arXiv:2304.01052v1 [cs.RO] 3 Apr 2023

无人机系统风险感知型马尔可夫决策过程与部分可观测马尔可夫决策过程应急管理自主性的研究

PrashinSharma *密歇根大学，*2505 Hayward St.*，安娜堡，*MI 48109*

BenjaminKraske †科罗拉多大学博尔德分校，*3775 Discovery Dr*，博尔德，*CO 80303*

JosephKim ‡密歇根大学，*2505 Hayward St.*，安娜堡，*MI 48109*

ZakariyaLaouar §，Zachary Sunberg ¶科罗拉多大学博尔德分校，*3775 Discovery Dr*，博尔德，*CO 80303*

Ella Atkins ‖弗吉尼亚理工大学，*Randolph Hall，460 Old Turner St.*，布莱克斯堡，*VA 24061*

无人机系统（UAS）正日益广泛地应用于各种领域。必须将无人机系统对人员和财产造成的风险控制在可接受水平。本文提出一种风险感知应急管理自主系统，以防止部件故障时发生事故，具体而言，针对推进单元失效和/或电池退化。该自主系统建模为马尔可夫决策过程（MDP），其求解结果是一种应急管理策略，可适时执行紧急着陆、飞行终止或继续执行计划中的飞行动作。鉴于故障指示器可能出错，本文研究了 MDP 状态空间的部分可观测性，并在高保真模拟器中分析了最优策略在不同可观测条件下的表现。结果表明，鉴于状态转移模型近乎确定，在不同状态可观测性标准下，部分可观测马尔可夫决策过程（POMDP）策略与最大后验马尔可夫决策过程策略的表现相近。

I. 引言

多旋翼无人机系统（UAS）正日益广泛地应用于巡检、监视和包裹投递等任务。低空城市无人机系统运行可能使飞越区域的人群面临不可忽视的风险，原因包括

*人工智能研究员，波音研究与技术部。本文所述工作完成于作者在密歇根大学安娜堡分校攻读博士期间，美国航空航天学会会员

†博士生，科罗拉多大学博尔德分校航空航天工程科学专业，美国航空航天学会学生会员

‡博士候选人，密歇根大学机器人系，安娜堡，美国航空航天学会学生会员 §博士生，科罗拉多大学博尔德分校航空航天工程科学专业

¶助理教授，科罗拉多大学博尔德分校航空航天工程科学专业，美国航空航天学会会员

‖Fred D. Durham 教授兼系主任，弗吉尼亚理工大学 Kevin T. Crofton 航空航天与海洋工程系，美国航空航天学会会士

执行器和电池性能的不确定性、外部干扰、链路丢失的可能性，以及迄今为止无人机系统领域在系统冗余或韧性方面缺乏行业标准。对1500家无人机系统或“无人机”公司开展的一项调查 [1] 显示，无人机系统每飞行小时的故障率 10^{-3} 相对较高。由于冗余度有限，且为实现较低的运行成本而采用的运行裕度较低，多旋翼飞行器的电池退化和电机故障是导致无人机系统事故的常见因素。

无人机系统组件故障率高，可以通过增加冗余系统并采用适当的切换逻辑来管理这些系统。但无人机系统在推力、重量和成本方面的限制，使其难以采用商业运输航空所要求的安全关键系统三重冗余。尽管当前的无人机系统自动驾驶仪能够制定并准确执行正常飞行计划，但它们无法抵御大多数系统故障及恶劣环境条件，例如降水、强阵风或风切变。要提高无人机系统的运行安全性，就需要自动化紧急着陆规划和应急管理，通过实时飞行规划以及制导、导航与控制（GNC）决策来实现。此类应急管理自主性将使无人机系统能够避免碰撞，并在安全、无人居住的地点着陆，而不是失控和／或失去动力地下降（例如借助降落伞），最终落入其飞行路径下方的任意地点。

本文提出一种基于马尔可夫决策过程（MDP）构建的风险感知型应急管理自主性（CMA）。CMA 根据可获取的电池和电机状态预测信息感知系统风险。尽管 CMA 被建模为完全可观测系统，但现实中的系统具有部分可观测性。因此，本文还研究了以部分可观测马尔可夫决策过程（POMDP）表示的 CMA。POMDP 为表示组件状态不确定性提供了直接方法，但这类模型的求解也需要更大的计算开销。因此，确定所考虑的事件是否需要用 POMDP 表示，还是采用 MDP 就已足够，十分重要。我们的研究探讨了在这一 CMA 实现中部分可观测性的价值。

本文的贡献如下：

- 针对小型无人机系统的应急管理自主性（CMA），设计并评估考虑风险的马尔可夫决策过程和部分可观测马尔可夫决策过程形式化方案。
- 基于经实验验证的无人机系统组件模型，开发以城市应急管理自主性（CMA）为核心的仿真框架。
- 比较马尔可夫决策过程与部分可观测马尔可夫决策过程应急管理自主性（CMA）模型及其性能。

本文结构如下：第二节 A 回顾了无人机系统应急管理方法和故障预测技术相关文献。第三节的问题陈述界定了无人机系统应急管理自主性的范围。第四节说明了马尔可夫决策过程和部分可观测马尔可夫决策过程应急管理自主性模型。第五节详细介绍了用于评估马尔可夫决策过程和部分可观测马尔可夫决策过程应急管理自主性策略的仿真架构。第六节和第七节分别介绍了最优策略的实现结果及讨论。第八节给出结论。

II. 文献综述

A. 应急管理自主性

实现安全的应急管理自主性（CMA）的关键，是确立适合具体情境的优化与评估指标 [2]，确保无人机系统始终处于可能已退化的安全运行包线内，飞行时与地形、建筑物及其他航空器保持足够间隔，并将飞越对人员与财产造成的风险降至最低 [3]。故障预测方法提供了一系列用于预测组件故障的工具。然而，除非将故障预测信息用于主动的预防性维护或应急管理系统，否则这些信息本身并无助益。参考文献 [4] 提出了基于预测的决策（PDM）架构，该架构由四个主要部分组成：诊断器（DX）、决策器（DM）、车辆仿真（VS）和车辆本身。该故障预测问题被建模为约束满足问题（CSP），并通过回溯搜索和粒子滤波加以求解。在此框架中，任务航路点预先设定；即使存在故障，也假定这些航路点均可到达。文献 [5] 提出了类似的预测架构，并将其应用于一辆无人地面车辆。文献 [6] 中，作者提出了一种多旋翼飞行器任务规划策略，将电池荷电状态（SoC）和健康状态（SoH）纳入考量，以生成更新后的任务计划。该规划问题被表述为优化问题，目标是在满足由无人机系统、无刷电机和电池动力学定义的非线性约束的前提下，尽量降低多旋翼飞行器的总能耗。文献 [7] 及后续研究 [8] 中，作者介绍了一种融合故障预测信息的自动应急规划器。通过在线优化，为系统及其组件确定成本最低的重构方案。文献 [9] 采用滚动时域规划器，将故障预测信息所确定的约束纳入规划。

[1] 进行的一项商用无人机系统安全性与可靠性调查显示，电池系统是最可能发生故障的关键子系统中排名第三的子系统。电池系统的应急管理始于预测与健康管理（PHM），多位研究人员针对根据电池荷电状态（SoC）计算的电池剩余使用寿命（RUL）进行了精确估算研究。此类方法采用了扩展卡尔曼滤波器（EKF） [10],无迹卡尔曼滤波器（UKF） [11], 无迹变换 [12],粒子滤波器 [13],神经网络 [14],以及高斯过程回归（GPR） [15, 16] 等方法。Sharma 等人 [17] 提出了一种基于预测信息的马尔可夫决策过程（MDP），用于无人机系统的多电池重构。该方法为配备两个电池组的无人机系统设计了最优电池切换MDP策略，以便根据观测到的电池组退化情况和剩余飞行时间做出适当响应。案例研究考察了多种无人机系统任务场景下最优电池重构的性能。

另一种关键子系统故障是电机和螺旋桨损坏/故障，这可能直接导致多旋翼飞行器不稳定。Brown 等人研究了此类损坏的特征描述与故障预测 [18]，方法是利用可测故障模式。通过控制信号、角速度等可测参数中的异常，可以间接检测损坏/故障。例如，监测自动驾驶仪的电机控制信号数据，可间接观察到特定电机的损坏，因为其余旋翼会补偿该故障。这类自监测故障模式

促进了无人机系统中结构健康监测（SHM）和损伤预测（DP）的应用。一旦检测到旋翼故障，容错控制（FTC）便可重新分配控制器，以维持安全飞行。Zhang 等人 [19]详细概述了无人旋翼飞行器系统中现有的故障检测与诊断（FDD）和容错控制（FTC）。特别是，文献 [20] 结合/复合模型参考自适应控制器（CMRAC），使发生单个推进器故障的四旋翼飞行器能够安全地维持飞行并着陆。文献 [21], 也研究了四旋翼飞行器两个或更多推进器失效时的容错控制，尽管由此导致的姿态运动学控制权限降低只能提供周期解，无法支持飞行器在特定着陆点着地。

Kim 等人 [22] 为先进空中交通（AAM）开发了一套系统化的自动缓解策略，以便在多旋翼飞行器发生旋翼故障时生成安全的应急动作。该论文提出了保障应急着陆管理（ACLM）子组件的逻辑流程，并推导了退化可控性和着陆点可达性的数学表达式。案例研究展示了如何使用离线和在线飞行规划器，根据基于风险的成本指标选择应急着陆点；这项研究扩展了此前针对电池能量衰减情况下多旋翼飞行器在线紧急着陆规划的工作 [23]。本文的不同之处在于，我们使用马尔可夫决策过程（MDP）和部分可观测马尔可夫决策过程（POMDP）建模，并通过概率模型检验（PMC）计算无人机系统的应急管理策略。

B. 可观测与部分可观测马尔可夫决策过程模型

正如 Sharma 等人 [17], 所展示的，MDP 框架支持对偶然不确定性进行推理，从而为安全关键应急管理决策提供依据。然而，当系统状态无法被完全观测时，则需要对认知不确定性进行推理。涉及这类不确定性的问题可以建模为 POMDP。Sunberg 等人指出，推断问题中的隐藏状态，有助于获得优于预定义基准的策略 [24]。与 MDP 相比，POMDP 的一项关键优势在于，借助观测所提供的信息对状态进行概率表示，从而增强了系统的鲁棒性 [25]。在这两种范式下，最优动作都会最大化预先定义的奖励函数；但只有在 POMDP 中，动作才能同时实现一项目标：收集信息以改进信念 [26]。当前称为 ACAS-X 的国家空域防撞系统采用 POMDP 框架，为航空飞行员提供稳健的监视和咨询逻辑 [27]。

自主系统中的故障检测问题可能很难，通常需要根据可能不完整且带有噪声的测量值推断故障。当将 FDD/FTC 问题建模为马尔可夫决策过程（MDP）时，无法恰当地描述部分可观测的故障状态。部分可观测马尔可夫决策过程（POMDP）框架可以通过故障状态上的概率分布，维持对故障的信念表示。 [28–30] 中的作者采用了一种框架，用于维持对故障状态或模式的信念，从而在发生故障后实现更稳健的恢复。

尽管部分可观测马尔可夫决策过程比马尔可夫决策过程更具表达能力和鲁棒性，但一般有限时域部分可观测马尔可夫决策过程的精确求解在最坏情况下属于 PSPACE 完全问题，因此需要采用近似方法 [31]。设计者必须权衡

满意的解与求解速度。 [32] 中的作者提出了 QMDP，它假设在第一个时间步之后完全可观测，从而为无限时域部分可观测马尔可夫决策过程求得近似最优策略。这种方法可能会因天真地假定完全了解状态过程，而导致对信息收集动作的信息价值近似不佳。许多方法都试图克服部分可观测马尔可夫决策过程的维数灾难，同时保持接近最优。SARSOP [33] 便是其中一种方法：Kurniawati 等人利用最优可达信念空间 的概念来提高计算效率。

据我们所知，尚未有人在高保真模拟器中全面研究并比较马尔可夫决策过程与部分可观测马尔可夫决策过程在应急管理场景下的有效性。本研究旨在弥补这一空白，并进一步探讨何时应将应急管理问题建模为马尔可夫决策过程，何时应建模为部分可观测马尔可夫决策过程。

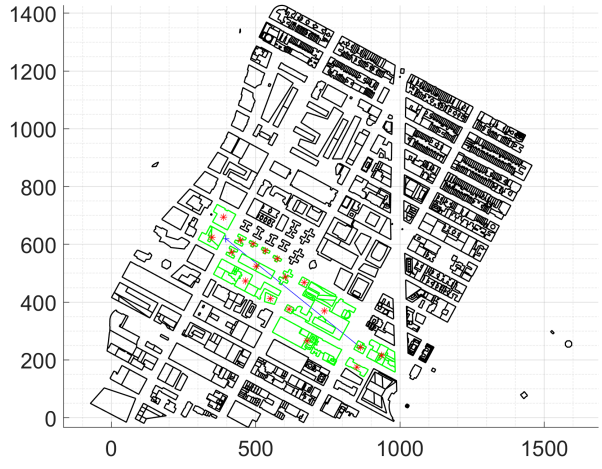
III. 问题陈述

本研究旨在为无人机系统开发应急管理自主性（CMA），利用系统状态和组件预后信息选择适当动作，以保障无人机系统的安全并防止高风险系统故障。为评估 CMA，我们选取一项具体任务进行深入研究，但本文提出的 CMA 方法也可应用于其他任务。本文研究的任务是：多旋翼无人机系统执行包裹投递任务时，发生飞行中推进装置故障、电池退化或电池电压骤降事件。我们假设已知飞行计划和紧急着陆点坐标。研究考虑了多种飞行计划，其可用电池能量的安全裕度各不相同。未知信息包括推进单元和/或电池发生故障的时间，以及阵风强度。

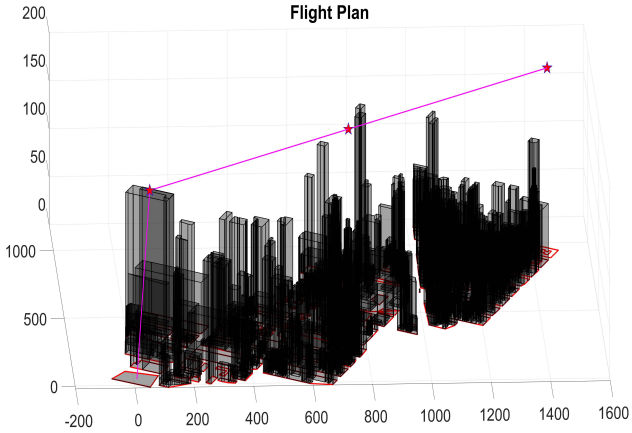
我们使用来自 [34] 的后处理 OpenStreetMap (OSM) 建筑数据，构建了一个用于评估应急规划模型的仿真环境。这些地图数据根据建筑高度、类型和轮廓坐标，构建了曼哈顿南部的三维建筑结构。我们使用 [34] 中经过修改的地理围栏多边形提取算法，寻找紧急着陆点。确定无人机系统飞行的起点和目的地后，该算法会生成一个感兴趣区域，并通过计算几何方法寻找潜在着陆点。我们使用一架在飞行中突发故障的六旋翼飞行器，在这张紧急着陆点地图上对 CMA 进行仿真。图 1a 展示了曼哈顿南部包裹投递任务的俯视示例，绿色线条和红色星形标出了离线着陆点。图 1b 展示了仿真中为包裹投递任务定义的三维飞行前应急数据库和一条正常飞行轨迹。

IV. 应急规划模型

CMA形式化的目的是，在考虑多电池组和电机中观测到的任何性能退化的同时，为无人机系统生成风险感知安全保障动作。在本研究中，CMA被建模为平稳无限时域马尔可夫决策过程 [35]。基于该马尔可夫决策过程模型，构建了一个具有不同



(a) 曼哈顿俯视图，名义飞行路径以蓝线表示，备选着陆点以红色星形标记表示



(b) 具有名义飞行路径的三维曼哈顿环境以洋红色线条表示，预先规划的检查点以红色星形标记表示。

图 1 例　用于马尔可夫决策过程和部分可观测马尔可夫决策过程 CMA 仿真的小型无人机包裹配送任务 ns.

状态可观测性 con　条件的POMDP模型。下一节将详细介绍CMA MDP和部分可观测马尔可夫决策过程形式化 s.

A. 马尔可夫决策过程模型

马尔可夫决策过程由四元组 $\langle S, A, p(s'|s,\ a),\ r(s,\ a)\rangle$ 定义，其中 S 是有限的系统状态集合，A 是允许的动作，$p(s'|s,\ a)$ 是状态转移概率张量，$r(s,\ a)$ 是在状态 s 下执行动作 a 所获得的奖励。马尔可夫决策过程根据贝尔曼方程，计算各状态下可使期望值最大化的动作。可使用值迭代或策略迭代等经典算法来确定最优马尔可夫决策过程策略 $\pi^{\star}$。

1. 状态空间描述

CMA MDP 旨在根据需要执行紧急着陆选项，以尽量降低无人机系统的风险。MDP状态空间包含与电机和电池健康状态、可用电池能量以及电机剩余使用寿命相关的信息，这些信息共同提供关键依据，以便在可能的情况下实现安全着陆。CMA MDP 状态定义为 $S = S_E \cup S_S$, $S_E = \{C, T, FL, E\}$ 和 $S_S = \{S = (FS, MH, MM, BH, RM) | FS \in \{N, ELASAP, ELPract\}\}$,$MH \in \{NF, SF, JF\}$, $MM \in \{MM0, MM1\}$, $BH \in \{G, M, P\}$, $RM \in \{RM0, RM1\}\}$。

MDP状态向量的各个要素定义如下：

- 飞行状态（FS）：由于马尔可夫假设，当前状态必须包含相关的系统状态信息。FS状态特征有三个可能值：N：正常飞行，$ELASAP$：尽快执行紧急着陆（尽可能快），$ELPract$：条件允许时执行紧急着陆。如果 $FS = ELPract$动作空间缩减为 $A_s = \{NoOp,\ Terminate,\ LandASAP\}$。如果 $FS = ELASAP$动作空间缩减为$A_s = \{NoOp,\ Terminate\}$。在 $FS = N$下，所有动作均可用。基于 FS的动作可用性已在马尔可夫决策过程中手动编码。

• 电机健康状态（MH）该特征假设具有以下取值： $MH = \{NF:\ NoFault,\ SF:\ SpallingFault,\ JF:\ JamFault\}$。电机性能退化有多种类型，但本研究仅关注剥落退化，这是电机轴承的一种关键故障状况。之所以选择这种故障，是因为已有大量关于剥落故障预测的文献 [36],[37]。假设剥落故障的发生完全可观测。当剥落区域超过阈值面积时，电机便发生故障。一旦电机停止运行，交互多模型（IMM）卡尔曼滤波器便会将其检测为卡滞故障，具体如第五节所述。

• 电机裕度（MM）对单个电机的数值定义如下：

$$MM = 1 - \sum_{i=1}^{n_f} w_i \frac{t_{FlightTime}}{t_{RUL_i}} \tag{1}$$

其中， $t_{FlightTime}$ = 飞行时间（秒）， t_{RUL} = 电机剩余使用寿命（RUL）（秒）， n_f = 不同类型的电机故障， $\sum_{i=1}^{n_f} w_i = 1$。本文仅考虑单一剥落故障，因此 $n_f = 1$。检测到剥落故障时， t_{RUL} 使用 Paris定律 [37]计算。在电机正常健康状态下，$t_{RUL} \geq 10 \times t_{FlightTime}$，因为在正常条件下，推进系统会按照较高的安全裕度进行设计。MDP 离散状态特征 MM 被赋予两个逻辑值： $MM0:\ MM < 0$ 和$MM1:\ 0 \leq MM < 1$。 $MM < 0$表示电机已停止运转或 $t_{RUL} < t_{FlightTime}$，即电机无法在剩余飞行时长内持续提供推力。范围 $0 \leq MM < 1$ 表示电机处于正常健康状态或正在发生剥落故障，但预计仍能在剩余飞行时长内提供推力。若考虑多种电机故障，则需要进一步研究如何离散化MM。本文仅考虑单个电机故障。

• 电池健康状况（BH）被抽象为三种可能的健康状态： $Good$、 $Medium$ 和 $Poor$。正常情况下 $BH = Good$。如果电池发生功率衰减或容量衰减，但两者并未同时发生 $BH = Medium$。如果电池组同时发生功率衰减和容量衰减 $BH = Poor$。由于电池退化是一个相对缓慢的过程，为简化起见，我们假设在本研究中， BH 在单次飞行期间保持不变。

• 可达裕度（RM）定义如下：

$$RM = 1 - \frac{t_{FlightTime}}{t_{EOD}} \tag{2}$$

其中， t_{EOD} = 为两块电池均在使用时串并联电池组的放电终止时间，其计算方法如 [17] 所述。 $t_{FlightTime}$ 是无人机系统完成当前飞行计划所需的时间。RM 离散化为 $RM0:\ RM < 0$ 和 $RM1:\ 0 \leq RM < 1$。 $RM0$ 表示电池组的能量不足以完成当前飞行计划。 $RM1$ 表示电池组的能量足以

完成飞行计划。

• 完成*(C)* 假定任务进行期间取值为 $False(0)$，无人机系统完成正常或紧急飞行计划后取值为 $True(1)$。所有 $s \in S$ 在 $C = 1$ 时均转移至 E（结束）这一吸收状态。

• 已终止 *(T)* 在执行飞行终止操作之前取值为 $False(0)$，执行该操作时锁存为$True(1)$。所有 $s \in S$ 在 $T = 1$ 时均转移至 E （结束）这一吸收状态。

• 故障 *(FL)* 取值为 $False(0)$，除非将其值设为 $True(1)$，这表示任务因控制器不稳定（通过轨迹跟踪误差测量）或电池能量不足（$RM < 0$）而失败。本研究不考虑其他系统故障，但在任何自主系统认证流程中，都有必要使用 CMA 对其进行建模和管理。所有 $s \in S$ 在 $FL = 1$ 时均转移至 E （结束）这一吸收状态。

该状态空间经过有意抽象，将应急管理的重要决策区域整合为复杂度最低的表示形式。该状态空间抽象得到 $|S_E| = 4$、 $|S_S| = |FS| \times |MH| \times |MM| \times |BH| \times |RM| = 3 \times 3 \times 2 \times 3 \times 2 = 108$，总计包含 $|S| = 112$ 个状态，其中包括 C， T， FL和 E。

2. 动作空间描述

CMA动作空间旨在应对本文所考虑的组件退化场景，同时最大限度地降低无人机系统带来的风险。CMA动作集为 $A = \{NoOp、 Terminate、 LandASAP、 LandPract\}$，定义如下：

• 不执行操作（*NoOp*）：执行此动作时，无人机系统继续执行当前飞行计划。

• 终止：此动作表示终止飞行，以降低失效无人机系统在着地时对人员与财产造成的风险。终止动作可以通过展开降落伞并切断电机功率来实现。此动作会立即使无人机系统无法运行。在执行终止动作时，无人机系统会失控着陆，但其造成的风险低于失控事件可能造成的风险，因为降落伞会降低动能，并且不会向螺旋桨施加扭矩。

• *LandASAP*： 执行此动作时，无人机系统会首先确定其大致可达范围内可用的紧急着陆点列表；该大致可达范围首次用于 [38] 中的航空器紧急着陆规划。大致可达范围是通过在横向和纵向定义一条最小控制量轨迹，计算无人机系统可飞行的最大距离得出的，如图2所示。如图所示，多旋翼飞行器的可达范围是初始速度和电池放电终止（EOD）时间的函数。图2中的锥体显示，在初始速度条件给定时，EOD值越高，大致可达范围越大。图2右半部分展示了锥体在不同EOD时间处的横截面。如图所示，EOD值较低且初始速度较高时，可达范围横截面会呈不对称，这是因为多旋翼飞行器需要时间减速并反向飞行。随后， $LandASAP$ 规划到达已识别的可达紧急着陆点的最短时间轨迹，并选择可达裕度最大的方案 RM。这一步在计算上可能

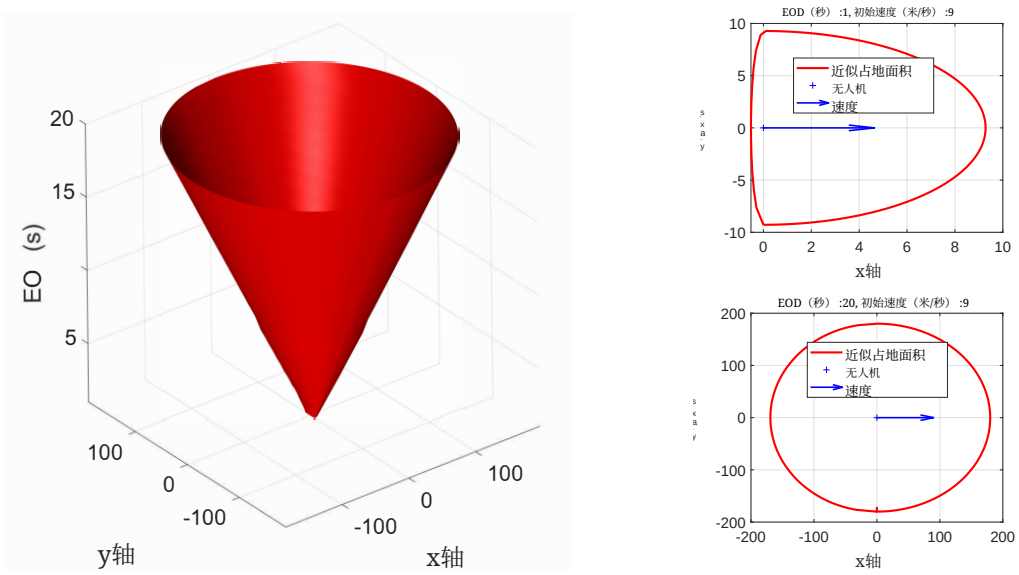


图 2（左）不同电池 EOD值以及沿 $+x$轴的初始速度为 $9m/s$ 时的多旋翼飞行器占用范围。（右上）EOD值为 $1sec$ 时的占用范围。（右下）EOD值为 $20sec$ 时的占用范围。

代价高昂，因此在可能的情况下优先选择下文定义的 *LandPract* 动作。

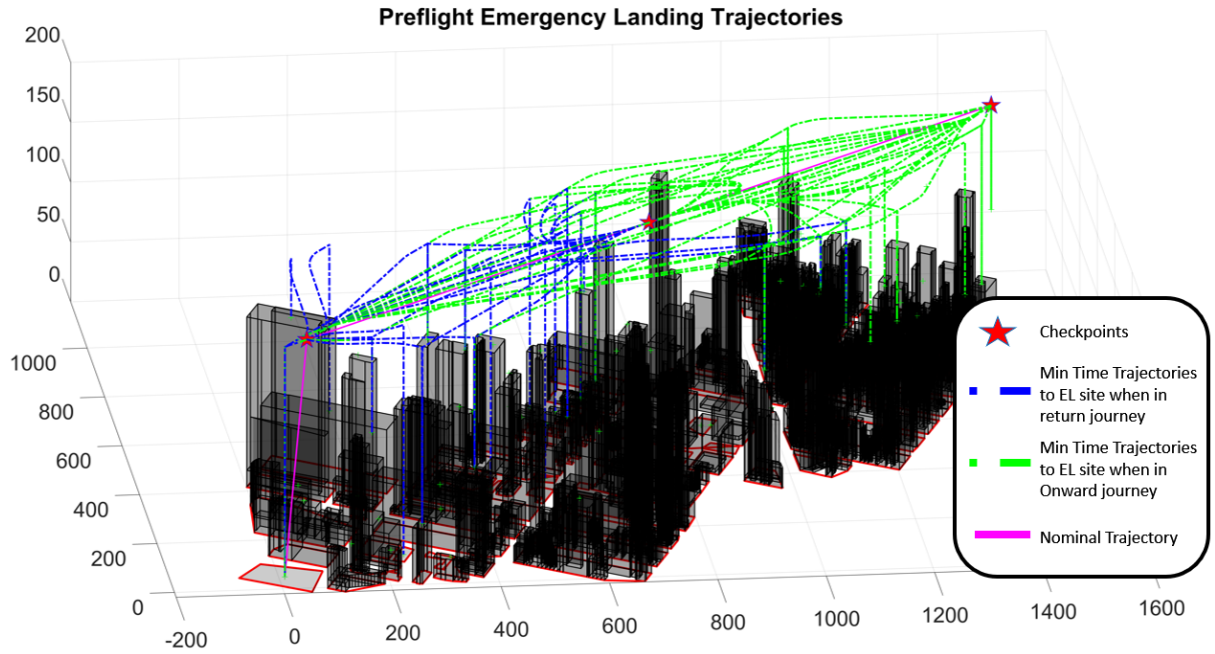


图 3 飞行前计算并存储的名义计划中各检查点对应的紧急着陆（EL）计划。

• *LandPract*： 使用此动作时，无人机系统会利用飞行前根据 [22] 计算并存储的着陆轨迹，前往从一组名义飞行计划检查点出发可到达的紧急着陆点；这些检查点如图3中的 ★ 所示。检查点的选取旨在将飞行计划划分为长度相等的路段，这也依据 [22]。基于

根据无人机系统当前位置及其与正常飞行计划中检查点的接近程度，从预先规划的紧急着陆轨迹列表中，选择具有最大 RM 的紧急着陆地点。图3展示了一个预先规划的紧急着陆轨迹示例。由于这些轨迹是从离线生成的数据库中选取的，因此与*LandASAP* 相比，此动作的计算开销更低。

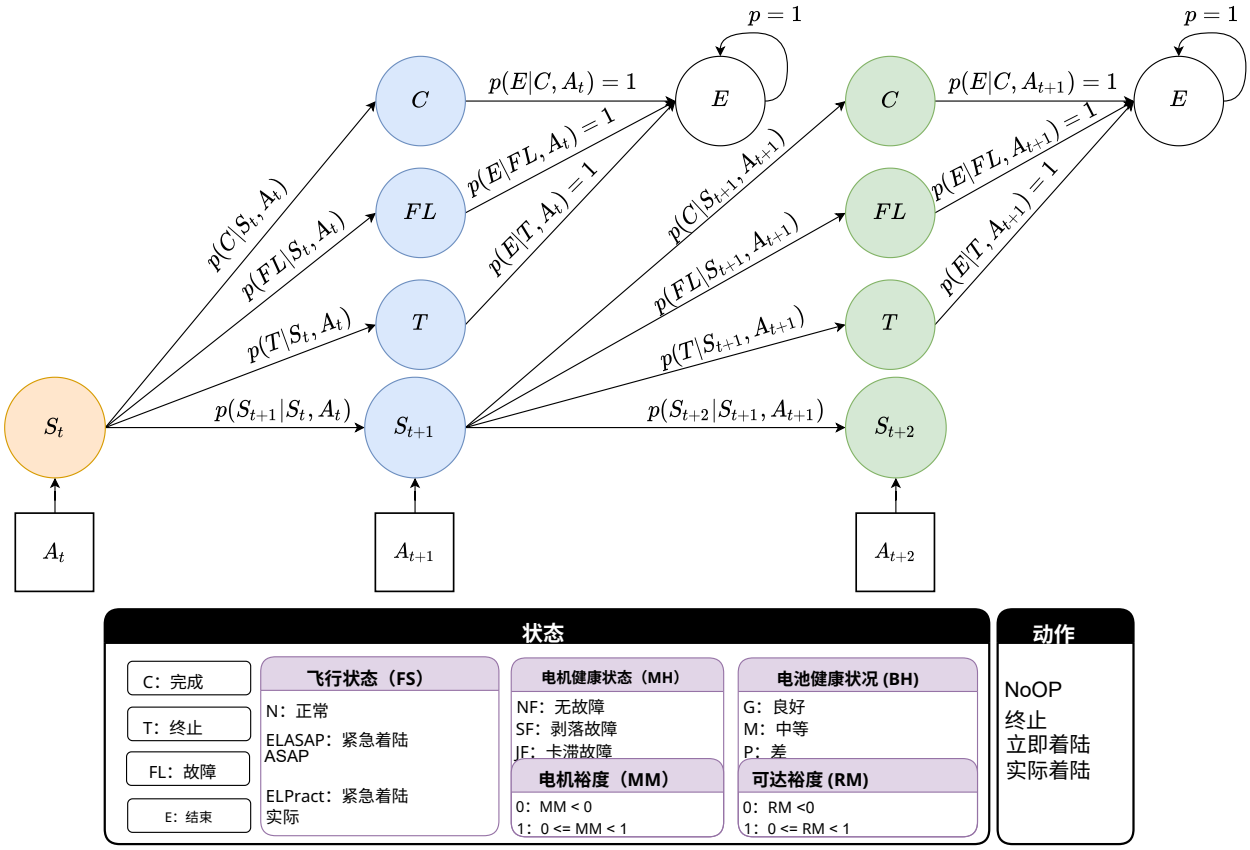
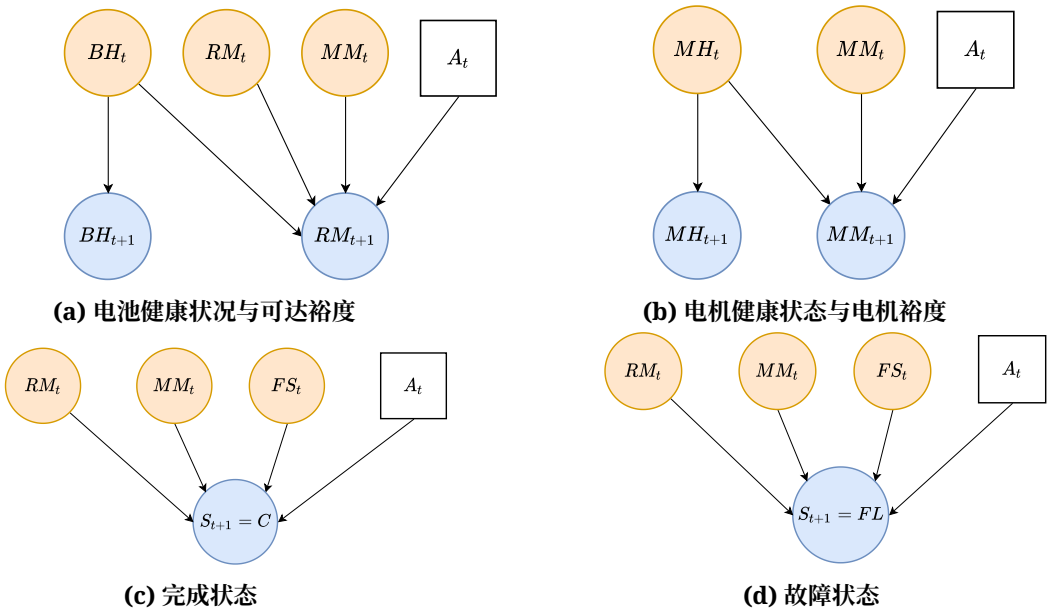


图4 CMA MDP 的状态转移图。 C：*Complete*，FL：*Failure*，T：*Terminate* **表示终止/吸收状态，** A_t：*Action*

3. 状态转移建模

CMA MDP 的状态转移建模为动态决策网络（DDN），如图 4 所示。为简化表示，每个暂态或吸收状态均以真实的单一特征标记（C 表示任务完成，FL 表示故障，T表示飞行终止，E表示结束）。时间 t 的任何其他状态均由特征集 $S_t=\{FS、MH、MM、BH、RM\}_t$ 定义。与手动定义表中每个元素相比，这种定义状态转移概率的方法能够利用条件独立性并支持可解释性。图 5 描述了时间 t 和 $t+1$ 时 MDP 状态特征之间的相互依赖关系。考虑图 5b 所示的 $t+1$ 时电机裕度（MM_{t+1}）情况。MM_{t+1} 根据 t_{RUL} 和 $t_{FlightTime}$ 的值，按照公式 1 定义。t_{RUL} 直接受到电机健康状态 MH_t 和 $t_{FlightTime}$ 的影响。因此，MM_{t+1} 取决于动作（A_t）、电机裕度（MM_t）和电机



(a) 电池健康状况与可达裕度 **(b) 电机健康状态与电机裕度**

(c) 完成状态 **(d) 故障状态**

图 5 展示马尔可夫决策过程状态元素随时间变化依赖关系的动态决策网络。

健康状态（MH_t）以及前一个时间步 t。表 1 和表 2 给出了状态转移概率的示例，其中假设状态更新频率为 $1Hz$。

表 1 MM_{t+1} 的状态转移概率

MM_t	MH_t	MM_{t+1}	$P(MM_{t+1}\|MM_t, MH_t, A_t = NoOp)$
$MM0$	NF	$MM0$	0
$MM0$	NF	$MM1$	1
$MM0$	SF	$MM0$	0.995
$MM0$	SF	$MM1$	0.005
$MM0$	JF	$MM0$	1
$MM0$	JF	$MM1$	0
$MM1$	NF	$MM0$	0
$MM1$	NF	$MM1$	1
$MM1$	SF	$MM0$	0.002809
$MM1$	SF	$MM1$	0.997191
$MM1$	JF	$MM0$	1
$MM1$	JF	$MM1$	0

表 2 MH_{t+1} **的状态转移概率**

MH_t	MH_{t+1}	$P(MH_{t+1}\|MH_t)$
NF	NF	0.9999525
NF	SF	0.0000475
NF	JF	0
SF	NF	0
SF	SF	0.997191
SF	JF	0.002809
JF	NF	0
JF	SF	0
JF	JF	1

4. 奖励函数

下面的奖励函数旨在鼓励采取能够防止无人机系统在性能退化条件下发生故障，或在可能时继续执行飞行计划的动作。可调权重 w_E、 w_S 和 w_A 根据用户偏好选定。

$$R(S,A) = R(S) + R(A) \tag{3}$$

$$R(S) = \begin{cases} w_E(S) f_E(S), & S \in S_E \\ w_S^T f_S(S), & S \in S_S \end{cases}, \; R(A) = w_A(A) f_A(A) \tag{4}$$

$$f_E(S) = \begin{cases} 1, & S = C \\ -0.1, & S = T \\ -1, & S = FL \\ 0, & S = E \end{cases}, \; f_A(A) = \begin{cases} 1, & A = NoOp \\ -1, & A = Terminate \\ -0.5, & A = LandASAP \\ 0.5, & A = LandPract \end{cases} \tag{5}$$

$$f_S(S) = [f_{S_{FS}}(FS), f_{S_{MH}}(MH), f_{S_{MM}}(MM), f_{S_{BH}}(BH), f_{S_{RM}}(RM)]' \tag{6}$$

$$f_{S_{FS}}(FS) = \begin{cases} 1, & FS = Nominal \\ -1, & FS = \{ELPract, ELASAP\} \end{cases}, f_{S_{MH}} = \begin{cases} 1, & MH = NF \\ 0, & MH = SF \\ -1, & MH = JF \end{cases} \tag{7}$$

$$f_{S_{MM}}(MM)=\begin{cases}1, & MM=MM1\\ -1, & MM=MM0\end{cases},\ f_{S_{BH}}=\begin{cases}1, & BH=G\\ 0, & BH=M\\ -1, & BH=P\end{cases} \tag{8}$$

$$f_{S_{RM}}(RM)=\begin{cases}1, & RM=RM1\\ -1, & RM=RM0\end{cases} \tag{9}$$

为马尔可夫决策过程状态设定的奖励值旨在优先确保标称任务不中断地完成。然而，如果出现电池性能下降或电机故障等非标称情景，则会对这类状态施加惩罚，以鼓励在附近安全着陆。同样，动作权重值的设定旨在不中断任务，并优先选择紧急着陆动作，而非飞行终止。为*LandASAP*和 *LandPract*动作赋予不同权重值，是因为与 *LandPract*相比，执行 *LandASAP*的计算成本更高。

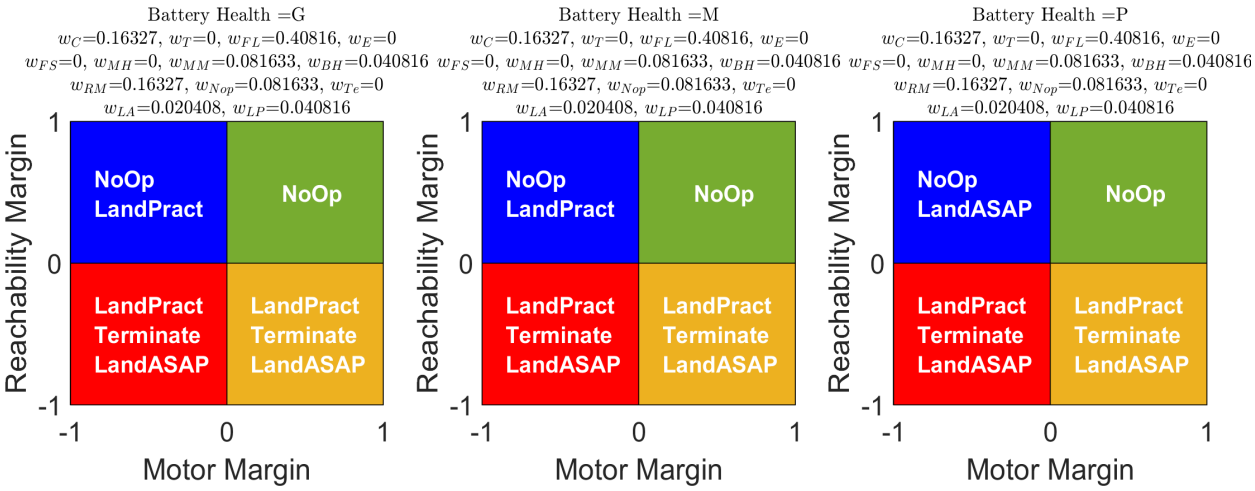


图6 奖励敏感性分析，象限颜色表示 ■：安全区域，■：危险区域，■：临界区域，■：超临界区域。

权重 w_E、w_S和 w_A经过调整，使马尔可夫决策过程的最优解（第IV.A.5节）在特定关键状态下符合我们期望的策略，如图6所示。各象限根据 RM 和 MM 的值定义。当电池健康状况较差时，即任务失败的可能性较高时，在危险区域内可用的紧急着陆为 *LandASAP*；相比之下，在电池健康状况良好或中等的情况下，可用的紧急着陆

动作是 $LandPract$。最终选定的权重为：

$$w_E(S)=\begin{cases}0.163, & S=C\\ 0.408, & S=FL\\ 0, & S=\{T,E\}\end{cases},w_A(A)=\begin{cases}0.082, & A=NoOp\\ 0, & A=Terminate\\ 0.02, & A=LandASAP\\ 0.041, & A=LandPract\end{cases} \tag{10}$$

$$w_S=[0,\ 0,\ 0.082,\ 0.041,\ 0.163]' \tag{11}$$

这种权重调整方法类似于一种逆强化学习（IRL）[39]方法，通过寻找奖励函数参数来恢复专家策略。对此方法的一种解释是，马尔可夫决策过程充当了一种稳健的泛化机制：专家指出少数状态下的恰当行为，而马尔可夫决策过程则将这种行为泛化到所有状态。自动逆强化学习常见的缺陷，例如问题本身描述不充分[39],，在此情形下得以避免，因为手动调整由专家监督。

5. 最优马尔可夫决策过程求解

鉴于无限时域折扣马尔可夫决策过程可以在多项式时间内求解（即求解时间关于状态空间和动作空间[40]的规模呈多项式关系），并且我们问题的状态空间和动作空间规模也有限，迭代算法可以在可处理的时间内轻松找到最优策略。CMA MDP 的最优策略 π^* 使用决策时刻为 $1Hz$ 的值迭代离线计算得出。

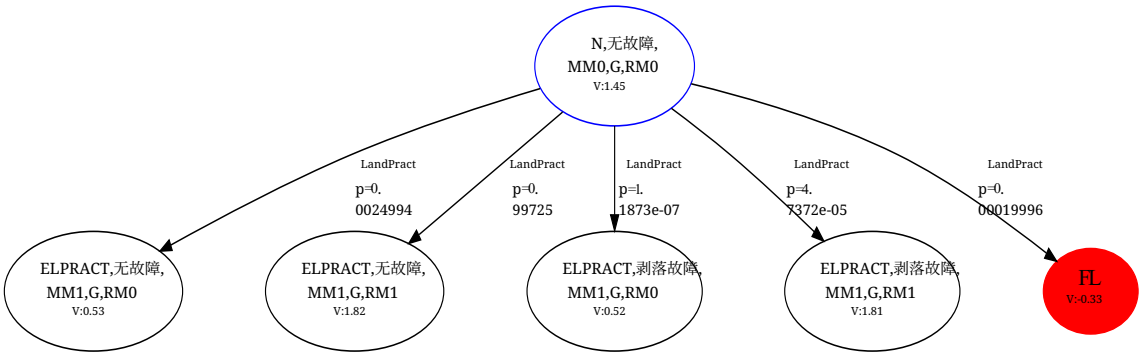


图 7 从 S_t ={N、 NF、 $MM0$、 G、 $RM0$} 执行的最优动作

例如，图 7 显示了从状态 S_t ={$N, NF, MM0, G, RM0$} 出发时的最优动作选择 $LandPract$。 $MM<0$ 和 $RM<0$ 都表明无人机系统处于危急状态。锂聚合物电池中的能量不足以完成标称飞行计划，同时，电机要么已经停止运转，要么其 RUL 太低，无法完成飞行计划。无人机系统的最佳动作是 $LandPract$，因为电池健康状况良好，而且此动作

会从数据库中选择一个时长远短于原标称飞行计划的紧急着陆飞行计划。

B. POMDP模型

许多系统被建模为完全可观测过程，但实际上，问题的某些部分是部分可观测的。本部分工作旨在通过构建一个部分可观测马尔可夫决策过程（POMDP），探讨马尔可夫决策过程的可观测性假设，并据此评估最优MDP策略和近似最优POMDP策略的性能。简而言之，本研究旨在确定部分可观测性在表征组件故障时的价值。

为了评估应急管理中完全可观测方法与部分可观测方法之间的权衡，我们构建了CMA MDP的部分可观测形式。该问题的POMDP表示提供了一个基准，用于评估MDP和POMDP求解方法。

POMDP由六元组 $\langle S, A, p(s'|s,\ a),\ r(s,\ a),\ O,\ p(o|s',\ a)\rangle$定义，其中 S、 A、 $p(s'|s,\ a)$、 $r(s,\ a)$在第IV.A节中定义， O 是系统观测的有限集合， $p(o|s',\ a)$是在给定状态转移至某状态且采取某动作的条件下，获得特定观测的概率。

1. 观测空间

部分可观测马尔可夫决策过程的观测空间 O 定义如下：

$$O = FS \times MotorObs \times RM \tag{12}$$

其中，$MotorObs = \{MM0, MM1, JF\}$

观测被定义为一个元组，由真实状态的某个子集构成；在本例中，该元组由飞行状态、可达裕度，以及根据下文说明而定的电机裕度或电机健康状态组成。

2. 观测概率

Ob测概率 $p(o|s',\ a)$ 取决于下文定义的真实底层状态：

- 飞行状态： 每个时间步都会确定性地观测到真实飞行状态。
- 可达裕度： 以 P_o 的概率观测到真实可达裕度，以 $1 - P_o$ 的概率观测到不准确的可达裕度。
- 电机健康状态： 由于交互多模型能够快速消除卡滞故障的不确定性（远快于马尔可夫决策过程的时间尺度），因此该故障可被确定性地观测到。观测到卡滞故障时，不会指示电机裕度，因为发生卡滞故障时电机裕度始终为零。除此之外，不提供电机健康状态的观测。
- 电机裕度：处于剥落故障状态时，真实电机裕度被观测到的概率为

P_o，观测到不准确电机裕度的概率为 $1-P_o$。仿真中，无故障状态下的电机裕度确定为 1。因此，处于无故障状态时，$MM1$ 被观测到的概率为 P_o，而 $MM0$ 被观测到的概率为 $1-P_o$。这些模型同时考虑了电机裕度和故障类型的不确定性，其中故障类型是不确定性的主要来源。在仿真中，只有电机处于剥落故障或卡滞故障状态时，电机裕度状态 $MM0$ 才有可能出现。因此，通过引入电机裕度的不确定性，也引入了故障类型的不确定性；而此前在无故障状态下，不可能出现错误的电机裕度指示。

• 电池健康状况：未提供关于此状态元素的观测。从部分可观测的意义上说，该状态元素仅由状态的初始信念和转移动力学确定（在此情况下，转移动力学不允许电池健康状况在不同状态之间转移）。

理想情况下，观测概率 P_o 应根据文献中的传感器精度设定。然而，由于马尔可夫决策过程的定义采用二元状态，因此需要将诊断误差转换为二元观测概率。此外，系统的可观测性程度可能因所用的具体设备而异。由于不存在唯一适用的观测概率，我们选择在多种观测概率下评估部分可观测马尔可夫决策过程策略与马尔可夫决策过程策略之间的性能差异。对观测概率进行这样的扫描，可为在此马尔可夫模型中，何种可观测性程度适合采用部分可观测马尔可夫决策过程表示提供更具普适性的指导。

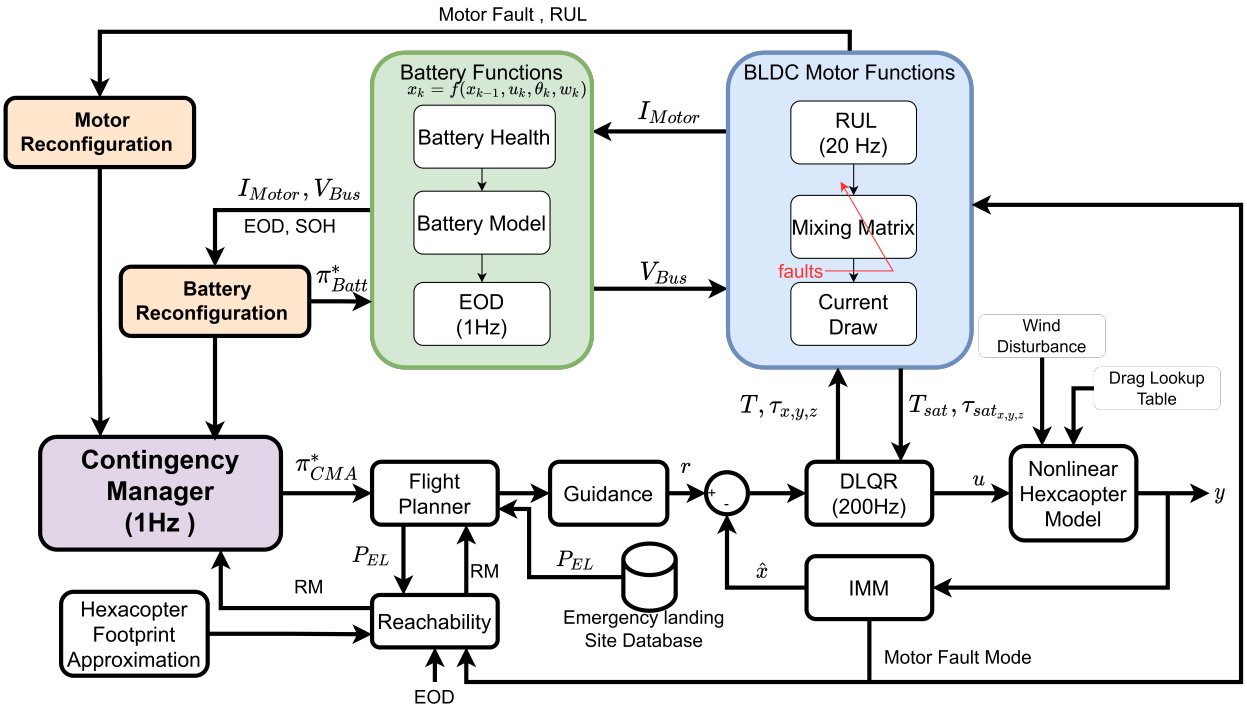


图 8 用于模拟马尔可夫决策过程和部分可观测马尔可夫决策过程策略的系统示意图。

V. 仿真

A. 系统架构

在将 CMA 部署到实际系统之前，必须在仿真环境中，利用系统组件的高保真模型对 CMA 进行严格测试。我们基于经过实验验证的六旋翼飞行器、锂聚合物电池和推进模块模型 [17], [41]，通过仿真评估 CMA MDP 和部分可观测马尔可夫决策过程的性能。图 8 展示了本文开发的高保真仿真控制回路。我们选择六旋翼无人航空系统，以便在推进系统复杂度最低的情况下实现冗余。仿真以输入无人机系统制导模块的包裹配送飞行计划启动。飞行规划器接收来自紧急着陆点数据库、离线开发的马尔可夫决策过程应急策略以及可达性模块的信息。飞行规划器根据马尔可夫决策过程动作的指示，向制导—导航—控制（GNC）六旋翼飞行器模块提供更新。

电池功能模块包含 LiPo 电池的等效电路电阻模型、电池健康状态识别模块和放电终止计算器。检测到剥落故障后，电机功能模块会确定电机剩余使用寿命（RUL）。根据剥落故障特性、电机推力饱和情况，以及满足指令推力要求时的电机电流消耗，对六旋翼飞行器的电机混控矩阵进行重构。

为诊断是否存在旋翼故障，我们开发了一种定制的交互多模型（IMM）滤波器。IMM 能够在保持计算可行性的同时提供足够的精度 [42]。有关 IMM 的深入介绍，请参阅 [43] 。所采用的 IMM 会为六旋翼飞行器运行的七种工况或模式维持一个概率分布。第一种模式对应标称动力学，其余六种模式则对应单个六旋翼旋翼发生故障时的动力学。在马尔可夫决策过程案例研究中，假设旋翼健康状态完全可观测，因此使用最大后验概率（MAP）估计来确定六旋翼飞行器所处的模式。由于当前实现的 IMM 能够以远快于部分可观测马尔可夫决策过程计算的时间尺度解析旋翼失效不确定性，因此在此情境下，卡滞故障的 IMM 诊断也被视为完全可观测。

这些系统模块以不同的速率进行仿真，以复现实际硬件。GNC模块以 $200Hz$运行。电池放电终止（EOD）计算以 $1Hz$ 更新，因为EOD不会频繁变化。应急策略执行以 $1Hz$进行仿真，与马尔可夫决策过程形式化相匹配。仿真中的故障定义为无人机系统处于不稳定状态，或可用电池电压低于阈值电压。

B. 策略与策略评估

上述部分可观测马尔可夫决策过程的求解方案通过近似最优离线部分可观测马尔可夫决策过程求解方法 SARSOP [33], 以及使用信念状态的不同度量来提供，这些度量允许在部分可观测马尔可夫决策过程中评估 MDP策略（通过值迭代获得）。这些信念度量采用两种不同的方法来处理状态中的不确定性。

之前的观测　　可观测 MDP 近似（$Obs\ MDP$）将每个时间步接收到的观测视为真　　e

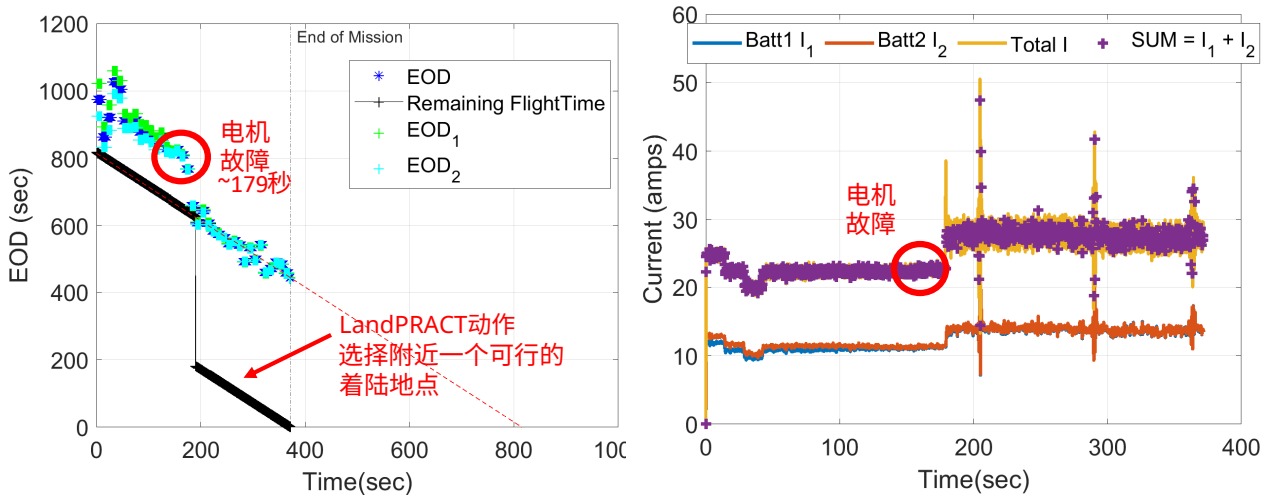
实状态，并基于这一假定状态采取最优 MDP 动作。实际上，由于观测空间是状态空间的子集，电池健康状况根据初始信念推断；当电机裕度为 1（$MM1$）时，假定电机健康状态为无故障（NF）；当电机裕度为 0（$MM0$）时，假定电机健康状态为剥落故障（SF）。针对重构状态，采取值迭代策略中的最优 MDP 动作。这种表示旨在复现所有观测均准确这一假设。第二种近似是最大似然或最大后验（MAP）表示（$MAP\ MDP$）。该表示在信念更新后，针对信念中似然最高的状态采取最优 MDP 动作。

观测概率从完美观测（以概率 $P = 1$ 观测到正确的状态子空间或状态特征）变化到近似均匀的概率，即观测到正确状态或错误状态的概率近似相等（$P = 0.6$）。针对每个观测概率运行大量仿真，以评估安全性（避免执行终止动作、成功着陆）和效率（达到初始目标）等指标，以及奖励。

对于在上述高保真仿真架构上执行的仿真，发生剥落故障的概率为 $P(SF|NF) = 0.5$。请注意，该概率高于根据表 2 为求得最优马尔可夫决策过程策略和 POMDP 策略（针对每个可观测度分别求得一个策略）所使用的概率。剥落故障会导致卡滞故障，可能需要采用修改后的轨迹（紧急着陆）或终止飞行计划。因此，与正常情况下的预期相比，结果反映了更高的故障发生频率。采用这一较高的采样频率，是为了减少评估策略性能所需的仿真次数。在完全可观测问题上运行的马尔可夫决策过程（直接提供状态）$True\ MDP$，用于提供性能上界。为便于比较，还运行了一个在每个时间步执行 $NoOp$ 的与状态无关策略。请注意，该 $Baseline\ NoOp$ 策略不依赖状态或观测，因此会在所有可观测度下显示。同样，完全可观测的马尔可夫决策过程不依赖观测，因此也会在所有可观测度下显示。

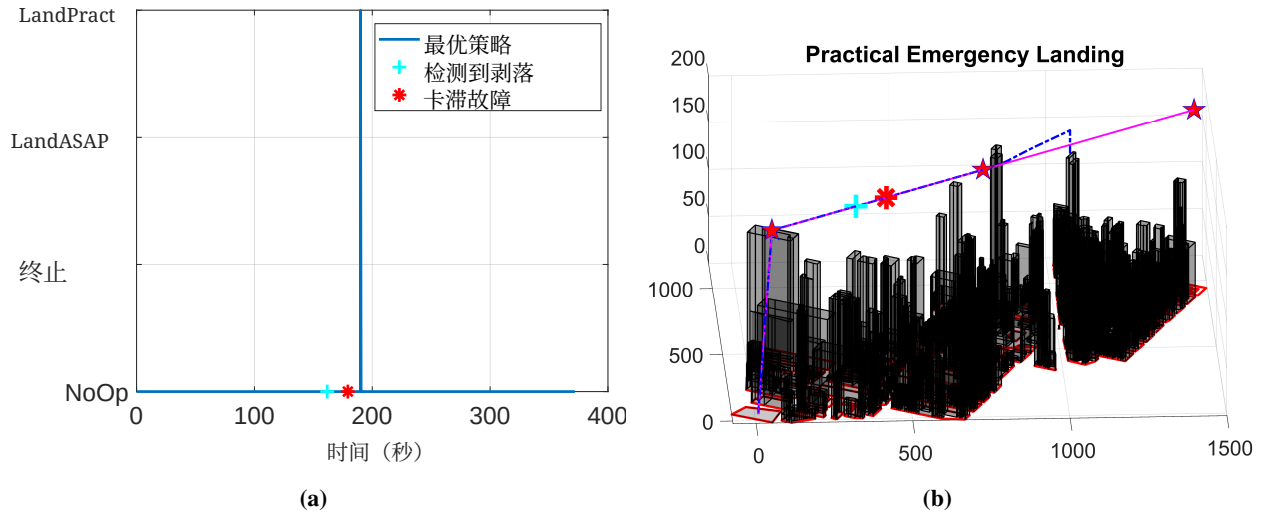
VI. 结果

通过多项案例研究，确定 CMA MDP 在无人机系统包裹配送任务中的效能。除案例研究外，我们还进行了蒙特卡洛（MC）仿真，以评估 MDP、POMDP 策略和 $Baseline\ NoOp$ 策略的整体性能。在 MC 仿真期间，对风速、风向、电池健康状况、电机故障位置和状态观测噪声等参数进行了变化。我们采用两项指标，即原始任务完成率和安全率，来评估性能。这些指标分别反映策略效率和策略安全性。原始飞行计划完成率表示无人机系统任务开始时由飞行规划器提供的已完成轨迹的归一化值。安全率由无人机系统仿真任务故障率的补数得出。我们针对三种电池健康状况初始条件——良好、中等和差——评估了这些指标。对于每种策略，在每种可观测性和电池健康状况组合下，均进行了 35 次蒙特卡洛仿真。



(a) (b)

图 9 (a) EODandr 实施策略时剩余飞行时间的变化，(b) 电机电流消耗因电机引起的变化重构。



(a) (b)

图10 (a) 应急管理策略执行的时间序列图，(b) 展示仿真中CMA马尔可夫决策过程策略执行情况的三维图，其中蓝色轨迹为实际飞行计划，品红色轨迹为标称飞行计划，红色星形标记 ★为检查点。

A. CMA案例研究

本节展示了一个示例，说明多旋翼飞行器发生电机性能退化时 CMA马尔可夫决策过程策略的执行情况。电机性能退化最终导致卡滞故障，并对 EOD 产生连锁影响，如图 9 和图 10 所示。在 $161sec$ 检测到剥落故障后，系统会进行故障预测以确定电机剩余使用寿命。CMA MDP 继续建议采取 $NoOp$ 动作，如图10a所示。电机在 $179sec$ 发生故障，交互多模型检测到了该故障。这使六旋翼飞行器重构为四旋翼构型。重构后，四台电机需要提供更大的推力，导致电池电流消耗增加，如图所示

图9b。电池电流消耗的突然增加使EOD值降至意料之外的低水平，如图9a所示，因此EOD值不足以完成飞行计划。CMA MDP识别出这一情况，并建议采取图10a所示的 $LandPract$ 动作。 $LandPract$ 动作会触发搜索合适的、飞行前计算的紧急着陆轨迹，该轨迹从最近的检查点出发，具有最大的可达裕度；最近的检查点在图10b中以红色星号标出。随后选定一条紧急着陆轨迹并发送给GNC。无人机系统执行紧急着陆飞行计划期间，CMA-MDP仍保持激活状态，但可用动作集缩减为 $A = NoOp$、$Terminate$、$LandASAP$。在本案例研究中，无人机系统成功降落在选定的紧急着陆地点。

B. 原飞行计划完成率与安全性

对于所有电池健康度指标，先前的观测MDP $Obs\ MDP$ 在部分可观测条件下表现很差，平均几乎无法完成任何飞行，同时安全性各不相同（图11–13）。连续收到多个错误故障观测时，这种行为有两种解释。第一，策略会因这些观测而选择在飞行很早阶段着陆；在这种情况下，飞行虽安全，但效率低下。第二，这些观测会导致策略在短时间内连续选择多个应急动作。在一些情况下，这种选择会使底层控制器失稳，或导致状态被错误标记为卡滞故障，尽管实际上并不存在卡滞故障，最终造成任务失败；而在现实中，底层控制器会阻止此类情况发生。如果控制器稳定性或误诊不是问题，那么我们预计，在部分可观测条件下，该策略会表现为始终安全但效率低下（无法完成原始任务），因为它将任何故障观测都视为准确无误，并采取应急动作。

相较于其他策略，可以观察到以下趋势。在电池健康状况不佳的情况下（图 11），对于所有可观测度大于 $P_{obs} = 0$。6 的情况，除 $Baseline\ NoOp$ 外，所有策略均表现出保守的策略行为。在这些情况下，$MAP\ MDP$ 的表现优于或可与 $POMDP$ 策略相媲美：在任务完成率相同的情况下，前者实现了更高数量的安全飞行，尽管由于试验次数有限，两者的结果均处于 95%置信区间内。在电池健康状况不佳且可观测度大于 $P_{obs} = 0$。6 的情况下，所有策略的表现均与能够访问完整状态的 $True\ MDP$ 相当。

在可观测度为 0.6 时， $POMDP$ 和 $MAP\ MDP$ 策略明显不那么保守，约有 40% 的概率完成预定任务，而仅有 40% 的飞行能够安全完成。这一表现与 $Baseline\ NoOp$ 策略相当。这种行为变化可能归因于策略更加依赖转移动力学而非观测，因为在这种情况下，观测提供的信息非常有限。由于转移矩阵模型假设旋翼故障的概率非常低，信念将持续集中于无故障状态，并沿着计划轨迹继续运行。

在中等电池健康状况下（图12），我们观察到类似的趋势（注意坐标轴变化）。在这种情况下，策略

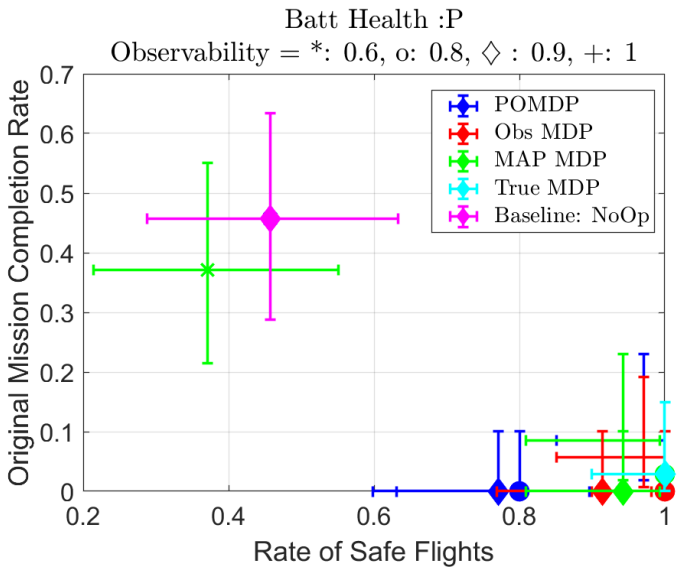


图 11 电池健康状况不佳时原始任务完成率与安全飞行率的关系，附 95% 置信界限。请注意， $P_{obs} = 0.6$ **处的 POMDP策略被** *Baseline NoOp* **策略遮挡。**

效率和安全性更高，这可能反映了电池状况更健康所带来的安全裕度提升。在较高的可观测度下（$P_{obs} = 1$, $P_{obs} = 0.9$, $P_{obs} = 0.8$），*MAP MDP*和 *POMDP*的表现与彼此以及 *True MDP* 策略相当（在95%置信界限内）。在最低可观测度 $P_{obs} = 0.6$时，也观察到同样激进的行为。

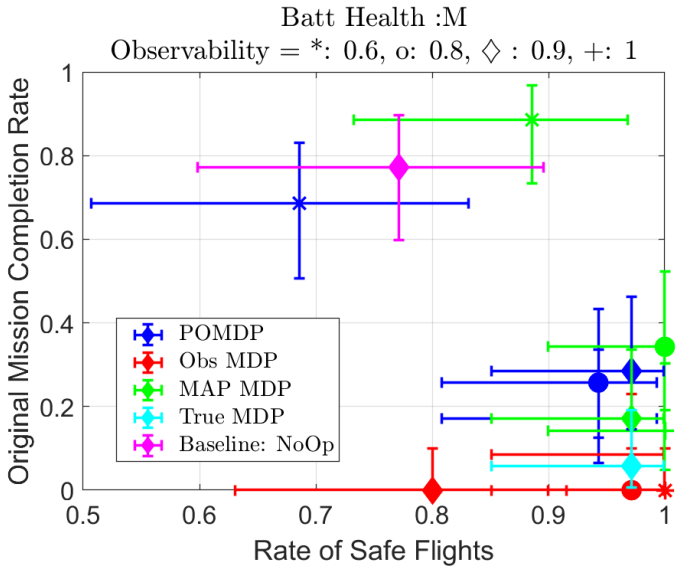


图 12 电池健康状况为中等时原始任务完成率与安全飞行率的关系，附 95% 置信界限。

在电池健康状况良好时（图 13），趋势出现了变化。在 $P_{obs} = 0.6$ 处，就任务完成而言，*MAP MDP* 和 *POMDP*的表现相当，且显著优于 *True MDP*策略。 *POMDP*

在其他可观测度下（P_{obs} = 1、 P_{obs} = 0.9、 P_{obs} = 0.8），就任务完成而言，其表现优于 $MAP\ MDP$策略，而 $MAP\ MDP$的表现则与 $True\ MDP$策略更为接近。有趣的是，部分可观测情形下的表现优于完全可观测性的 $POMDP$和 $MAP\ MDP$策略。这可能延续了此前图表中低可观测性下观察到的同类激进行为，P_{obs} = 0.6。不过，在电池状况良好时，无人机系统更有可能保持安全并完成任务，这与先前的情况不同。值得注意的是，在电池状况良好时， $POMDP$策略优于 $Baseline\ NoOp$策略。

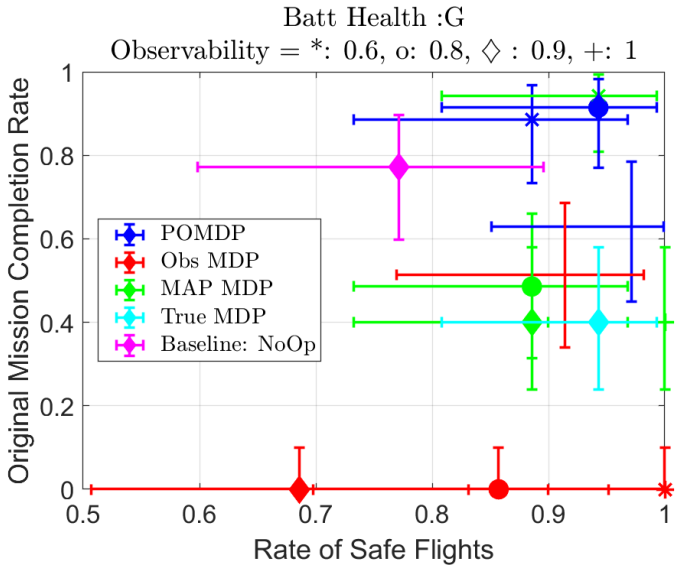


图 13 电池健康状况良好时，原始任务完成率与安全飞行率的关系，置信度为 95%。

C. 基于电池健康状况的 CMA 性能

本节展示了图 14 所示不同电池健康状况和可观测性条件下应急管理策略性能的分析结果。

在完全状态可观测性下，所有应急策略在状态可观测性较差（0.6、0.8）以及完全状态可观测性下的原始飞行计划完成率分别如图 14a、14c 和 14e 所示。观察完全状态可观测性情形（图 14e）可知，随着电池健康度下降，所有策略的表现均有所下降。正如预期， $POMDP$， $Obs\ MDP$、$MAP\ MDP$在完全可观测性下与 $True\ MDP$策略的表现相近。与 $Baseline\ NoOp$ 策略相比，这些策略的任务完成表现均相对较差。这是因为，当 $MM < 0$（即单个电机即将发生故障或已经发生故障）且 $RM > 0$（即能量足以完成飞行计划）时，应急策略会在这些情形下积极执行紧急着陆。在中等和较差电池健康状况下，也表现出类似的积极安全行为。在图 14a 和 14c 所示的状态可观测性较差情形中， $True\ MDP$和 $Baseline\ NoOp$仅作为比较参考，因为它们

不受状态可观测性的影响。 $Obs\ MDP$ 的任务完成率最低，在状态观测噪声很大的情况下，会在飞行计划初期执行紧急着陆。此外，随着状态可观测性下降， $POMDP$和 $MAP\ MDP$策略的表现有所提升。这归因于它们采取了积极完成飞行计划的行为，对安全性的顾虑较少。

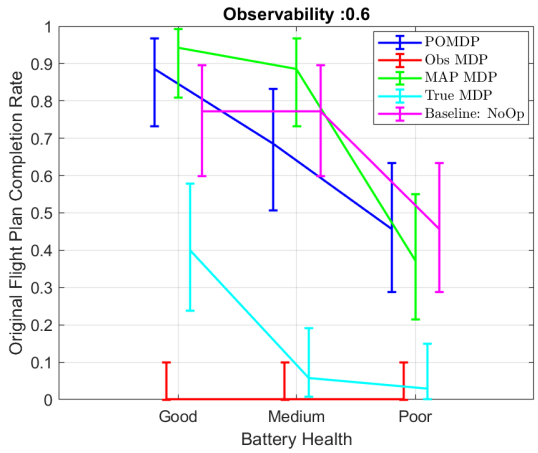
通过分析不同电池健康状况和不同状态可观测性下的安全率， $True\ MDP$和 $Baseline\ NoOp$ 策略分别代表上限和下限。在图 14f 所示的完全可观测性下，CMA MDP 和 POMDP 策略收敛于 $True\ MDP$，并且其安全性高于 $Baseline\ NoOp$策略。在状态可观测性较差时，CMA MDP 和 POMDP 策略收敛于 $Baseline\ NoOp$ 策略。如图 14b 所示，$Obs\ MDP$ 策略可能看起来最安全，但这是因为在状态可观测性较差的场景中，该策略从未完成任何任务，而是在飞行计划早期着陆。总之，应急策略的保守行为带来了更高的安全率，但任务完成率较低。

D. 转移矩阵仿真

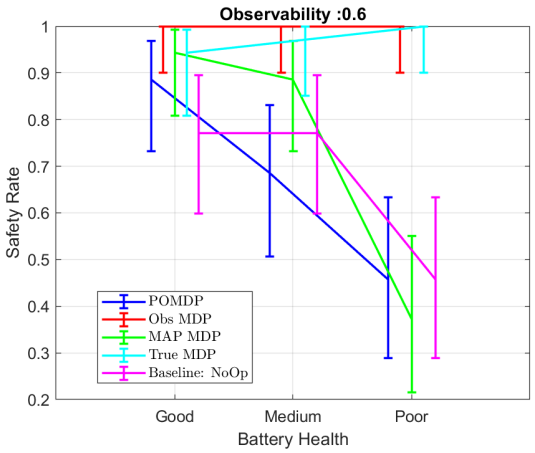
除上述高保真仿真外，还执行了“转移矩阵仿真”。这些仿真将用于计算最优策略的转移矩阵作为评估仿真器。与之相对的是高保真模拟器，后者更接近策略在现实世界中的部署，并使用不同的状态转移动力学。两个模拟器使用相同的观测矩阵。这些转移矩阵仿真可以隔离将抽象 POMDP 模型解决方案应用于高保真仿真时的误差来源。相关的潜在误差来源包括：将现实世界动态抽象为 POMDP 状态转移动力学时存在偏差，以及将期望结果转换为目标时存在偏差。在此前描述的实验中，动力学和目标都可能存在不匹配。相比之下，以下仿真通过在转移矩阵模拟器上评估上述策略的性能，控制了动力学差异的影响，因为在该模拟器中，解决方案与问题的动力学必然匹配。使用转移矩阵模拟器对这些策略完成了 5000 次蒙特卡罗评估，最大仿真步数为 100。

图 15a 展示了任务完成率与安全飞行率的关系，采用与图 11–13 相同的指标，同时控制动态不匹配。在这些仿真中， $POMDP$和 $MAP\ MDP$在转移矩阵仿真中的表现相近，而 $Obs\ MDP$策略虽然安全，但效率低下。在 $POMDP$和 $MAP\ MDP$无法比较的情况下， $POMDP$策略表现得更为保守：完成的任务较少，但安全性更高。在这种不存在动态差异的情况下，我们观察到可能存在目标不匹配，因为 POMDP策略的行为比其他策略更为保守。

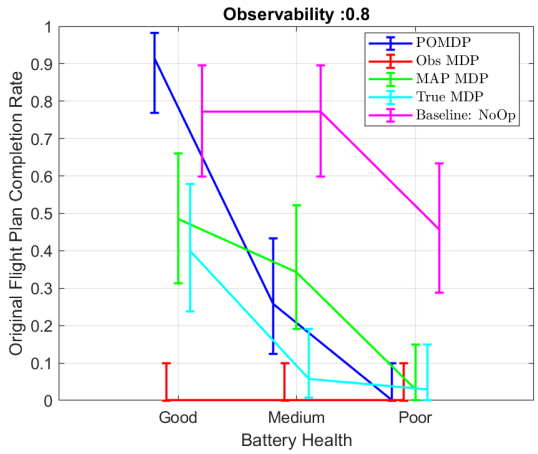
图 15b 展示了控制动态不匹配和目标不匹配后，各策略在不同可观测性程度下的平均累积奖励。因此，该图表明这些策略相对于抽象动态以及分别由状态转移函数和奖励函数定义的目标，是否接近最优。
$MAP\ MDP$和



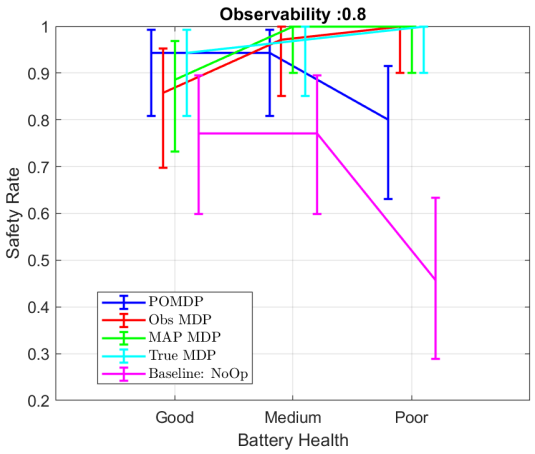
(a) 60% 部分状态可观测性



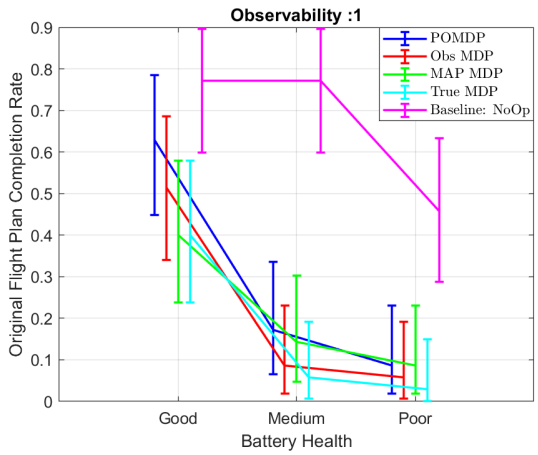
(b) 60% 部分状态可观测性



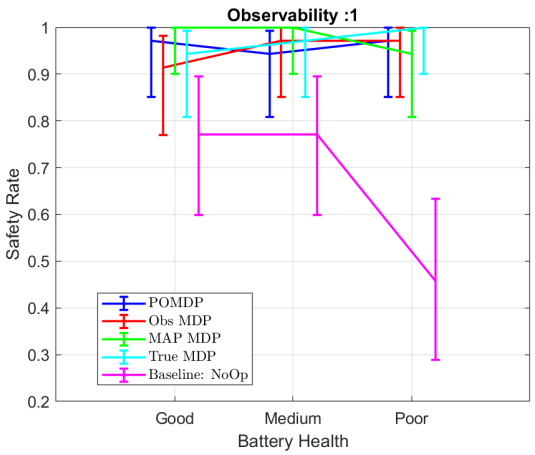
(c) 80% 状态可观测性



(d) 80% 部分状态可观测性



(e) 完全状态可观测性

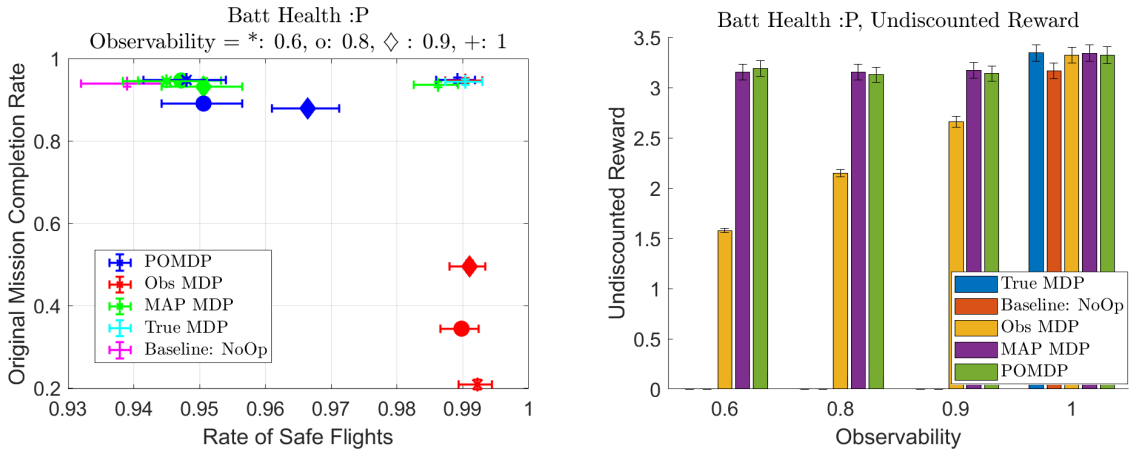


(f) 完全状态可观测性

图 14：高保真仿真中不同电池健康状况下应急策略的安全率和原始飞行计划完成率。 *Baseline NoOp* **和** *True MDP* **策略不受可观测性的影响。**

*POMDP*各策略在不同可观测度下的奖励表现相近。 *POMDP*和 *MAP MDP*在完全可观测时的表现与 *True MDP*和 *Baseline NoOp*策略相当，而在可观测度降低时则明显优于 *Obs MDP*策略。就抽象化部分可观测马尔可夫决策过程表述的奖励而言， *MAP MDP*和 *POMDP*策略接近最优。

请注意，奖励方面的表现相近，并不意味着安全性或效率方面的表现也相当。这是奖励函数设计所致，该函数并非为明确考量上述任务完成和安全性指标而设计。鉴于图 15b 所示的奖励表现接近最优，且图 15a 所示的任务完成率和安全飞行率更高，本研究中的误差最可能源于抽象化的状态转移动力学。



(a) **(b)**

图 15 电池健康状况不佳时的转移矩阵仿真。(a) 原始任务完成率与安全飞行率，含 95% 置信界限。请注意，两轴的刻度不同。(b) 相同仿真的平均（未折扣）奖励，含 2σ 均值标准误界限。请注意， *True MDP* 和 *Baseline NoOp* 策略不受可观测度影响，因此仅显示完全可观测时的结果。

VII. 讨论

A. 使用高保真仿真进行策略评估

上述结果表明，在电池状况不佳时，无人航空系统任务的完成率较低；对于 Poor 和 Medium Battery Conditions 中的*POMDP*和 *MAP MDP*，情况均是如此。CMA 的这种安全保障行为确实使故障数量少于 *Baseline NoOp* 策略。不过，人们会期待最优策略不要那么保守（使标称任务完成情况更接近 *Baseline NoOp* 策略），同时仍能保持较低的故障率。

值得注意的是，高保真模拟器生成了若干马尔可夫决策过程转移矩阵未涵盖的状态转移。有时，由于系统行为分类错误，系统似乎会从无故障直接转变为卡滞故障

。未表示或建模不准确的状态转移可能是观察到的完成率和安全率较低的原因之一。

这些底层状态中的意外状态转移，可能解释了在部分可观测性更高的状态下观察到的更高任务完成率。在低 P_o的仿真中，异常状态会被归因于错误观测；其中，具有 $P_o = 1$ 的确定性观测模型无法将意外状态转移解释为观测噪声。意外的状态转移进一步支持了应考虑含噪观测这一观点。鉴于高保真模拟器所指示的状态本身似乎受到一定噪声的影响，在模型中考虑这种噪声就更加重要了。

除这些趋势外，除了电池健康状况良好的初始条件，在所有其他初始条件下， $POMDP$策略与 $MAP\ MDP$表示的性能都相近，这与转移矩阵仿真的结果一致。尽管$Baseline\ NoOp$ 策略的安全率通常最差，但就原始飞行计划完成率而言，该策略的表现与其他策略相当，甚至更好。尤其值得注意的是，能够直接访问状态（不存在不确定性）的 $True\ MDP$在完成的标称任务数量方面，表现反而不如 $Baseline\ NoOp$ 策略。我们希望看到马尔可夫决策过程的任务完成率表现更接近 $Baseline\ NoOp$ ，同时仍保持低得多的故障率，不过这些指标不可避免地相互制约。由于CMA策略的首要目标是保障无人机系统的安全性，而不是任务完成，因此这些结果并不令人意外。

如 V.B 所述，高保真仿真中使用的剥落故障发生概率高于转移矩阵中的概率。剥落故障发生概率较低时（这也符合现实世界系统中的预期），安全性和效率预计都会提高。因此，这些结果反映了某种最坏情况：在现实环境中，故障发生的可能性远不应如此之高。由于高保真测试所用的样本量相对较小，各项策略和电池状况下，可能发生剥落故障的仿真次数以及实际出现剥落故障的仿真次数并不一致。

B. 使用转移矩阵仿真进行策略评估

如第六节 D 部分所述， $POMDP$和 $MAP\ MDP$策略在奖励、安全性和效率方面表现相近。对此的一种可能解释是，随着时间推移，状态信念始终集中于某一个状态。在这种情况下，我们预计，基于状态最大后验估计的策略与基于完整信念的策略会取得相近的表现。为量化信念分布的集中程度，式 13 引入了最小最大后验状态概率 P_{MinMax}。该值表示一次仿真在整个时间范围内的最小最大后验状态概率。在每个时间步，找出信念中的最高概率，然后从这组最大后验概率中取最小值。由此可以了解信念的

在给定仿真过程中集中于某一状态的程度。

$$P_{MinMax}(i) = \min_{t \in h} \{ \max_{s \in b_t} b_t(s) | t \in h \} \tag{13}$$

其中，h 是仿真的时域，t 是时间步，s 是状态，b 是信念中为各状态分配的概率，i 是仿真次数。

在对固定时域为 100 个时间步（如上所述）的 $MAP\ MDP$ 策略进行的 60000 次转移矩阵仿真中，每种初始电池状态和可观测性各进行 5000 次仿真，该指标的均值为 0.991，标准差为 $\sigma = 0.031$，而最小值为 0.5073。因此，大多数仿真都保持了较低的不确定性信念。这可归因于转移矩阵中编码的向故障状态转移的概率。以确定性转移为例进行考虑。无论观测结果如何，确定性转移都会导致信念集中，因为与确定性转移到的状态相冲突的观测结果，只会被归因于错误观测的概率。因此，近似确定性的概率同样会导致信念集中。

鉴于这些结果，$MAP\ MDP$ 策略可能是适用于此情境的最佳策略。找到该策略所需的计算复杂度，仅相当于求解马尔可夫决策过程的值迭代解所需的计算复杂度。该策略能够维持对状态的信念，因此对含噪观测具有更强的鲁棒性。

VIII. 结论

本文提出了一种基于强大的马尔可夫决策过程框架构建的风险感知应急管理自主系统（CMA）。我们将经过实验验证的六旋翼无人航空系统、推进单元以及电池性能与退化模型纳入马尔可夫决策过程和部分可观测马尔可夫决策过程形式化中。我们在更贴近现实应用的高保真模拟器以及可用于研究模型准确性的转移矩阵模拟器上，对应急管理策略的有效性进行了评估。

我们针对各种组件退化场景和状态可观测度对这些策略进行了评估。我们使用两个指标，即原始飞行计划完成率和安全率，来分析不同策略的表现。在我们的方法中，不考虑主动信息收集和长期不确定性，因此 $MAP\ MDP$ 更为有利。与 $POMDP$ 相比，该策略的计算复杂度更低。$MAP\ MDP$ 策略的表现也优于一种朴素MDP策略，该策略假设所有观测都准确无误。我们认为，在存在观测不确定性且状态转移近乎确定的形式化中，最大后验近似或类似方法可能是最佳折中：这类方法在基于信念进行推理的同时采用MDP策略，从而兼顾解的保真度与计算复杂度。

在未来的工作中，我们旨在完善状态转移模型，使其准确反映模拟器的转移概率。目前，CMA 的成功与否通过仿真进行评估。为进一步验证我们的结果，下一步将

be t开展基于硬件的实验，以确定 CMA 在实践中的有效性。

致谢

本研究经费由行业赞助商通过无人机系统中心（美国国家科学基金会工业—大学合作研究中心）提供，资助编号1650468。

参考文献

[1] Osborne, M, Hutchison, Z, Kabbani, A, Winiarski, T, Clothier, R, Johnson, K, 《无人机系统操作员安全性与可靠性调查：面向自主无人机系统认证的新兴技术》，*2019*年第四届系统可靠性与安全国际会议（*ICSRS*），IEEE，2019年，第203–212页。

[2] Ochoa, C. A.，以及Atkins, E. M.，《小型无人机系统运动规划的城市度量地图》，航空航天信息系统期刊，第19卷，第1期，2022年，第37–52页。

[3] Di Donato, P. F., and Atkins, E. M., “航空器紧急着陆规划中人员与财产风险的评估”，《航空航天信息系统期刊》，第14卷，第5期，2017年，第259–278页。

[4] E.Balaban, and J.J.Alonso, “预测性决策建模框架及其在无人机任务规划中的应用”，故障预测与健康管理协会年会，2013年，第1–12页。

[5] Balaban, E., S.Narasimhan, M.Daigle, I.Roychoudhury, A.Sweet, C.Bond, and G.Gorospe, “移动机器人测试平台的开发及预测性决策算法验证方法”，《国际预测与健康管理期刊》，第4卷，第1期，2013年，第87页。

[6] Schacht-Rodríguez, R., Ponsart, J. C., García-Beltrán, C. D., Astorga-Zaragoza, C. M. 和 Theilliol, D.，《基于飞行续航时间估算的多旋翼无人机任务规划策略*》，无人机系统国际会议（*ICUAS*），2019年，第778–786页。URL https://doi.org/10.1109/ICUAS.2019.8798292。

[7] Tang, L.、Kacprzynski, G. J.、Goebel, K.、Saxena, A.、Saha, B. 和 Vachtsevanos, G.，《面向先进自主系统的预测增强型自动应急管理》，故障预测与健康管理国际会议，2008年，第1–9页。URL https:/doi/org/10.1109/PHM.2008.4711448。

[8] Tang, L., Kacprzynski, G. J., Goebel, K. 和 Vachtsevanos, G.，《面向航空器系统的预测增强型自动应急管理案例研究》，*IEEE* 航空航天会议，2010年，第 1–11 页。网址：https://doi.org/10.1109/AERO. 2010.5446844。

[9] Zhang, B., L.Tang, J.Decastro, M.Roemer 和 Goebel, K.，《基于自主车辆电池荷电状态预测的增强型任务规划》，《国际预测与健康管理期刊》，第 5 卷，第 8 期，2014年。

[10] Schacht-Rodríguez, R., Ponsart, J.-C., Garcia-Beltran, C. D. 和 Astorga-Zaragoza, C. M.，《用于预测无人机飞行续航时间的故障预测与健康管理》，*IFAC-PapersOnLine*，第 51 卷，第 24 期，2018年，第 983–990 页。

[11] He, W., Williard, N., Chen, C. 和 Pecht, M.，《采用无迹卡尔曼滤波进行电动汽车电池荷电状态估计》，*Microelectronics Reliability*，第 53 卷，第 6 期，2013年，第 840–847 页。[12] Daigle, M. J. 和 Goebel, K. F.，《利用无迹变换提高基于模型的故障预测中的预测计算效率》，预测与健康年度会议，2010年。[13] Dalal, M., Ma, J. 和 He, D.，《采用粒子滤波框架的锂离子电池寿命预测与健康管理系统》，英国机械工程师学会论文集，*O* 部分：风险与可靠性期刊，第 225 卷，第 1 期，2011 年，第 81–90 页。

[14] Obeid, A., Tariq, U., and Mukhopadhyay, S., “用于早期、准确检测电池端电压骤降的监督学习”，*IET*电路、器件与系统, Vol.14, No. 3,2020年, pp. 347–356.

[15] Wu, L., Fu, X., and Guan, Y., “使用数据驱动方法综述车辆锂离子电池的剩余使用寿命预测”，*Applied Sciences*, Vol. 6, No. 6, 2016, p. 166.

[16] Liu, D., Pang,J., Zhou, J., Peng, Y., and Pecht,M., “基于组合高斯过程函数回归的锂离子电池健康状态估计预测”, *Microelectronics Reliability*, Vol. 53, No. 6, 2013年, pp. 832–839.

[17] Sharma, P., and Atkins, E., “多电池小型无人机系统能源系统中的预测信息驱动电池重构”，*2021*年国际无人机系统会议 *(ICUAS)*, IEEE, 2021, pp.423–432.

[18] Brown, J.M., Coffey, J. A., Harvey, D. 和 Thayer, J. M.，“多旋翼故障的表征与故障预测”，结构健康监测与损伤检测，第 *7* 卷，Springer，2015年，第 157–173 页。

[19] Zhang, Y.、Chamseddine, A.、Rabbath, C. A.、Gordon, B. W.、Su, C.-Y.、Rakheja, S.、Fulford, C.、Apkarian, J. 和 Gosselin, P.，“先进故障检测与诊断（FDD）及故障容错控制（FTC）技术的开发及其在无人四旋翼直升机试验台中的应用”，富兰克林研究所期刊，第 350 卷，第 9 期，2013年，第 2396–2422 页。

[20] Dydek, Z. T.、Annaswamy, A. M. 和 Lavretsky, E.，“四旋翼飞行器 UAV 的自适应控制：一项结合飞行评估的设计权衡研究”，*IEEE* 控制系统技术汇刊，第 21 卷，第 4 期，2012年，第 1400–1406 页。

[21] Mueller, M. W., and D’Andrea, R., “尽管一、两个或三个螺旋桨完全失效，四旋翼飞行器的稳定性与控制”，*2014*年*IEEE* 国际机器人与自动化会议（*ICRA*），IEEE，2014年，第45–52页。

[22] Kim, J., Sharma, P., Atkins, E., Neogi, N., Dill, E., and Young, S., “先进空中交通的保障应急着陆管理,” *2021*年*IEEE*/美国航空航天学会第 *40* 届数字航空电子系统会议（*DASC*）, IEEE, 2021, pp. 1–12.[23] Ten Harmsel, A. J., Olson, I. J., and Atkins, E. M., “能量受限多旋翼飞行器的紧急飞行规划,” 智能与机器人系统期刊, Vol. 85, No. 1, 2017年, pp. 145–165.[24] Sunberg, Z., Ho, C., and Kochenderfer, M., “推断交通参与者内部状态对高速公路自动驾驶的价值,” , 2017年. https://doi.org/10.48550/ARXIV.1702.00858, URL https://arxiv.org/abs/1702.00858.

[25] Kurniawati, H., “部分可观测马尔可夫决策过程与机器人学,” 《控制、机器人与自主系统年度评论》, Vol. 5, 2022年.[26] Kaelbling, L. P., Littman, M. L., and Cassandra, A. R., “部分可观测随机领域中的规划与行动,”人工智能, Vol. 101, No. 1-2, 1998年, pp. 99–134.[27] Kochenderfer, M. J., Holland, J.E., and Chryssanthacopoulos, J. P., “新一代机载防碰撞系统,” Tech. rep., 麻省理工学院林肯实验室，列克星敦，美国, 2012年.

[28] Goel, P.、Dedeoglu, G.、Roumeliotis, S. I. 和 Sukhatme, G. S.，“使用多模型估计和神经网络进行移动机器人故障检测与识别”，*IEEE* 国际机器人与自动化会议论文集，第 3 卷，2000年，第 2302–2309 页。https://doi.org/10.1109/ROBOT.2000.846370。

[29] Mehra, R.、Rago, C. 和Seereeram, S.，“面向航空航天应用的自主故障检测、识别与容错估计”，*IEEE*航空航天会议论文集，第 2 卷，1998年，第 133–138 页。https://doi.org/10.1109/AERO.1998.687904。

[30] Zhang, Y. 和 Li, X. R.，“使用交互多模型估计器检测和诊断传感器与执行器故障”，*IEEE TransactionsonAerospaceandElectronic Systems*，第 34 卷，1998年，第 1293–1313 页。https://doi.org/10.1109/7.722715。

[31] Kochenderfer, M. J., Wheeler, T. A., and Wray, K. H.,决策算法，麻省理工学院出版社，2022年。

[32] Littman, M. L., Cassandra, A. R., and Kaelbling, L. P.，《部分可观测环境中的策略学习：扩展规模》，机器学习论文集 *1995*年，爱思唯尔，1995年，第362–370页。

[33] Kurniawati, H., Hsu, D., and Lee, W. S.，《SARSOP：通过近似最优可达信念空间实现高效的基于点的部分可观测马尔可夫决策过程规划》，机器人学：科学与系统会议论文集，2008年。

[34] Kim, J., and Atkins, E.，《低空城市小型无人机系统的空域地理围栏与飞行规划》，*Applied Sciences*，第12卷，第2期，2022年，第576页。

[35] M.L.Putterman, 马尔可夫决策过程：离散随机动态规划，John Wiley & Sons，2005年。URL https://doi.org/10.1002/9780470316887.[36] Qiu, C., Wu, X., Xu, C., Qiu, X., and Xue, Z.,“通过跟踪电流中的多种振动频率，近似估算感应电动机剥落滚动轴承的故障尺寸”，*Sensors*，第20卷，第6期，2020年。https://doi.org/10.3390/s20061631，URL https://www.mdpi.com/1424-8220/20/6/1631.[37] Zhang, B., Sconyers, C., Patrick, R., and Vachtsevanos, G.,“一种用于工程系统故障诊断和故障预测的多故障建模方法”，故障预测与健康管理协会年会，第1465号，2009年。[38] Atkins, E. M., Portillo, I. A., and Strube, M. J.,“推力完全丧失情况下的紧急飞行规划”，*Journal ofaircraft*，第43卷，第4期，2006年，第1205–1216页。[39] Ng, A. Y., Russell, S., et al.,“逆强化学习算法”。*ICML*，第1卷，2000年，第2页。

[40] Ye, Y.,“对于固定折扣率的马尔可夫决策问题，单纯形法和策略迭代法具有强多项式时间复杂度”，运筹学数学，第36卷，第4期，2011年，第593–603页。https://doi.org/10.1287/moor.1110. 0516，URL https://pubsonline.informs.org/doi/10.1287/moor.1110.0516，出版商：INFORMS。

[41] Sharma, P. 和 Atkins, E.，“牵引式与推进式六旋翼飞行器性能的实验研究”，*Journal of Aircraft*，第 56 卷，第 5 期，2019年，第 1920–1934 页。https://doi.org/10.2514/1.C035319，网址：https://doi.org/10.2514/1.C035319。

[42] Bar-Shalom, Y.、Kirubarajan, T. 和 Li, X.-R.，《用于跟踪与导航的估计》，John Wiley & Sons，2002 年，第 11 章，第 453–457 页。网址：https://doi.org/10.1002/0471221279。

[43] Blom, H. A. P. 和 Bar-Shalom, Y.，“具有马尔可夫切换系数的系统的交互式多模型算法”，*IEEE* 自动控制汇刊，第 33 卷，第 8 期，1988 年 8 月，第 780–783 页。https://doi.org/10.1109/9.1299。