BRIEF PAPER

Integrated Clustering and Mission Planning for Agile Earth Observation Satellite Constellations*

Xiaohe HE[†,††,†††a)], ***Member***, **Junyan XIANG**[†,††,†††], **Mubiao YAN**[†,††], **Chengxi ZHANG**[††††], **Zhuochen XIE**[†,††b)], ***and*** **Xuwen LIANG**[†,††,†††c)], ***Nonmembers***

SUMMARY The Agile Earth Observation Satellite Constellations Mission Planning (AEOSCMP) problem seeks to maximize global cumulative reward by optimizing task selection and scheduling across the Earth's surface while adhering to the intricate resource constraints of individual satellites. This optimization challenge is further complicated by the diverse observation intervals required for different targets and the necessity for coordinated action among multiple satellites, introducing complexities in synchronization, data consistency, and overall mission planning. Deep reinforcement learning (DRL) and target clustering represent two complementary methodologies that synergistically enhance the autonomy and observation efficiency of AEOSCMP. This letter introduces an innovative approach that elegantly unifies these two methodologies - the Integrated Clustering and Planning with Proximal Policy Optimization Algorithm (ICP3O). This sophisticated framework seamlessly preserves the intelligent decision-making capabilities inherent to DRL while delivering substantial improvements in observation efficiency.

key words: *Earth observation, Agile satellite constellation, Mission planning, Task scheduling, Target clustering, DRL*

1. Introduction

Earth observation satellites (EOS) play a crucial role in global monitoring and management, providing essential data for various applications such as environmental change tracking, weather prediction, disaster management, and agricultural optimization. These satellites are fundamental to numerous fields, including natural resource management, urban planning, environmental protection, and scientific research advancement. Agile Earth observation satellites (AEOS) are equipped with advanced maneuverability, allowing them to make rapid and precise attitude adjustments along the roll, pitch, and yaw axes. This critical capability sets them apart from traditional EOS, significantly enhancing their operational flexibility. As the demand for high-resolution, timely observational data grows and satellite constellations expand, the need for developing and implementing autonomous, efficient mission planning methods for agile satellite constellations has become increasingly critical [1].

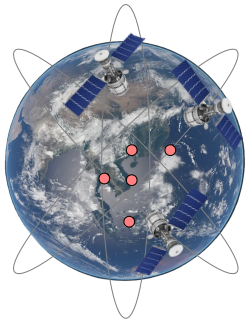

Fig. 1 Description of AEOSCMP problem.

The satellite mission planning problem has been extensively studied for over two decades, beginning with its initial formulation by Lemaitre [2]. Researchers have proposed various solutions, including exact algorithms[3], [4], heuristic approaches, and meta-heuristic techniques[5]. However, both exact and heuristic methods face limitations in dynamic environments due to their open-loop nature. When an observation plan fails or new requirements arise, these methods require recalculations of the entire plan. Recent studies [6]–[9] have shown the considerable potential of DRL in addressing the AEOSCMP problem. Notably, He et al. [9] reformulated the AEOSCMP problem as a sequential decision-making process, achieving monotonic improvement.

Besides, target clustering represents a methodological advancement in Earth observation. This technique strategically groups targets exhibiting similar characteristics into clusters, effectively reducing the problem's computational complexity. The approach has demonstrated considerable success across various Earth observation applications, particularly in remote sensing [10]–[12]. However, their methods typically treat target clustering and scheduling as separate processes, potentially compromising overall system performance. In contrast, our work organically integrates target clustering with scheduling optimization, resulting in substantially enhanced observation efficiency.

2. Problem Formulation

This section presents a comprehensive formulation of the

Manuscript received December 13, 2024.
Manuscript revised April 12, 2025.

[†]University of Chinese Academy of Sciences, No.1 Yanqihu East Rd, Huairou District, Beijing,101408, China
[††]Innovation Academy for Microsatellites of CAS, No.1 Xueyang Rd, Pudong District, Shanghai,201304, China
[†††]Shanghaitech University,393 Middle Huaxia Rd, Pudong District, Shanghai,201210, China
[††††]Jiangnan University, Lihu Avenue, Wuxi, 214122, China

*This work is supported by the National Key Research and Development Program of China (2022YFB3902801).

a) E-mail: hithxh@gmail.com
b) E-mail: xiezc@microsat.com (Corresponding Author)
c) E-mail: liangxw@shanghaitech.edu.cn (Corresponding Author)

DOI: 10.1587/transele.E108.C.1

AEOSCMP problem. We begin by elucidating the problem objectives and underlying simulation models. Subsequently, we formalize the environment as a decentralized, partially observable Markov decision process (Dec-POMDP).

2.1 AEOSCMP Problem Description

As depicted in Figure 1, the AEOSCMP problem involves multiple AEOS satellites arranged in a walk-delta constellation. These satellites collaboratively observe terrestrial targets to maximize the cumulative value of uniquely imaged targets.

Each AEOS is equipped with three essential subsystems: The Attitude Determination and Control System (ADCS), the Power Management System (PMS), and the Data Management System (DMS). Especially, the ADCS employs reaction wheels and a thruster to adjust the satellite's attitude across various operational modes precisely. The satellite is steered toward targets using an attitude controller based on modified Rodrigues parameters (MRP) and rate servos for the reaction wheels [13]. The full simulation setup resembles that presented in [14].

In the AEOSCMP problem, observation targets are points or simple polygons. For polygon targets, a spherical geometry-based tessellation algorithm divides them into a set of feasible tiles, as proposed by Eddy et al. [15]. This study primarily focuses on point targets. Let c_j represent a target in the set N. Initially, all requests are in the unfulfilled set $U = N$. When a request is imaged, it is removed from U and added to the fulfilled set F. The set F' represents the fulfilled set in the next step.

The environment and satellite models are implemented using the Basilisk framework [16], which offers several advantages over traditional mathematical models: 1) It provides precise satellite orbit simulation accounting for atmospheric drag and J2 perturbations, thus offering a realistic representation of Low Earth Orbit (LEO) dynamics. 2) It enables detailed modeling of critical satellite subsystems, particularly the ADCS. 3) It facilitates the development of robust algorithms that are both theoretically sound and practically viable, enhancing the transferability from simulation to real-world implementation. 4) It allows for an integrated examination of the AEOSMP problem, considering the complex interplay between orbital mechanics, satellite subsystems, and environmental factors, providing a comprehensive and high-fidelity simulation environment.

2.2 Dec-POMDP Formulation

We model the AEOSCMP problem as a Dec-POMDP, where each agent can observe only a limited subset of the state space. Agents make decisions based on their individual, potentially noisy, or incomplete observations and must collaborate with other agents to maximize the expected cumulative reward over time. Formally, a Dec-POMDP is represented by a tuple $(S, A, \Omega, T, R, O, \gamma)$, where:

- **State Space (S):** A finite set representing the environment's states, defined as $S = S_1 \times \cdots \times S_s \times S_{env}$, where S_i represents the state space of satellite i, and S_{env} represents the state of the overall environment. Specifically, the state space includes satellite position and velocity in the Earth-Centered, Earth-Fixed (ECEF) frame, target information, attitude change velocity, battery remaining fraction, and hard disk storage fraction.
- **Action Space (A):** The combined action space of all satellites $A = A_1 \times \cdots \times A_s$, where A_i represents the action space of satellite i. The action space encompasses three primary operations: imaging one of the next ten unfulfilled targets, initiating a battery charging sequence, or executing reaction wheel desaturation maneuvers. Each satellite autonomously selects its actions based on its current state and the learned policy parameters. The success of imaging actions is contingent upon satisfying both temporal and resource constraints.
- **Observation Space (Ω) and Observation Function (O):** The combined observation space $\Omega = \Omega_1 \times \cdots \times \Omega_s$ is composed of individual observation spaces Ω_i for each satellite i. Observations are subsets of the state space, represented as $O(s) \in s$. The observation function O specifies the probability of obtaining a combined observation Ω given the new state s' and the combined action a.
- **Transition Function (T):** The transition function defines the probabilistic state transition. In this study, the transitions are deterministic and generated by the Basilisk simulator, resulting in $T(s'|s, a) = 1$.
- **Reward Function (R):** Reward function specifies the immediate reward received after performing action a in state s. The reward function in a step is:

$$R(s, s') = \sum_{c_j \in N} \begin{cases} p_j & c_j \in U \text{ and } c_j \in F' \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

  where p_j is the priority of target c_j.
- **Discount Factor (γ):** $\gamma \in [0, 1]$ models the trade-off between immediate and future rewards, reflecting the relative value of future rewards.

3. Methodology

To tackle the AEOSCMP problem, we introduce the Simultaneous Clustering and Planning with Proximal Policy Optimization Algorithm (ICP3O) as shown in Algorithm 2.

First, targets close to each other are clustered. For simplicity, we use the distance between targets as the clustering criterion. The spatial relationships between targets are represented with a binary adjacency measure based on the inter-target distance $l_{jj'}$ and a predefined threshold l (set to 100 km in this study).

$$\text{adj}(c_j, c_{j'}) = \begin{cases} 1 & \text{if } l_{jj'} < l \text{ and } j \neq j' \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

This measure creates a spatial graph of targets, serving as

Algorithm 1 PPO for AEOSCMP

1: Initialize PPO parameters θ
2: **for** each episode **do**
3: Observe current state s_t
4: Select and execute action a_t based on policy π_θ
5: Observe reward r_t and next state s_{t+1}
6: Update policy π_θ using Eq. 4
7: **end for**

Algorithm 2 ICP3O for AEOSCMP

1: Initialize adjacency threshold and PPO parameters θ
2: Perform adjacency marking (Eq. 2) and clustering (Eq. 3)
3: **for** each episode **do**
4: Observe current state s_t
5: Select and execute action a_t based on policy π_θ
6: Observe reward r_t and next state s_{t+1}
7: Update cluster observations: If target c_j is observed, mark all unobserved targets in cluster(c_j) as potentially observable
8: Update policy π_θ using Eq. 4
9: **end for**

the foundation for our clustering approach.

Then, for each target c_j, its cluster is defined as:

$$\text{cluster}(c_j) = \{c_{j'} \mid \text{adj}(c_j, c_{j'}) = 1\} \tag{3}$$

This clustering approach enables overlapping clusters, effectively capturing the continuous spatial relationships in satellite observation scenarios.

Finally, algorithm 1 is employed to find the best combination of targets to observe. Specifically, the policy π_θ is optimized using the PPO objective, which includes policy loss $L_i^{CLIP}(\theta)$, value function loss $L_i^{VF}(\theta)$, and an entropy $S[\pi_\theta](s_i)$ to ensure sufficient exploration[17]:

$$L_i(\theta) = \hat{\mathbb{E}}_i \left[L_i^{CLIP}(\theta) - \alpha_1 L_i^{VF}(\theta) + \alpha_2 S[\pi_\theta](s_i) \right] \tag{4}$$

where

$$L_i^{CLIP}(\theta) = \hat{\mathbb{E}}_i \left[\min \left(r_i(\theta)\hat{A}_i, \text{clip}\left(r_i(\theta), 1-\epsilon, 1+\epsilon\right)\hat{A}_i \right) \right] \tag{5}$$

$$L_i^{VF}(\theta) = \hat{\mathbb{E}}_i \left[(V_\theta(s_i) - V)^2 \right] \tag{6}$$

$$S[\pi_\theta](s_i) = -\sum_{a \in \mathcal{A}} \pi_\theta(a|s_i) \log \pi_\theta(a|s_i) \tag{7}$$

$L_i^{CLIP}(\theta)$ stabilizes policy updates by using a clipped probability ratio $r_i(\theta) = \frac{\pi_\theta(a_i|s_i)}{\pi_{\theta_{old}}(a_i|s_i)}$ to constrain update magnitudes. $\hat{A}_i$ is the advantage function, and ϵ is the clip range of the probability ratio. $L_i^{VF}(\theta)$ enhances state-value estimations by minimizing the squared difference between the predicted state value $V_\theta(s_i)$ and the target value V. α_1 and α_2 are hyperparameters.

4. Simulation and Discussion

In this section, we present the simulation scenarios, results, and analyses to evaluate the performance of the proposed approach in addressing the AEOSCMP problem.

Table 1 ICP3O parameters.

Parameter	Value
Learning rate	0.003
Discount factor (γ)	0.999
Batch size	5000
Number of epochs	10
Hidden layer of actor and critic networks	[512, 512]
GAE parameter (λ)	0.95
α_1	0.5
α_2	0.01
Clip range ϵ	0.1

4.1 Simulation Scenarios

To rigorously evaluate the efficacy of our proposed approach, we designed and implemented a comprehensive experimental framework comprising multiple simulation scenarios.

Our experimental design incorporates two fundamental dimensions that characterize the simulation scenarios: the scale of observation targets and the satellite constellation configuration. This dual-dimensional approach systematically evaluates our algorithm's capabilities under varying operational conditions.

In terms of observation targets, we conducted experiments across three distinct scales: 1000, 2000, and 3000 targets. These targets were distributed uniformly across the Earth's surface, providing a controlled environment for assessment. This systematic increment in target quantity facilitates a thorough analysis of our algorithm's performance characteristics as the problem complexity increases.

The satellite constellation configurations were designed to examine different operational paradigms. We implemented three distinct configurations:

- 1sat: A single-satellite configuration serving as the baseline scenario, representing minimal resource availability
- 3s3p: A distributed configuration of three satellites across three orbital planes, introducing complex inter-satellite coordination requirements
- 3s1p: A configuration of three satellites in a single orbital plane, presenting unique challenges in coverage optimization and resource allocation

Specifically, the satellite constellation operates in a walk-delta configuration at an orbital altitude of 500 kilometers with a 45-degree inclination. To ensure optimal coverage of the observation area, the satellite cluster is configured with a single unit maintaining 30-degree spacing. The simulation duration corresponds to one complete orbital period, approximately 96 minutes, allowing for a thorough assessment of the system's performance over a full orbital cycle.

In our simulation, Algorithm configuration are shown in Table 1. Specifically, the learning rate is set to 0.003 for stable training, and the discount factor (γ) is set to 0.999 to prioritize long-term reward optimization. The actor and critic networks of ICP3O are both equipped with two hidden

Table 2 Comparative analysis of PPO and ICP3O performance in observation reward.

Targets	Sats	PPO	ICP3O	Improv (%)
1000	1sat	19.84	21.17	6.70
1000	3s1p	58.63	62.12	5.95
1000	3s3p	59.86	63.77	6.52
2000	1sat	20.00	22.42	12.10
2000	3s1p	59.54	66.84	12.26
2000	3s3p	59.93	67.32	12.33
3000	1sat	20.00	23.74	18.70
3000	3s1p	59.82	69.78	16.65
3000	3s3p	59.98	70.24	17.11

layers, each comprising 512 units. GAE parameter (λ) is set to 0.95 to reduce the variance of the advantage estimate. The coefficients of loss function and entropy are set to $\alpha_1 = 0.5$ and $\alpha_2 = 0.01$ to balance optimization objectives. The clip range ϵ is set to 0.1 to prevent excessive policy updates.

4.2 Simulation Results

To compare the performance of PPO and ICP3O, we calculate the observation reward of targets in a one-orbital-period simulation. The results are shown in Table 2.

We first analyze the performance of ICP3O and PPO in terms of target quantity. From table 2, we could find that as the number of targets increases from 1000 to 3000, ICP3O outperforms PPO and becomes more pronounced, ranging from approximately 6.39% to 17.49%. And through different satellite configurations, ICP3O outperforms PPO in all scenarios, indicating its adaptability.

From the collected simulation information, taking the 3000-target scenario as an example, approximately 250 targets (8.3%) have 1-2 neighboring targets within the clustering threshold distance. When examining the actual observation tasks executed by the satellites, we found that about 18% (3-4 out of 20) of the observed targets had neighbors, indicating that ICP3O effectively improved the observation efficiency of spatially correlated targets.

ICP3O's superior performance is due to three key factors: First, clustering allows satellites to capture multiple nearby targets in one pass. Second, considering spatial relationships during planning leads to better observation sequences. Finally, integrating clustering and planning enables dynamic adaptation between strategies and decisions.

5. Conclusion

In this study, we proposed a novel approach to tackle the AEOSCMP problem that combines target clustering with DRL policies. Our method significantly enhances observation efficiency, especially as the number of targets increases, allowing satellites to dynamically adapt to rapidly changing mission requirements and respond swiftly to evolving environmental conditions. Specifically, ICP3O outperforms PPO from approximately 6.39% to 17.49% as the number of targets increases from 1000 to 3000, showing consistent performance improvements.

References

[1] X. Wang, G. Wu, L. Xing, and W. Pedrycz, "Agile earth observation satellite scheduling over 20 years: Formulations, methods, and future directions," IEEE Systems Journal, vol.15, no.3, pp.3881–3892, 2020.

[2] M. Lemaıtre, G. Verfaillie, F. Jouhaud, J.M. Lachiver, and N. Bataille, "Selecting and scheduling observations of agile satellites," Aerospace Science and Technology, vol.6, no.5, pp.367–381, 2002.

[3] G. Peng, G. Song, L. Xing, A. Gunawan, and P. Vansteenwegen, "An exact algorithm for agile earth observation satellite scheduling with time-dependent profits," Computers & Operations Research, vol.120, p.104946, 2020.

[4] X. Chu, Y. Chen, and Y. Tan, "An anytime branch and bound algorithm for agile earth observation satellite onboard scheduling," Advances in Space Research, vol.60, no.9, pp.2077–2090, 2017.

[5] X. Liu, G. Laporte, Y. Chen, and R. He, "An adaptive large neighborhood search metaheuristic for agile satellite scheduling with time-dependent transition time," Computers & Operations Research, vol.86, pp.41–53, 2017.

[6] A. Herrmann, M. Stephenson, and H. Schaub, "Reinforcement learning for multi-satellite agile earth observing scheduling under various communication assumptions," AAS Rocky Mountain GN&C Conference, 2023.

[7] M. Stephenson and H. Schaub, "Reinforcement learning for earth-observing satellite autonomy with event-based task intervals," AAS Rocky Mountain GN&C Conference, Breckenridge, CO, 2024.

[8] L. Dalin, W. Haijiao, Y. Zhen, G. Yanfeng, and S. Shi, "An online distributed satellite cooperative observation scheduling algorithm based on multiagent deep reinforcement learning," IEEE Geoscience and Remote Sensing Letters, vol.18, no.11, pp.1901–1905, 2020.

[9] X. He, J. Xiang, M. Yan, C. Zhang, Z. Xie, and X. Liang, "Agile earth observation satellite constellation mission planning based on multi-agent transformer," IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences, 2025.

[10] G. Wu, J. Liu, M. Ma, and D. Qiu, "A two-phase scheduling method with the consideration of task clustering for earth observing satellites," Computers & Operations Research, vol.40, no.7, pp.1884–1894, 2013.

[11] L. Xiaolu, B. Baocun, C. Yingwu, and Y. Feng, "Multi satellites scheduling algorithm based on task merging mechanism," Applied Mathematics and Computation, vol.230, pp.687–700, 2014.

[12] J. Long, S. Chen, C. Li, and J. Liu, "A task clustering method for multi agile satellite based on clique partition," 2018 11th International Conference on Intelligent Computation Technology and Automation (ICICTA), pp.332–336, IEEE, 2018.

[13] H. Schaub and S. Piggott, "Speed-constrained three-axes attitude control using kinematic steering," Acta Astronautica, vol.147, pp.1–8, 2018.

[14] M. Stephenson, L. Mantovani, and H. Schaub, "Intent sharing for emergent collaboration in autonomous earth observing constellations," AAS Rocky Mountain GN&C Conference, Breckenridge, CO, 2024.

[15] D. Eddy and M.J. Kochenderfer, "A maximum independent set method for scheduling earth-observing satellite constellations," Journal of Spacecraft and Rockets, vol.58, no.5, pp.1416–1429, 2021.

[16] P.W. Kenneally, S. Piggott, and H. Schaub, "Basilisk: A flexible, scalable and modular astrodynamics simulation framework," Journal of aerospace information systems, vol.17, no.9, pp.496–507, 2020.

[17] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint arXiv:1707.06347, 2017.