman (1957).

This paper first formulates the autonomous search problem in an unknown environment into a control problem. The cost function is defined as the expected error between the robot’s future position and the *predicted* estimation of the target location (after taking into account information gain due to hypothesised future measurements). Since the true chemical concentration measurements at future positions of the robot are unknown, this predicted estimation is generated by a dispersion model with the current estimated source location as a

prior and randomly simulating sensor readings under described sensor characteristics. It is shown that the cost function actually consists of two parts: the first part is about driving the robot moving towards the *predicted* estimated source location with hypothesised measurements, while the second term quantifies the uncertainty level of the *predicted* estimated source location. In other words, the first term is related to *exploitation* by making good use of the estimated source location in generating control actions, whereas the second term is about *exploration* by using control to probe the unknown environment in order to reduce the uncertainty level of the location estimation and the belief of the environment. By minimising this cost function, we are able to optimally trade-off between these two effects, i.e. simultaneously driving the robot towards the source and reduce the uncertainty in source location estimation. It is pointed out that, to a large extent, the current MPC and information theoretic approach could be considered as special cases of our framework.

Bayesian inference is used to develop the control algorithm. The whole control framework consists of two steps: Bayesian estimation of the source location (and other associated environmental parameters) after a new chemical concentration measurement is taken, and the calculation of a control action by minimising the above cost function where Bayesian estimation is also embedded for calculating the predicted posteriors. However, it is difficult to implement this Bayesian estimation based control framework, particularly when a large number of source and environment related parameters are unknown. We resort to particle filtering for implementing the proposed control algorithm in both simulation and experimental tests.

There are 4 main contributions in this paper. 1). This paper formulates the autonomous search of airborne hazardous substance release as a new type of control problem from a control theoretic perspective. The main feature of this control problem is that the target is at an unknown location and there is also no predefined reference trajectory. In other words, the goal of the control system is well defined, i.e. finding the source, but the information about the target and its operational environment is *incomplete*. 2). Inspired by a dual control concept, a new framework is proposed to trade-off between making use of the current estimate to driving a robot towards the believed location of the target (exploitation) and reducing the uncertainty of the estimation (exploration), namely Dual Control for Exploration and Exploitation (DCEE). It is shown that optimal autonomous search is realised through this approach (in the sense of Bellman optimality Bellman (1957)). This work is extended to deal with an unknown environment. 3). Simulation and physical experiments have been developed to demonstrate the performance of the proposed DCEE. Numerical implementation of Bayesian estimation and online optimisation have been presented. The comparison with

alternative approaches including information theoretic approaches and classic MPC clearly shows superior performance and effectiveness of the proposed DCEE. 4). This work brings insights into a number of related areas such as information theoretic approaches (e.g. IPP, active sensing), dual control, exploration and exploitation, autonomous search, and MPC. It shows that the existing approaches are biased to (or only focus on) either exploitation or exploration so may lead to less efficient or poorer searching performance.

The remainder of the paper is organised as follows. Section 2 introduces the autonomous search problem for atmospheric dispersion of hazardous substances. A isotropic plume dispersion model is presented and chemical sensor behaviour with misdetection and sensor errors are also introduced. The control algorithm development is presented in Section 3. An one-step-ahead cost function is proposed and the control algorithm is derived with discussions. Particle filtering implementation of the proposed control algorithm is presented to facilitate real-time applications. This work is further extended in several directions in Section 4 including operating in an unknown environment (for example, wind speed and direction has a significant effect on the dispersion) and a multi-stage cost function. The relationship between DCEE and other relevant control/search strategies is also discussed in this Section to provide more insight. Simulation is conducted in Section 5 and comparison with IPP and MPC clearly demonstrates superior performance of the proposed framework. Furthermore, physical experiments have been conducted on a ground robot in an indoor environment in Section 6. This paper ends with conclusion in Section 7.

2 Autonomous search and its formulation

2.1 Agent modelling

Consider an autonomous agent (e.g. a robot or a UAV) is tasked to search an area for finding a possible source of airborne hazardous substance release. It is supposed that there is a lower level controller that drives the agent to following any planned path so the detailed dynamics of the agent are ignored for the path planning purpose. The behaviour of the agent is modelled as Hutchinson (2019)

$$\mathbf{p}_{k+1} = \mathbf{p}_k + \mathbf{u}_k, \quad (1)$$

where $\mathbf{p}_k \in R^3$ is the current position of the agent, $\mathbf{u}_k \in \mathcal{U}$ is the movement of the agent and \mathcal{U} is a set of admissible control actions.

2.2 Dispersion model and environment

The dispersion model of an airborne chemical release can be formulated as an Isotropic plume model Holmes

& Morawska (2006). In this model, the expected concentration at the robot position $\mathbf{p}_k = [p_{x,k}, p_{y,k}, p_{z,k}]^\top$ from a release source positioned at $\mathbf{s} = [s_x, s_y, s_z]^\top$ with a releasing rate of q_s is given by:

$$\mathcal{M}(\mathbf{p}_k, \Theta) = \frac{q_s}{4\pi\zeta_{s1}\|\mathbf{p}_k - \mathbf{s}\|} \exp\left[\frac{-\|\mathbf{p}_k - \mathbf{s}\|}{\lambda}\right] \times \exp\left[\frac{-(p_{x,k} - s_x)u_s \cos\phi_s}{2\zeta_{s1}}\right] \exp\left[\frac{-(p_{y,k} - s_y)u_s \sin\phi_s}{2\zeta_{s1}}\right], \quad (2)$$

where $\lambda = \sqrt{\frac{\zeta_{s1}\zeta_{s2}}{1+(u_s^2\zeta_{s2})/(4\zeta_{s1})}}$. A number of parameters are involved to characterise this dispersion, including the average particle lifetime ζ_{s2} , the diffusivity ζ_{s1} , the mean wind speed u_s and wind direction ϕ_s . These parameters of the source and local environment (e.g. wind field, temperature) are assumed to be unknown but may have certain prior information.

2.3 Sensor modelling

When an agent moves to a location, a point-wise measurement of chemical/biological concentration at the current location is taken by onboard chemical (or biological) sensors/detectors. Due to the nature of these sensors, an agent needs to hold at the location for a short period in order to get a reliable reading, which is referred to as the sampling interval. This is also why the detailed dynamics of an agent are not required in plan planning. Due to the power, size and payload of mobile platforms (e.g. small UAVs), only portable chemical detectors could be installed on the agent which quite often implies a limited accuracy and poor dynamic measurement capability. Furthermore, local turbulence in airflow has a significant effect on the distribution of concentration, causing quick perturbation in the dispersion, which causes intermittent sensor readings. Therefore, it is important to model sensor characteristics and the influence of the local environment on its behaviour. Both detection and non-detection events are considered due to the sparsity of the measurements and complication in local flow. A non-detection event may be caused by three hypothesised scenarios:

- The source is not present at all or present but not within range of the chemical detector. Any concentration measurement recorded is only a result of background and instrument noise.
- A source is present and in detectable concentrations but there is a non-detection as a result of intermittency caused by turbulence or a missed detection, typically exacerbated by the short sampling intervals of the agent.
- The source is present but the concentration (plus any addition from background and instrument noise) is below a pre-specified concentration threshold z_{thr} . The

threshold is set high enough to minimise false detections, whilst maintaining sufficient sensitivity Yee (2017).

In a detection event, the sensor reading consists of the true concentration and sensor noises. In summary, the sensor behaviour is modelled as

$$z_k = \begin{cases} \mathcal{M}(\mathbf{p}_k, \Theta) + v_k; & D = 1 \\ \bar{v}_k; & D = 0 \end{cases} \quad (3)$$

where D denote a detection event. $P_d = Pr\{D = 1\}$ is the detection probability encapsulating all the three hypothesised scenarios. A Poisson distribution is used to represent the non-detection event in this study with its non-detection probability as the sum of all these three scenarios. \bar{v}_k and v_k satisfying Gaussian distribution represent the noise in a non-detection event and detection event, respectively.

2.4 Autonomous search as a control problem

Traditionally the source search problem is referred to as an STE problem. This is to estimate the key parameter related to release so emergency responders or disaster management teams could be well informed. Broadly speaking, there are two approaches in developing information theoretic based search strategies to this problem: optimisation based approaches where gradients of the concentration are exploited and Bayesian approaches which is based on a probability framework Hutchinson et al. (2017). Various information metrics such as mutual information on entropy, Kullback-Leibler divergence and variance have been used as a reward function to develop cognitive search strategies such as ‘infotaxis’ Ristic et al. (2017), Ristic et al. (2016), ‘entrotaxis’ Hutchinson, Oh & Chen (2018). They provide a more reliable and robust solution with a shorter search time than the optimisation based search methods and traditional search methods such as ‘chemotaxis’ Adler (1966). In essence, these information theoretic methods mainly focus on exploration, i.e., to explore the unknown environment to find more information about sources, although some inherent trade-off behaviour between exploitation and exploration has been observed since it also drives the agent towards the target Hutchinson, Liu & Chen (2018).

Departing from the information theoretic approaches, we reformulate this problem from a control theoretic perspective. Basically, we consider the autonomous search problem as a process to driving the agent from its start location to a target. Once the target is reached, the mission is completed. However one challenge is that this is a non-conventional control problem since the location of the target is unknown and the environment is also unknown. It cannot be formulated in the traditional control framework such as a regulation or tracking problem

as there is no pre-defined reference trajectory. Another challenge from the control system point of view is that measurements are sporadic and intermittent. In actuality, for the majority of the search process, there is no sensor reading. To best of the authors' knowledge, there is no work in investigating this kind of autonomous search from a control theoretic perspective. We will show our new formulation provides a much better way to balance exploitation (make use of information collected) and exploration (reduce uncertainty in information by exploring the environment) as to lead to a substantially improved search performance.

3 Control algorithm development

We will start control algorithm development with only unknown parameters related to the source but will extend the work to deal with an unknown environment in Section 4. Consequently, the unknown or uncertain parameters we consider at this stage are the location of the source and its release rate, i.e. $\Theta = \begin{bmatrix} \mathbf{s}^\top & q_s \end{bmatrix}^\top$.

The goal of the control system for autonomous search is to drive the agent to the source at an unknown location. Up until now, a typical approach in the current control system setting would be to drive the robot towards the believed location derived from prior information and collect measurements. However, this would not be able to take into account the current level of uncertainty in the belief and how the future move could reduce it as illustrated in Section 4.2.2. This motivates our DCEE framework.

3.1 DCEE framework

When the source term Θ is unknown, a probability density function $p(\Theta_k)$ can be used to represent the belief of Θ at time k . Let \mathcal{Z}_k denote the vector of measurements collected at different locations by the robot up to time step k , i.e. $\mathcal{Z}_k := \{z_1(\mathbf{p}_1), z_2(\mathbf{p}_2), \dots, z_k(\mathbf{p}_k)\}$.

We define $\rho_{k|k} := p(\Theta|\mathcal{Z}_k)$ as the posterior distribution at time k . Now we consider any control input \mathbf{u}_k , which moves the robot to a new location where a future new measurement \hat{z}_{k+1} could be made at time $k+1$ (where \hat{z} is used to distinguish a predicted variable from a real variable). This future measurement can be considered as a random variable given the control input, i.e. $\hat{z}_{k+1} \sim p(\hat{z}_{k+1}|\mathbf{u}_k)$. We define a hypothetical posterior distribution $\hat{\rho}_{k+1|k}$ conditioned on this possible future measurement \hat{z}_{k+1} , i.e. $\hat{\rho}_{k+1|k} := p(\Theta|\mathcal{Z}_k, \hat{z}_{k+1})$. Therefore, the control input not only affects the future location of the robot but also affects its future belief about where the target might be.

Inspired by the above discussion, we would like to move the robot's future position closer to the predicted pos-

terior of the target's location, $\hat{\rho}_{k+1|k}$, conditional upon the control input \mathbf{u}_k . Given a possible future measurement \hat{z}_{k+1} induced by control input \mathbf{u}_k , the conditional cost function for approaching a target at an unknown location \mathbf{s} , can be formulated as follows

$$J(\mathbf{u}_k) = \mathbb{E}_\Theta \left[\mathbb{E}_{\hat{z}_{k+1}} \left[\|\mathbf{p}_{k+1|k} - \mathbf{s}\|^2 | \mathcal{Z}_k, \hat{z}_{k+1} \right] \right] \quad (4)$$

In this formulation, the control input \mathbf{u}_k not only determines the future robot position $\mathbf{p}_{k+1|k}$, but also affects the possible future measurement $\hat{z}_{k+1}(\mathbf{p}_{k+1|k})$. Let $\Theta_{k+1|k} = [\mathbf{s}_{k+1|k}^\top, q_{s,k+1|k}]^\top$ denote the estimate of the source location and the release rate at time k with the measurements \mathcal{Z}_k and the virtual measurement $\hat{z}_{k+1}(\mathbf{p}_{k+1|k})$. It is clear that the move \mathbf{u}_k affects the predicted posterior $\hat{\rho}_{k+1|k} := p(\Theta_{k+1|k})$ of the source location and the release rate. Taking the expectation with respect to the future measurement rewards the exploration effect in reducing the level of uncertainty in unknown source parameter estimation, as we will show later.

The optimisation problem for DCEE can be formulated as follows:

$$\min_{\mathbf{u}_k} J(\mathbf{u}_k) = \min_{\mathbf{u}_k} \mathbb{E}_\Theta \left[\mathbb{E}_{\hat{z}_{k+1}} \left[\|\mathbf{p}_{k+1|k} - \mathbf{s}\|^2 | \mathcal{Z}_{k+1|k} \right] \right] \quad (5a)$$

subject to

$$\mathbf{p}_{k+1|k} = \mathbf{p}_k + \mathbf{u}_k \quad (5b)$$

$$\mathbf{u}_k \in \mathcal{U} \quad (5c)$$

where $\mathcal{Z}_{k+1|k} := \{\mathcal{Z}_k, \hat{z}_{k+1}\}$.

We define the nominal estimated source location as the mean of the posterior distribution of the source location estimation. Similarly, the nominal predicted estimation of the source location $\bar{\mathbf{s}}_{k+1|k}$ is defined as the mean of the predicted distribution, i.e. $\hat{\rho}_{k+1|k}$, with $\mathcal{Z}_{k+1|k}$, which is given by

$$\bar{\mathbf{s}}_{k+1|k} := \mathbb{E} [\mathbf{s}_{k+1|k}] = \mathbb{E} [\mathbf{s} | \mathcal{Z}_{k+1|k}] \quad (6)$$

With $\bar{\mathbf{s}}_{k+1}$, we define $\tilde{\mathbf{s}}_{k+1|k}$ conditional on $\mathcal{Z}_{k+1|k}$ as $\tilde{\mathbf{s}}_{k+1|k} = \mathbf{s} - \bar{\mathbf{s}}_{k+1|k}$. Therefore, the cost function (5) can be reformulated as

$$J(\mathbf{u}_k) = \mathbb{E}_{\Theta, \hat{z}_{k+1}} \left[\|\mathbf{p}_{k+1|k} - \bar{\mathbf{s}}_{k+1|k} - \tilde{\mathbf{s}}_{k+1|k}\|^2 | \mathcal{Z}_{k+1|k} \right] \quad (7)$$

Theorem 1 *The cost function of DCEE defined in (4) is equivalent to the following cost function*

$$J(\mathbf{u}_k) = \|\mathbf{p}_{k+1|k} - \bar{\mathbf{s}}_{k+1|k}\|^2 + P_{k+1|k} \quad (8)$$

where $P_{k+1|k} = \mathbb{E}_{\Theta, \hat{z}_{k+1}} \left[\tilde{\mathbf{s}}_{k+1|k}^T \tilde{\mathbf{s}}_{k+1|k} | \mathcal{Z}_{k+1|k} \right]$ is the predicted covariance matrix of $\mathbf{s}_{k+1|k}$.

PROOF. Expanding the square and collecting terms of the right hand side of (7) gives

$$\begin{aligned}
J(\mathbf{u}_k) = & \mathbb{E} \left[\|\mathbf{p}_{k+1|k} - \bar{\mathbf{s}}_{k+1|k}\|^2 | \mathcal{Z}_{k+1|k} \right] \\
& - 2\mathbb{E} \left[\tilde{\mathbf{s}}_{k+1|k}^T (\mathbf{p}_{k+1|k} - \bar{\mathbf{s}}_{k+1|k}) | \mathcal{Z}_{k+1|k} \right] \quad (9) \\
& + \mathbb{E} \left[\tilde{\mathbf{s}}_{k+1|k}^T \tilde{\mathbf{s}}_{k+1|k} | \mathcal{Z}_{k+1|k} \right]
\end{aligned}$$

Given that both $\mathbf{p}_{k+1|k}$ and $\bar{\mathbf{s}}_{k+1|k}$ are deterministic and that $\mathbb{E} \left[\tilde{\mathbf{s}}_{k+1|k}^T | \mathcal{Z}_{k+1|k} \right] = 0$, results in (8) can be concluded.

Remark 2 *The cost function in the form of (8) clearly reveals the dual control effect of our approach. The first term in the cost function (8) drives the robot to the estimated location of the source, which is related to exploitation. The second term is about the level of uncertainty of the estimated target location, captured by the predicted covariance of the source location in the next time step. The influence of future control action on both the distance to a believed target location and the current information uncertainty is quantified. It is quite intuitive and natural. Optimising this cost function over admissible control actions gives the best next move that balances the probing effect and performing certain task.*

Remark 3 *This optimal control problem gives the best strategy in trade-off between exploration and exploitation so it is the optimal autonomous search strategy for this problem in the sense of the principle of optimality (Bellman, 1957). The existing IPP based autonomous search is only concerned about the second term while the classic MPC approach only considers the first term. In many areas, there is a big challenge in how to balance exploration and exploitation, particularly in artificial intelligence, optimisation and decision making for complicated problems. Normally weights on the cost/reward functions have to be introduced to balance these two effects. Sometimes, in order to reflect the dual effect, a related term is also artificially added to the cost function Heirung et al. (2015). There is no such a requirement in choosing weights to trade-off in our formulation. They are derived from a physically meaningful cost function in (4). The balance between them is naturally embedded. More discussion about this approach and the existing work will be made in Section 4.*

Bayesian estimation plays a key role both in the inference engine for estimating the parameters related to the source and in the planning loop for calculating predicted posteriors. It is implemented using using particle filtering while the control a finite set of one-step action, i.e. $\mathbf{u}_k \in \mathcal{U} := \{\uparrow, \downarrow, \leftarrow, \rightarrow, \swarrow, \searrow, \nearrow, \nwarrow\}$ Hutchinson, Liu & Chen (2018). In the following sections, we will discuss the implementation issues related to this framework.

3.2 Implementation of Bayesian estimation

Bayesian estimation plays a key role both in the inference engine for estimating the parameters related to the source and in the planning loop for calculating predicted posteriors.

The conditional probability of the source terms can be obtained via recursive Bayesian estimation, such that

$$p(\Theta_k | \mathcal{Z}_k) = \frac{p(z_k | \Theta_k) p(\Theta_k | \mathcal{Z}_{k-1})}{p(z_k | \mathcal{Z}_{k-1})} \quad (10)$$

where

$$p(z_k | \mathcal{Z}_{k-1}) = \int p(z_k | \Theta_k) p(\Theta_k | \mathcal{Z}_{k-1}) d\Theta_k. \quad (11)$$

The initial prior distribution $p(\Theta_0 | \mathcal{Z}_0) = p(\Theta_0)$ is assumed to be given. In this work, the parameters associated with the source term are assumed to be unknown but fixed. The likelihood function in the Bayesian estimation $p(z_k | \Theta_k)$ is determined by using the dispersion model (2) and the sensor model (3).

Given the nonlinear nature of this Bayesian estimation problem, a particle filtering approach is used to approximate the Bayesian estimation. In the particle filter, the posterior distribution from Eq. (10) is approximated by a set of weighted random samples $\{\Theta_k^{(i)}, w_k^{(i)}\}_{i=1}^N$ such that

$$p(\Theta_k | \mathbf{z}_{1:k}) \approx \sum_{i=1}^N w_k^{(i)} \delta(\Theta_k - \Theta_k^{(i)}), \quad (12)$$

where $\delta(\cdot)$ is a Dirac delta function, $\Theta_k^{(i)}$ is a sample representing a potential source term realisation and $w_k^{(i)}$ is the corresponding normalised weighting such that $\sum_{i=1}^N w_k^{(i)} = 1$.

The process of recursively calculating the posterior distribution at sampling instance k is summarised in Algorithm 1 and more details can be found in Hutchinson, Liu & Chen (2018).

A similar process is employed in calculating predicted posteriors with hypothesised measurements under a candidate control action but with some simplifications to reduce online computational load.

3.3 Implementation of optimisation

To reduce the computational load, in this paper we choose the optimal control input from a finite set of one-step action, i.e. $\mathbf{u}_k \in \mathcal{U} := \{\uparrow, \downarrow, \leftarrow, \rightarrow, \swarrow, \searrow, \nearrow, \nwarrow\}$. The step size can be determined based on the operation environment.

Require: prior samples: $\{\Theta_{k-1}^{(i)}, \omega_{k-1}^{(i)}\}_{i=1}^N$; sensor measurement z_k at location \mathbf{p}_k ;

- 1: **for** $i = 1, 2, \dots, N$ **do**
- 2: draw sample $\Theta_k^{(i)} \sim q(\Theta_{k-1}^{(i)})$
- 3: Assign weight $\bar{w}_k^{(i)} = w_{k-1}^{(i)} \cdot \frac{p(z_k | \Theta_k^{(i)}) p(\Theta_k^{(i)} | \Theta_{k-1}^{(i)})}{q(\Theta_k^{(i)} | \Theta_{k-1}^{(i)}, \mathbf{z}_{1:k})}$
- 4: **end for**
- 5: normalise weight $w_k^{(i)} = \bar{w}_k^{(i)} / (\sum_{i=1}^N \bar{w}_k^{(i)})$
- 6: calculate effective sample size $N_{eff} = 1 / \sum_{i=1}^N (w_k^{(i)})^2$
- 7: **if** $N_{eff} < N_T$ **then**
- 8: resample $\{\Theta_k^{(i)}, \omega_k^{(i)}\}_{i=1}^N$
- 9: apply a MCMC (Monte Carlo Markov Chain) move
- 10: **end if**

Ensure: posterior samples: $\{\Theta_k^{(i)}, \omega_k^{(i)}\}_{i=1}^N$

Algorithm 1. Particle filter for source parameter estimation

The possible future measurements for a given control input \mathbf{u}_k can be generated using the agent model (1), the dispersion model (2) and the sensor model (3) with $\Theta_k \sim \rho_k | k$. Based on (12), a set of samples to represent the distribution of $\rho(k|k)$ are generated. For each sample Θ^i , the dispersion model (2) is run to generate the chemical concentration at the new agent location $\mathbf{p}_{k+1|k}$ due to the move \mathbf{u}_k . A set of measurements \hat{z}_{k+1} are obtained by randomising the measurements using the sensor model (3) with the described noise characters v_k and \bar{v}_k . The optimal control action is selected as the one minimising the cost (4) or (8).

4 Extension and relationships with other methods

This section first extends the basic DCEE framework presented in Section 3 in a number of directions (including the incorporation of an unknown environment). The relationships with several other methods will then be discussed. It is shown that DCEE covers the cost/reward functions in both classic MPC and IPP, and is better fitting with autonomous search and control system design for autonomous systems in general.

4.1 Extension

4.1.1 Unknown environment

We have only considered unknown parameters related to the target such as its location in Section 3. In real operation, the operation environment is also more likely to be partially known. For example, we may only know wind direction and speed within a certain range from weather forecast or meteorologic data. We now extend our DCEE algorithm to cope with uncertainty in both the source and environment. Physical parameters in the dispersion model depend on the type of chemical release and the

environment. We now consider that all the related parameters in the Isotropic plume model (2) are unknown but with certain prior information. More specifically, in addition to the position and release rate of the source, the parameters to be estimated during the search process also include wind direction ϕ_s , wind speed u_s , the diffusivity ζ_{s1} , and the average particle lifetime ζ_{s2} . That is, $\Theta = [\mathbf{s}, q_s, \phi_s, u_s, \zeta_{s1}, \zeta_{s2}]^T$. These parameters are related to the environment (e.g. wind field, temperature) and specific chemicals. It is assumed that all the parameters are unknown but constant during the search process.

By the virtue of the Bayesian estimate framework, we are able to make use of any prior information of these unknown parameters in our estimation. This will improve the accuracy of the parameter estimation of the target. There are no major changes to the structure of Algorithm 1 except the computational burden. This extended version of DCEE will be implemented and tested in real-time search operation on a robot in Section 6.

4.1.2 Multistage DCEE

For the convenience of illustrating the basic concept of DCEE, only a one-step-ahead cost function is considered in Section 3. It is straightforward to extend it to DCEE with a multistage cost function as

$$J(\mathbf{U}_k) = \mathbb{E}_{\Theta} [\mathbb{E}_{\hat{z}_{k+1}, \dots, \hat{z}_{k+N}} [\|\mathbf{p}_{k+N|k} - \mathbf{s}\|^2 | \mathcal{Z}_{k+N|k}]] \quad (13)$$

where $\mathbf{U}_k = [\mathbf{u}_k, \dots, \mathbf{u}_{k+N-1}]$, $\mathcal{Z}_{k+N|k} = [\mathcal{Z}_k, \hat{z}_{k+1}(\mathbf{p}_{k+1|k}), \dots, \hat{z}_{k+N}(\mathbf{p}_{k+N|k})]$, where $\mathbf{p}_{k+i|k}$, $i = 1, \dots, N$, are the predicted locations of the agent under the control sequence \mathbf{U}_k with the agent model (1). This implies that the conditional distance between the robot position and the location of the source in the next N th step is of interest. In other words, the predicted posterior of N th step with the hypothesised future N step measurements under possible control sequence \mathbf{U}_k is used to improve our autonomous search strategy. Then this cost function is minimised over the set of admissible control \mathbf{U}_k to give the optimal control sequence \mathbf{U}_k^* but only the first move \mathbf{u}_k^* is implemented and the process is repeated in a receding horizon fashion. Although the multistage strategy may improve the performance of the autonomous search further, it brings a significant increase of computation burden with it.

4.2 Relationship with other methods

4.2.1 Information theoretic approaches

As briefly discussed in Section 3, information theoretic approaches have been used to develop autonomous search methods where the search process is treated as an information gathering problem. More specifically, the aim of motion planning is to reduce the uncertainty of

the estimation of the target location and other unknown environmental factors. Therefore, the next move of the robot is determined to maximise a reward function related to information. Mutual information on entropy, Kullback–Leibler divergence and other metrics are used to measure the uncertainty of the information and the corresponding search schemes have been presented. More specifically, Entrotaxis enforces the maximum entropy sampling principle which dictates the agent moves to the position that is least certain Hutchinson, Oh & Chen (2018). Instead of reducing predicted entropy, infotaxis aims to minimise the predicted posterior variance of the source location Ristic et al. (2016) Ristic et al. (2017). This directly links to exploration effect of DCEE, i.e. the second term of (8).

Since Entrotaxis exhibits a slightly better performance than Infotaxis Hutchinson, Oh & Chen (2018), we implement Entrotaxis as a benchmark of IPP for comparison with DCEE in our simulation and experimental study.

4.2.2 Model Predictive Control (MPC)

As discussed in Introduction, currently there is little work looking into autonomous search from a control engineering perspective. If following our idea of considering the search as a process to drive the agent towards the source, it could be formulated in a *classic* stochastic MPC framework as

$$J(\mathbf{u}_k) = \mathbb{E}_\Theta [\|\mathbf{p}_{k+1|k} - \mathbf{s}\|^2 | \mathcal{Z}_k] \quad (14)$$

where only the measurements up to the current time are used in estimating the location of the source. A simple interpretation of this cost function is that a right strategy is to drive the robot towards the best estimate of the source location with all the information we have so far.

In a similar fashion, this cost function could be rewritten as

$$J(\mathbf{u}_k) = \|\mathbf{p}_{k+1|k} - \bar{\mathbf{s}}_{k|k}\|^2 + P_{k|k} \quad (15)$$

where $\bar{\mathbf{s}}_{k|k}$ is the mean estimated location of the source at time k conditional on \mathcal{Z}_k , and $P_{k|k}$ is the covariance matrix of the estimation $\mathbf{s}_{k|k}$ defined as $P_{k|k} = \mathbb{E} [\tilde{\mathbf{s}}_{k|k}^T \tilde{\mathbf{s}}_{k|k}]$.

Since the control \mathbf{u}_k does not affect the current estimation so its uncertainty, the second term in (15) is not relevant. The control action only affects the future position of the agent. So broadly speaking, the cost function (14) in classic stochastic MPC corresponds to the exploitation effect of DCEE, i.e. the first term of (8), i.e., to drive the robot to the best estimate location of the source based on all the current available information. This is not surprising since control is about to make use of information to take action. During the process of driving towards the believed source location, the robot also collects new measurements so its belief about the source

location is updated at each step with these new measurements. However, this is *accidental* or *passive* learning to reduce the uncertainty. It shall be highlighted that recently there are works in MPC with *active* learning using dual control effort that will be discussed in the next section Mesbah (2018).

In summary, the MPC and information theoretic approaches can be considered as special cases of the the proposed DCEE: the former is biased to exploitation while the latter exploration. The simulation and experimental comparisons between MPC, Entrotaxis and DCEE will be presented in Section 5 and 6.

4.2.3 Dual control of uncertain systems

According to Bar-Shalom and Tse Bar-Shalom & Tse (1974), a control input is said to have dual control effect if it can affect, with nonzero probability, at least one r -th-order central moment of a state variable ($r > 1$). In a series of seminal papers, Feldbaum first recognised when controlling an uncertain system, control inputs have a dual effect; i.e. not only a directing effect on system states but also a probing effect on system uncertainty Feldbaum (1960a) Feldbaum (1960b) Feldbaum (1961a) and Feldbaum (1961b). Most of the work in dual control is devoted to a system with unknown or immeasurable states. It has been shown in the so-called *separation principle* of control engineering, that a state estimator and a controller could be designed separately does not hold for most of systems, except, i.e. Linear Quadratic Gaussian (LQG). Various works have been presented to exploit the dual effect of control inputs. However, although it is very conceptually attractive, dual control is computationally intractable even for moderately sized systems Wittenmark (2008). Recently triggered by the advances in MPC and active learning, there is a renewed interest in dual control in the context of stochastic MPC with active learning Mesbah (2018). The probing effect of control inputs is investigated to actively learn the system state so to reduce the level of uncertainty in state estimation in an MPC setting.

Another area concerning the dual effect of control signals is adaptive dual control, e.g. Wittenmark (2008), Filatov & Unbehauen (2000), Heirung et al. (2017). Most of the adaptive control methods are developed based on the Certainty Equivalence (CE) principle where a control law is developed by treating the estimated parameters as the true parameters and a parameter updating mechanism is adopted to online estimate the unknown parameters. However, except in very special cases, control signal affects not only the output but also the quality of the parameter estimation. This separation principle implied by CE does not hold for most of adaptive control. There is also an intrinsic conflict between them since one side, information about the process increases with the level of perturbation and on the other side, the

system output shall vary as little as possible. Adaptive dual control was proposed to address these issues; see Wittenmark (2008), Filatov & Unbehauen (2000).

To address the computationally intractable challenge of dual control, a number of methods have been proposed to find suboptimal solutions. Broadly speaking, there are two types of MPC with dual control effect: indirect dual control and explicit dual control. The former is to approximate the dynamic programming Bertsekas (2005) while the latter adds the probing activity by deliberately modifying the cost function. A number of suboptimal approaches have been developed, but so far there are only a few reports about the applications of adaptive dual control or MPC with active learning Mesbah (2018).

Although DCEE is similar to these dual control methods that explicitly exploit the dual effects of control signals, there are clear differences between them. In the MPC with active learning or other dual control methods, the dual effect concerns the probing effect on a control system itself (i.e. unknown states or parameters).

Rather than about a control systems itself, DCEE exploits the probing effect on unknown parameters of the tasks a control system is designed to perform (e.g. the source location in this study), or an unknown environment that a control system operates in. So it goes beyond the traditional use of dual control effects, e.g. Wittenmark (2008), Filatov & Unbehauen (2000), Heirung et al. (2015), Heirung et al. (2017), Mesbah (2018). DCEE aims to increase the level of autonomy, or deal with control problems where only high level specifications (goals) are defined. The control system itself has to find more information about the goals and the environment in order to perform the tasks by probing the environment.

5 Simulation study

To test the efficacy of the proposed solution, a simulation study is performed which compares DCEE against the competitive solutions, most notably, MPC described in Section 4.2.2 and IPP. It shall be highlighted that there is no existing work about applying MPC in autonomous search which itself also forms a new contribution of this paper. For IPP, we implement a typical search algorithm, Entrotaxis proposed in Hutchinson, Oh & Chen (2018). In all the simulation studies and the experiment studies in the next section, we keep everything the same except the cost/reward function. By testing in this manner, the joint exploitation/exploration characteristics of DCEE can be thoroughly evaluated.

5.1 Simulation Scenario

The simulation is performed under the setting of a single source release within an open bounded environment.

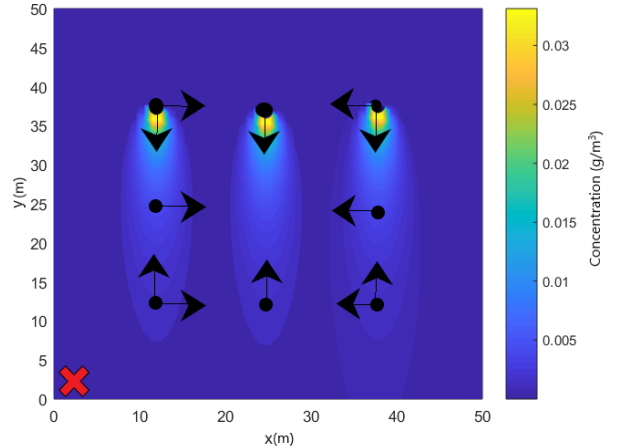


Fig. 1. 12 sources differing in $\mathbf{s}_x, \mathbf{s}_y$ and their respective wind directions ϕ_s (indicated by the black arrow) within the bounded search domain. Three example plumes are shown with a coloured scale. Red 'X' represents the UAV's start location.

The agent deployed to search the area is a UAV. The simulated source is modelled with a constant release rate using the IP model (2). The gas sensing model (3) outlined in Section 2 is used for taking simulated measurements from the IP ground truth model. To make the results meaningful, in total 360 runs across 12 source configurations have been conducted (see Fig. 1).

Prior values of Θ are initialised using simulation model parameters fitted with a PDF to encompass initial uncertainty. The prior θ_{xy} area is set uniformly as the entire domain of Fig. 1. UAV operational parameters are defined as a 2m sampling step size with directions at 45° intervals. At each potential sampling location, 40 measurement predictions are made. In this simulation study, we follow Section 3 that only the location $(\mathbf{s}_x, \mathbf{s}_y, \mathbf{s}_z)$ and the release rate q_s of the source are considered to be unknown.

5.2 Results and discussion

To measure the performance of each search method over time, the RMSE of estimated $(\mathbf{s}_x, \mathbf{s}_y, \mathbf{s}_z)$ against that of the true source is recorded at each sampling event. The average RMSE for each method across all sources is shown in Figure 2.

Initially, MPC demonstrates a comparable gradient of RMSE reduction as DCEE. This is because in MPC, the location estimation is given by the mean of the probability distribution $p(\Theta_k)$, i.e. $\bar{\mathbf{s}}_{k|k}$, and at the beginning, a uniform distribution of the possible locations of the source is used. Its mean is therefore located in the middle of the search area so naturally MPC drives the UAV towards the middle of the area. Given a plume intersects with this path, then early sampling can occur.

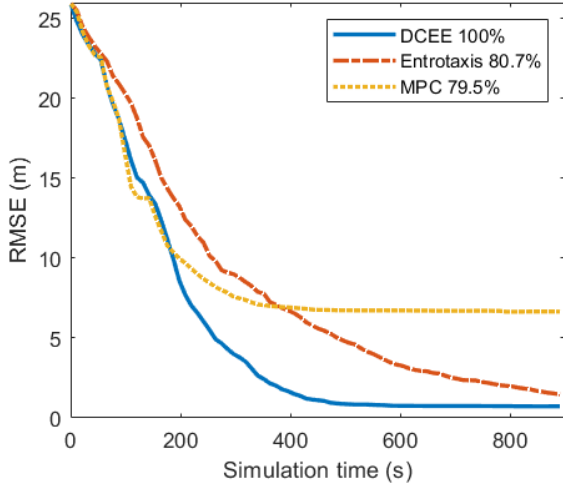


Fig. 2. Average RMSE over time for all 120 simulation runs. Also pictured, each method’s source acquisition rate and plume acquisition rate respectively.

Evidently, it accidentally occurs that early sampling is achieved with MPC, not intentionally. Entrotaxis, however, shows a shallower gradient at the beginning due to the tendency to travel perpendicular to wind direction in order to reduce the entropy of the estimation in the particle filter. At approximately 200s, MPC fails to continue to adequately resolve the source whereas DCEE continues gathering useful data and convergence to a RMSE of 0.5m at approximately 500s. Entrotaxis also manages to converge to a similar accuracy but at the end of the UAV budgeted flight time of 900s.

The source acquisition rate and the plume acquisition rate are also recorded in Fig. 2. For a control method to be defined as having acquired the source, the final RMSE of an individual run must be below 3m. The UAV is deemed to have acquired the plume when there is a sampling event yielding a reading greater than the minimum threshold h_{thr} of the sensor model. These are both important metrics to consider during evaluation as although performance is primarily dictated by an ability to reduce the distance to the true source quickly, the success rate of finding the plume (exploration) and source (exploitation) gives a greater insight as to operational qualities of each control method. Both DCEE and Entrotaxis achieve 100% plume acquisition (therefore good exploration characteristics) but MPC only achieves 80%. DCEE achieves 100% success in source acquisition rate in all 120 simulations (thus also showing good exploitation characteristics), compared to 80% of Entrotaxis and MPC. The lack of plume acquisition for MPC is caused by local minimum during the search due to a large mismatch between the prior information and the ground truth of certain source configurations.

It is clear that DCEE outperform these two methods significantly in all aspects. This is because DCEE is able to

generate the best autonomous search strategy in trading off between exploration and exploitation in all the scenarios. It shall be mentioned that DCEE and Entrotaxis has a similar computational burden but MPC has a much less computational demand than these two other methods.

6 Experimental Study with a ground robot

To further validate the findings in the simulation, these three autonomous search algorithms are implemented in an indoor environment using a ground robot. They are tested and evaluated through a large number of experiments with the system setup remaining largely unchanged (changes are only made in appropriate operational parameters due to the size of the search area and the agent model from UAV of the third order to the ground robot of a second order). However, one major difference between the simulation and experimental tests is that all parameters associated with the sources and the environment are now treated as unknown and their estimation uncertainty is considered through dual control effect, i.e. the algorithm presented in Section 4 is applied. More specifically, the parameters estimated online in the search process consist of the source location (s_x, s_y) , the release rate q_s , wind direction ϕ_s , wind speed u_s , the diffusivity ζ_{s1} , and the average particle lifetime ζ_{s2} . The source of the airborne release used in the experiments is a small beaker of acetone, which is agitated by the air flow of two fans that create a wind field of $\sim 1.5\text{m/s}$ across the test area. A fresh air intake as well as exhaust fans are used to ensure that the plume is stable over the 60 runs for a fair comparison.

An equipped ground robot is used to collect concentration measurements of the acetone vapour (see Fig. 4). Since one of the aims of the experimental study is to closely recreate real-world deployment, it is also not assumed that localisation and low-level control/planner are known or perfect. Mapping and localisation are performed using the popular simultaneous localisation and mapping (SLAM) technique of Hector SLAM Kohlbrecher et al. (2011), whilst the low-level planner is achieved by using a dynamic window approach (DWA) algorithm Fox et al. (1997).

The main functional difference between the simulation and experiments is that there is greater model uncertainty between the dispersion model used in Bayesian estimation and the real dispersion event. Whereas the estimation algorithm could guarantee convergence in the simulations (due to using the same plume model, certain localisation and perfect control), the source parameters in the experimental scenario are unknown and values have to be estimated during the autonomous search. These factors make the online estimation aspect much more challenging. Furthermore, since the DCEE strategy is deeply coupled with the Bayesian estimation to

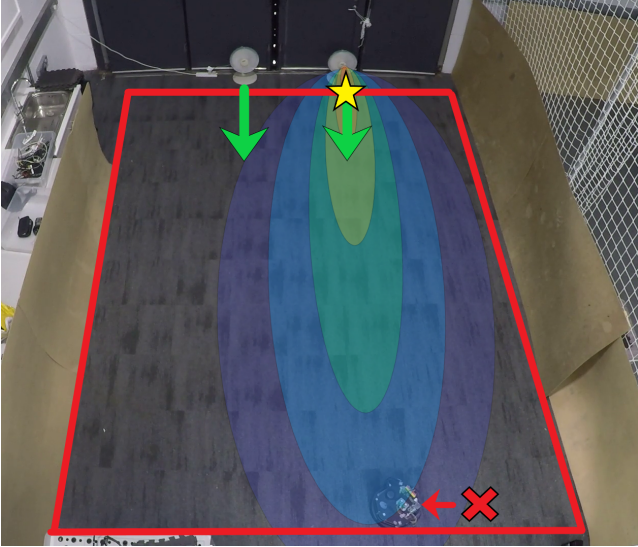


Fig. 3. Experimental scenario. Red boundary shows the prior search boundary limits, green arrows show the wind field source and direction, the yellow star marks the acetone source location and the red X marks the downwind start location for the search.

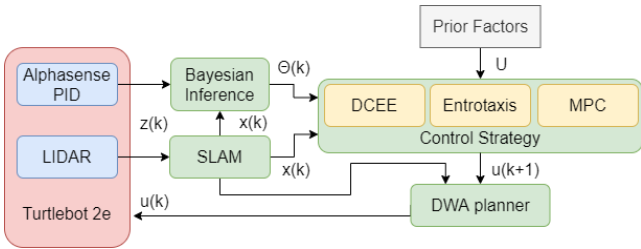


Fig. 4. Data flow for the experimental testing of autonomous search algorithms using a ground robot

decide the best next move based on the level of estimation uncertainty, it is important to investigate the performance of DCEE under dispersion model mismatching and unmodelled dynamics in real applications. Given these factors (plus the new uncertainty in localisation, mapping and low-level planning/control actions), the experiment provides a challenging stage upon which any weaknesses of DCEE will become apparent. Successful application in the experimental trial will also show that the DCEE framework is suitable for real-world deployment on autonomous search systems.

Since the indoor experiment proposed is relatively constricted in the number of configurations that can be tested, two start locations of the robot are explored. The first is downwind of the source but close to the plume and the second is situated away from the plume in the perpendicular wind direction. Testing these two extremes will show the relative merits and demerits of each control method and also test the versatility of DCEE. Each location is tested 10 times per control method (60 total runs).

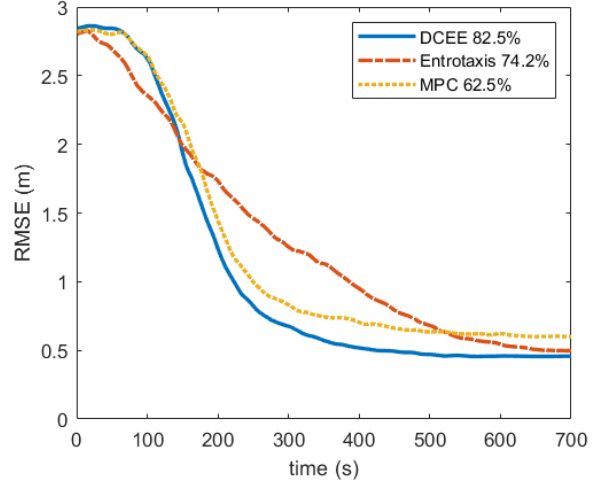


Fig. 5. Average RMSE over time for each control method across both experimental source configurations.

6.1 Results

The RMSE over time combined from the two source configurations are shown in Figure 5. The trends echo those seen in the simulation tests, with DCEE showing its ability to quickly localise and converge on the source. Due to the constricted nature of the test (and therefore certainty of sampling the plume), MPC shows more competitive performance however even in this small scenario, it still lacks the ability fully converge to the same degree as DCEE and Entrotaxis.

The DCEE framework has now been proven to be suitable for real-world deployment on autonomous search systems.

7 Conclusion

For an autonomous search problem, a control theoretic approach is proposed in this paper. Inspired by the dual control concept, DCEE is able to take into account both exploitation and exploration effect of control/decision/planning actions. It is suitable for designing high level control systems for a system operating in an unknown environment and (or) with unknown parameters associated with a task (i.e. unknown location of a source in this paper). Simulation and experiment results show this new framework outperforms classic MPC, a popular method in control engineering, and IPP, an information theory approach widely used in robotics and autonomous systems. For the specific autonomous search of hazardous sources, it is shown that DCEE maintains an optimal balance (in the sense of the principle of optimality (Bellman, Bellman (1957))) between the probing activity and control activity of control inputs, which are naturally in conflict. In the DCEE framework, stochastic MPC, active learning,

IPP, exploitation and exploration, and autonomous search are closely related and integrated. Much more work is required in further exploring and exploiting this new framework.

The level of autonomy of a system can be measured in terms of the level of complexity of tasks it is able to perform and the level of uncertainty it is able to cope with Antsaklis (2020). Hence it could be argued that any autonomous system shall be able to perform certain tasks and achieve defined goals based on its belief, but also update its belief and reduce the level of uncertainty by actively exploring the environment. Consequently any decision making/planning/control action shall have (dual) effect in these two aspects. The DCEE framework proposed in this paper provides a promising vehicle in realising these desirable properties for autonomous systems. It will inspire more research into developing similar goal-oriented control systems.

Acknowledgements

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) Established Career Fellowship “Goal-Oriented Control Systems: Disturbance, Uncertainty and Constraints” under the grant number EP/T005734/1.

References

- Adler, J. (1966), ‘Chemotaxis in bacteria’, *Science* **153**(3737), 708–716.
- Antsaklis, P. (2020), ‘Autonomy and metrics of autonomy’, *Annual Reviews in Control* **49**, 15–26.
- Bar-Shalom, Y. & Tse, E. M. (1974), ‘Dual effect, certainty equivalence, and separation in stochastic control’, *IEEE Transactions on Automatic Control* **19**, 494 – 500.
- Bellman, R. E. (1957), *Dynamic Programming*, Princeton University Press, New Jersey.
- Bertsekas, D. P. (2005), ‘Dynamic programming and suboptimal control: a survey from adp to mpc’, *European Journal of Control* **11**, 310 – 334.
- Bieringer, P. E., Rodriguez, L. M., Vandenberghe, F., Hurst, J. G., Bieberbach, G., Sykes, I., Hannan, J. R., Zaragoza, J. & Fry, R. N. (2015), ‘Automated source term and wind parameter estimation for atmospheric transport and dispersion applications’, *Atmospheric Environment* **122**, 206–219.
- Boström-Rost, P. (2019), *On Informative Path Planning for Tracking and Surveillance*, Vol. 1838, Linköping University Electronic Press.
- Feldbaum, A. A. (1960a), ‘Dual control theory. i’, *Automation Remote Control* **21**, 874 – 880.
- Feldbaum, A. A. (1960b), ‘Dual control theory. ii’, *Automation Remote Control* **21**, 1033 – 1039.
- Feldbaum, A. A. (1961a), ‘Dual control theory. iii’, *Automation Remote Control* **22**, 1 – 12.
- Feldbaum, A. A. (1961b), ‘Dual control theory. iv’, *Automation Remote Control* **22**, 109 – 121.
- Filatov, N. M. & Unbehauen, H. (2000), ‘Survey of adaptive dual control methods’, *IEE Proc Control Theory Application* **147**, 118 – 128.
- Fox, D., Burgard, W. & Thrun, S. (1997), ‘The dynamic window approach to collision avoidance’, *IEEE Robotics & Automation Magazine* **4**(1), 23–33.
- Heirung, T. A. N., Foss, B. & Ydstie, B. E. (2015), ‘Mpc-based dual control with online experiment design’, *Journal of Process Control* **32**, 64–76.
- Heirung, T. A. N., Ydstie, B. E. & Foss, B. (2017), ‘Dual adaptive model predictive control’, *Automatica* **80**, 340–348.
- Holmes, N. S. & Morawska, L. (2006), ‘A review of dispersion modelling and its application to the dispersion of particles: an overview of different dispersion models available’, *Atmospheric environment* **40**(30), 5902–5928.
- Hutchinson, M. (2019), On the use of autonomous unmanned vehicles in response to hazardous atmospheric release incidents, PhD thesis, Loughborough University.
- Hutchinson, M., Liu, C. & Chen, W.-H. (2018), ‘Information-based search for an atmospheric release using a mobile robot: Algorithm and experiments’, *IEEE Transactions on Control Systems Technology* **27**(6), 2388–2402.
- Hutchinson, M., Liu, C., Thomas, P. & Chen, W.-H. (2019), ‘Unmanned aerial vehicle-based hazardous materials response: Information-theoretic hazardous source search and reconstruction’, *IEEE Robotics & Automation Magazine* .
- Hutchinson, M., Oh, H. & Chen, W.-H. (2017), ‘A review of source term estimation methods for atmospheric dispersion events using static or mobile sensors’, *Information Fusion* **36**, 130–148.
- Hutchinson, M., Oh, H. & Chen, W.-H. (2018), ‘Entro-taxis as a strategy for autonomous search and source reconstruction in turbulent conditions’, *Information Fusion* **42**, 179–189.
- Kohlbrecher, S., Von Stryk, O., Meyer, J. & Klingauf, U. (2011), A flexible and scalable slam system with full 3d motion estimation, in ‘2011 IEEE international symposium on safety, security, and rescue robotics’, IEEE, pp. 155–160.
- Mesbah, A. (2018), ‘Stochastic model predictive control with active uncertainty learning: A survey on dual control’, *Annual Reviews in Control* **45**, 107 – 117.
- Platt, N. & Deriggi, D. (2010), Comparative investigation of source term estimation algorithms using fusion field trial 2007 data, in ‘8th Conference on Artificial Intelligence Applications to Environmental Sciences at AMS Annual Meeting, Atlanta, GA, Jan’, Vol. 1.
- Ristic, B., Angley, D., Moran, B. & Palmer, J. (2017), ‘Autonomous multi-robot search for a hazardous source in a turbulent environment’, *Sensors* **17**(4), 918.
- Ristic, B., Morelande, M. & Gunatilaka, A. (2010), ‘In-

formation driven search for point sources of gamma radiation', *Signal Processing* **90**(4), 1225–1239.

Ristic, B., Skvortsov, A. & Gunatilaka, A. (2016), 'A study of cognitive strategies for an autonomous search', *Information Fusion* **28**, 1–9.

Simpkins, A., De Callafon, R. & Todorov, E. (2008), 'Optimal trade-off between exploration and exploitation', in '2008 American Control Conference', IEEE, pp. 33–38.

Singh, S. K., Sharan, M. & Issartel, J.-P. (2015), 'Inverse modelling methods for identifying unknown releases in emergency scenarios: an overview', *International Journal of Environment and Pollution* **57**(1-2), 68 – 91.

Vergassola, M., Villiermaux, E. & Shraiman, B. I. (2007), 'infotaxis' as a strategy for searching without gradients', *Nature* **445**(7126), 406–409.

Wittenmark, B. (2008), Adaptive dual control, in H. Unbehauen, ed., 'Control systems, robotics and automation', Vol. X, EOLSS Publisher, pp. 122 – 132.

Yee, E. (2017), 'Automated computational inference engine for bayesian source reconstruction: Application to some detections/non-detections made in the ctb international monitoring system', *Applied Mathematical Sciences* **11**(32), 1581–1618.

a Reader in Unmanned Vehicles. He has considerable research experience in autonomous vehicles and Bayesian estimation and their applications in different domains.

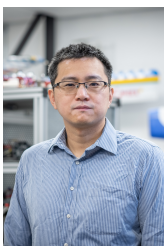


We-Hua Chen holds a Chair in Autonomous Vehicles with the Department of Aeronautical and Automotive Engineering, Loughborough University, U.K. He is interested in control, signal processing and artificial intelligence and their applications in robots, aerospace, and automotive systems. He is a Chartered Engineer, a Fellow of IEEE, the Institution of Mechanical Engineers and the

Institution of Engineering and Technology, U.K. He has authored or coauthored near 300 papers and 2 books.



Callum Rhdoes received the MEng (Hons) degree in automotive engineering from Loughborough University in 2018. He is currently pursuing his PhD within the Loughborough University Centre for Autonomous Systems research team, with a research interest in path planning techniques for autonomous robots operating in hazardous environments.



Cunjia Liu received a Ph.D. in autonomous vehicle control from Loughborough University in 2011. He was a research associate with the Department of Aeronautical and Automotive Engineering at Loughborough University, where he is currently