Resilient Operation Planning for CubeSat Using Reinforcement Learning

Shuntaro Kuroiwa[1], Nozmu Kogiso[1]

[1] *Osaka Metropolitan University, 1-1 Gakuen-Cho, Naka-Ku, Sakai, Osaka, 599-8531, Japan*
sk22175c@st.omu.ac.jp
kogiso@omu.ac.jp

ABSTRACT

This study proposes an autonomous operation procedure for a CubeSat by applying reinforcement learning based on resilient engineering. The CubeSat requires rapid judgement in every visible window based on sufficient understanding of the health conditions of the satellite from limited telemetry data due to the limited communication performance and poor protection functions from the harsh environment. This study first performs a risk analysis by using System Theoretic Process Analysis (STPA) to evaluate the risk scenario of the CubeSat. In order to successfully operate the missions with avoiding the risk scenarios, the reinforcement learning is applied to learn adequate behaviors according to the satellite situations such as temperature and voltage of the installed battery, the sunlight and eclipse phase and the mission progress and plan. Through numerical examples, validity of the proposed method is illustrated.

1. INTRODUCTION

The Small Spacecraft System Research Center at Osaka Prefecture University (now Osaka Metropolitan University) and the Aerospace System Research Center at Muroran Institute of Technology jointly developed a CubeSat called "Hirogari" from 2016 to 2020 (Iida et al., 2018; Hashiwaki et al., 2019). It was successfully operated from March 2021 to April 2022 and achieved full success (Osaka Prefecture University, 2021). However, the CubeSat experienced some failures during its early operation stage, because our team did not sufficiently consider operational risks. Our team struggled to identify the causes and determine the feasible recovery method. (Nakase et al., 2021).

CubeSats have limitations in protecting against harsh environmental conditions in orbit due to size and weight constraints, and also have strict limitations such as power margins. Operating a CubeSat with limited operational capabilities for a designed operational period, even without any failures, places a significant burden on the operations team.

To develop an efficient operational plan for the next developing CubeSat that reduces such operational burdens, our research group applied System Theoretic Process Analysis (STPA) (Leveson, 2012), one of the hazard analyses, to "Hirogari"(Yamada et al., 2022). Yamada et al. first defined the loss and hazard of the CubeSat and the mission, then focused on the information flow between the ground station and the CubeSat to extract "unsafe" commands or scenarios. Through the STPA process, the research extracted unsafe control actions (UCAs) that activate hazards. Then, the study identified hazard activating factors and scenarios that are difficult to identify using conventional hazard analysis such as Fault Tree Analysis (FTA). Finally, the study demonstrated that resilient operation of the CubeSat by considering these loss scenarios.

However, the identification of loss scenarios alone does not lead to a reduction in the operational burden if the efficient hazard inactivation methods cannot be implemented. Therefore, in this study, we further subdivided the elements of STPA to seek more concrete methods for mission operations and aimed to automatically generate mission operation methods using the results obtained by STPA and reinforcement learning.

2. STPA

2.1. Characteristics and Analytical Procedure of STPA

In modern systems where software plays a crucial role, accidents can occur even when no components fail because the software may function correctly but have incorrect requirements, that lead to accidents. STPA is effective for such modern systems, because the method is developed to focus on the interactions between system components. Furthermore, STPA can model various loss factors, including not only software but also human actions, making it applicable to complex systems. Additionally, STPA is a top-down method, enabling the identification and analysis of unexpected scenarios that could arise during operation at the conceptual design stage.

The analytical procedure of STPA is as follows:

1. Define “losses” that are unacceptable and enumerate the “hazards” that lead to those losses.
2. Create a “control structure diagram” that shows the interactions between system components.
3. Extract “unsafe control actions (UCAs),” which are control instructions that can lead to hazards, from the control instructions between system components.
4. Identify “loss scenarios” that could lead to UCAs and hazards.

The scenarios obtained from this analysis can be used to design appropriate operational methods to protect the system.

2.2. Application of STPA to “Hirogari”

The operational system is divided into a ground station and a satellite. To achieve the mission, a control structure diagram was created by focusing on the flow of necessary information from the ground station operator and the equipment used to achieve the mission or prevent satellite loss. This is shown in Figure 1, where the red lines and words indicate control instructions and the blue lines and words represent feedback. Incorporating the equipment used for satellite maintenance and mission execution into the control structure diagram, which was not included in previous studies (Yamada et al., 2022), allows for the identification of potential loss scenarios that may occur during satellite operation when the hazard is activated, as shown in Table 1.

Using the control structure diagram, following the STPA procedure, loss scenarios are identified related to unsafe control actions (UCAs) that could lead to satellite loss, loss of satellite status monitoring, and loss of mission. Then, countermeasures to avoid these loss scenarios should be determined. Some of the countermeasures can be captured as operational guidelines. Table 2 shows some loss scenarios and their corresponding countermeasures. However, it should be noted that for issues that cannot be addressed during operation, STPA should be applied from the conceptual design phase and a design that takes STPA into consideration should be performed.

3. Satellite model

The previous section described that the loss scenarios and countermeasures obtained by STPA can be utilized not only in design but also in operation. However, the obtained countermeasures interact with each other and with the satellite operation, situational judgments such as whether to continue the mission, abort it, or wait, depending on the health status of the CubeSat, must be made in a rapidly changing situation. Therefore, this study aim to develop an efficient operational method that avoids loss scenarios using a deep Q-network (DQN)(Mnih et al., 2015). In order to demonstrate the applicability of the reinforcement learning, this study will focus on a simple “Hirogari” mission to check communication performance.

3.1. Agent model

The agent chooses one of the three actions “Wait”, “Satellite Status Data Downlink”, and “Benchmark Downlink”. The power consumption, required voltage, and time associated with these actions are depicted in Table 3. The state of the satellite agent encompasses the following: voltage, current, whether the ”Satellite Status Data Downlink” has been executed in the current pass, remaining pass time, and mission progress rate.

3.2. Battery model

First, the State of Charge (SOC) is determined by the current integration method using the following equation.

$$\mathrm{SOC}(t) = \mathrm{SOC}_0 + \frac{1}{\mathrm{FCC}} \int_0^t I(\tau) d\tau \tag{1}$$

where SOC_0 is the initial SOC, FCC is the full charge capacity, and $I(t)$ is the current. The terminal voltage V is then computed by the following equation:

$$V(t) = f_{\mathrm{OCV}}(\mathrm{SOC}(t)) - R(\mathrm{SOC}(t), T(t))I(t) \tag{2}$$

where f_{OCV} represents the SOC-OCV characteristics, with OCV standing for Open Circuit Voltage, and R is the resistance value obtained by the current interruption method. The value of R depends on both SOC and temperature $T(t)$ (Aoki, Matsuyama, Miayata, Tsuruda, & Yamagata, 2021).

Furthermore, the current I is represented as follows.

$$I = \begin{cases} (P_{\mathrm{w}} - P \cdot \eta)/V & (V < V_{\mathrm{max}}) \\ 0 & (V \geq V_{\mathrm{max}}) \end{cases} \tag{3}$$

where P_{w} is the power consumption, P is the generated power, and η is the conversion efficiency of the solar cell. V_{max} is the voltage threshold by the shunt circuit.

Lastly, we will explain the settings used during training. For the implementation of DQN, we used Stable Baselines3 (Raffin et al., 2021). The parameters used for training are shown in Table 4. For all unspecified parameters, we used the default values provided by Stable Baselines3.

Table 1. Identified losses and hazards for the CubeSat, “Hirogari”.

Accidents	Hazards
Loss of satellite (and operation)	Loss of satellite integrity Failure to maintain the minimum power required for satellite operation Inability to measure bus status data for house keeping Inability to send/receive status data Inability to make operational and mission execution decisions based on status data
Loss of satellite missions	Inability to measure measure mission data Inability to send/receive mission data

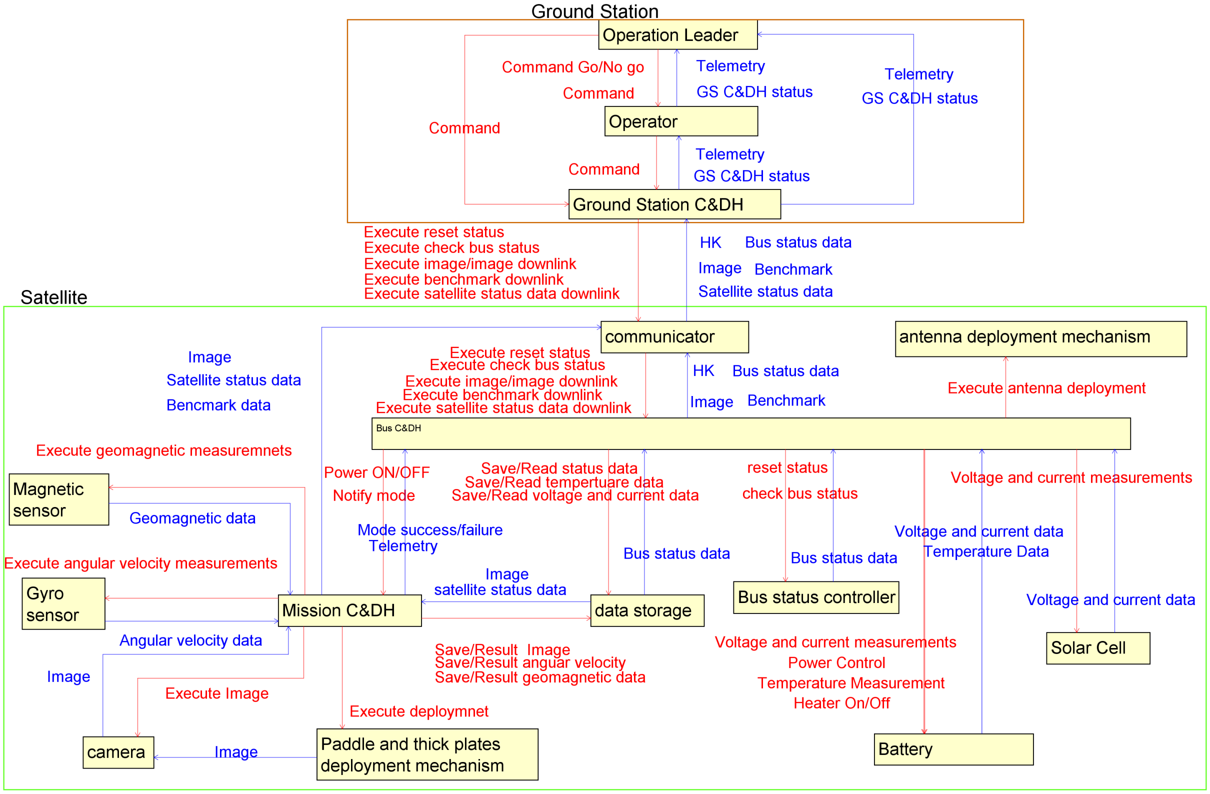


Figure 1. Control Structure Diagram for operation of “Hirogari”

4. Numerical Examples

4.1. Rewards and learning conditions

Rewards were designed based on countermeasures with IDs of 1, 2, 8, and 10 in Table 2. The rewards during mission continuation are shown in Table 5, and the rewards at the time of mission completion are shown in Table 6. The “Action Valid Status” is introduced to indicate whether the previous action was executed without the voltage dropping below the required threshold before or after the operation. Furthermore, “Whether Satellite Status Data Downlink is done” is used to denote whether the Satellite Status Data Downlink has been executed in the current path. This allows operators to verify the satellite’s status before initiating the mission. In addition, to motivate the agent to complete the mission as quickly as possible, a time penalty is introduced by subtracting 0.1 from all rewards.

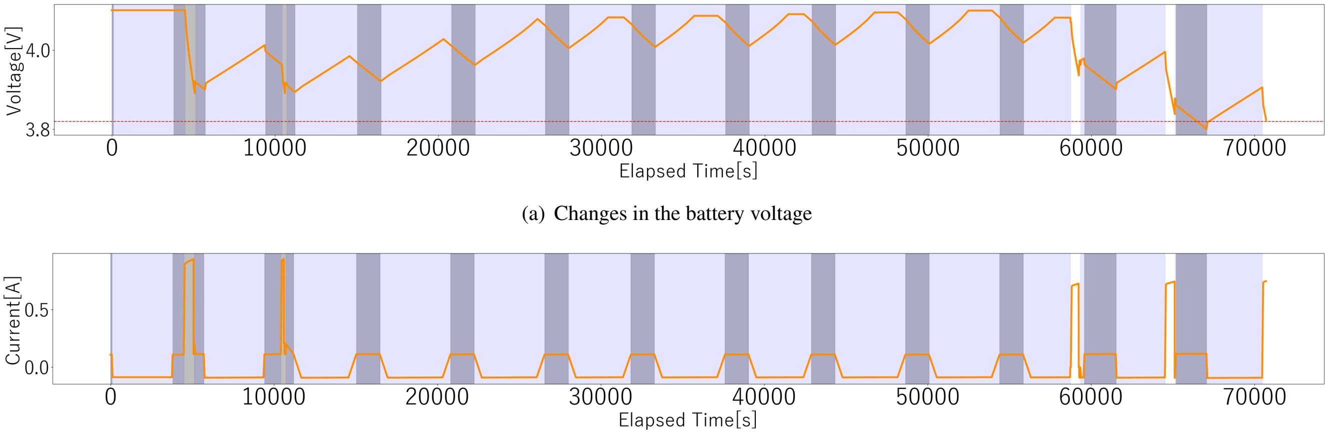
Then, the initial battery voltage was randomly selected between 3.75 and 4.1, and the temperature was set at 25°C. The termination criteria for learning were set as follows: the battery voltage falling below 3.65V, 40 hours elapsing since the start of the mission, or the “Benchmark Downlink” being conducted for a cumulative duration exceeding 1800 seconds.

4.2. Learning Result

The changes in voltage and current when acting on the learned model are shown in Figure 2. Figure 3 also shows the behavior of the satellite when it was able to communicate with the ground station. The blue areas in these figures represent the periods when communication between the satellite and the ground station is not possible, while the gray areas

Table 2. Possible countermeasures based on loss scenarios derived from STPA for “Hirogari”.

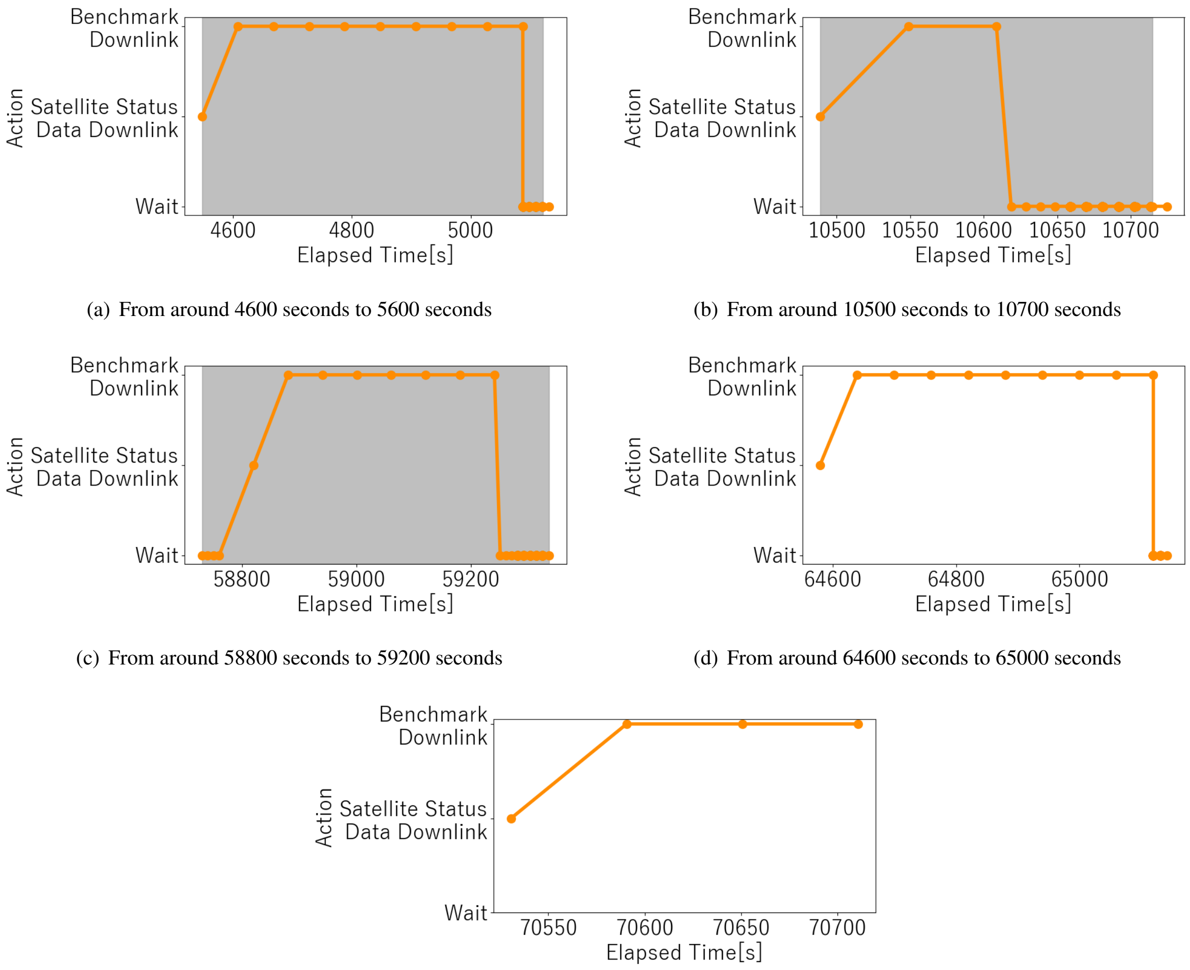
ID	Loss Scenario	Countermeasure
1	The satellite stops functioning due to low voltage, resulting in a missed mission opportunity.	Execute the function only when the battery voltage is higher than the allowable voltage for some function, V_{w}.
2	The satellite deploys paddle and thick plate despite low voltage, resulting in further low voltage and loss of the satellite.	Execute the deployment function only when the battery voltage is higher than the allowable voltage at the end of the deployment mission, $V_{\min}$.
3	The satellite deploys paddle and thick plate while battery temperature is low, resulting in voltage drop and loss of the satellite.	Execute the deployment function only when the battery temperature is higher than the allowable temperature at the end of the deployment mission, $T_{\min}$.
4	Downlink is instructed despite the antenna not being deployed, resulting in a failed downlink and a missed mission opportunity.	Issue a downlink command after the antenna deployment command. However, since it is not possible to directly confirm the success of the antenna deployment after the command, approve the antenna deployment only when downlink (FM communication) is successful. If not, reissue the antenna deployment command.
5	Image capture is instructed without deploying the paddle and thick plate. The correct image cannot be obtained, resulting in a missed mission opportunity.	Issue a paddle and thick plate deployment command before the image capturing command. However, since it is not possible to directly confirm the success of paddle and thick plate deployment after the command, confirm paddle and thick plate deployment only by capturing the image. If unsuccessful, reissue the paddle and thick plate deployment command.
6	Antenna or paddle deployment is instructed during the shadow period, which prevents the nichrome wire temperature from rising, This results in failure to cut the nylon line and deploy the paddle.	Execute deployment during the sunlit period.
7	Image capture is instructed during the sunlit period, resulting in the thick plate to be overexposed and a missed mission opportunity.	Execute image capturing during the shadow period.
8	The operator does not instruct the satellite to perform functional performance data downlink, making it impossible to monitor satellite status.	Adopt an operation policy of providing instructions on a regular basis.
9	The operator does not instruct the satellite to perform image capture and downlink, making it impossible to perform the mission.	Adopt an operation policy of actively providing instructions.
10	The operator does not instruct the satellite to perform benchmark data downlink.	Adopt an operation policy of actively providing instructions.



(a) Changes in the battery voltage

(b) Changes in the battery current

Figure 2. Changes in the battery voltage and current when acting on the learned model



(a) From around 4600 seconds to 5600 seconds

(b) From around 10500 seconds to 10700 seconds

(c) From around 58800 seconds to 59200 seconds

(d) From around 64600 seconds to 65000 seconds

(e) From around 70550 seconds to 70700 seconds

Figure 3. Actions of the satellite when it can communicate with the ground station

Table 3. Action Characteristics.

Action	**Power Consumption [W]**
Wait	0.439
Satellite Status Data Downlink	3.658
Benchmark Downlink	3.658
Action	**Required Voltage [V]**
Wait	-
Satellite Status Data Downlink	3.82
Benchmark Downlink	3.82
Action	**Time [s]**
Wait	10
Satellite Status Data Downlink	60
Benchmark Downlink	60

Table 4. Parameter settings for DQN.

Learning Rate	0.0001
Final Exploration Epsilon	0.00001
Initial Exploration Epsilon	1.0
Batch Size	128
Total Time Steps	10000000

represent the periods when the satellite is in shadow.

The figures indicate that the model has successfully learned the operational flow, first executing the “Satellite Status Data Downlink” and then executing the “Benchmark Downlink”. Moreover, the model demonstrates the ability to execute “Wait” to conserve power when the voltage starts to decrease.

The results show that the operational policy is obtained as intended by the rewards. Since the rewards are designed based on the measures required by the STPA, the loss scenario can be avoided by operating according to this operating policy.

The cumulative reward during learning is shown in Figure 4. The cumulative reward seems to converge, but it cannot

Table 5. Reward during mission continuation.

Situation		Rewards		
Action Status	**Whether Satellite Status Data Downlink is done**	**Wait**	**Satellite Status Data Downlink**	**Benchmark Downlink**
Valid	Not yet	-0.1	0.9	-1.1
	Done	-0.1	-1.1	0.9
Invalid	-	-1.1		

Table 6. Reward at mission completion.

Situation	Rewards
The battery voltage drops below the threshold.	-1.1
The predetermined time has passed.	-1.1
Mission success	0.9

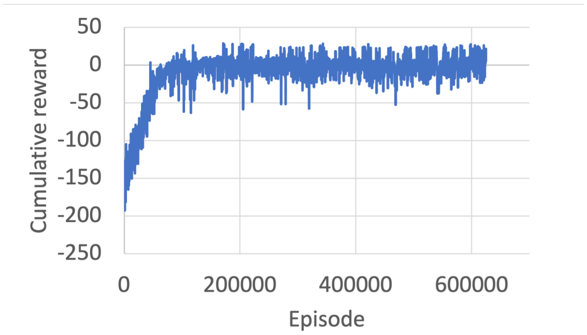


Figure 4. Cumulative reward

indicate that sufficient measures have been obtained. More detail investigation is required for the future.

5. Conclusion

In this study, a more detailed STPA is performed based on the previous study (Yamada et al., 2022) to obtain more detailed loss scenarios and determine the specific countermeasures. This allowed us to identify operational constraints that can avoid loss scenarios.

Additionally, reinforcement learning is applied to the CubeSat to achieve an efficient operational policy that follows the rewards depending on the state of the CubeSat. It is possible to automatically obtain a resilient operational policy by designing appropriate rewards.

A future challenge is to consider the battery temperature. The strategies derived using reinforcement learning in this study could be manually designed as a simple set of rules. However, introducing battery temperature into the equation complicates the voltage drop, thereby enhancing the significance of employing reinforcement learning.

Another future challenge is the application to actual satellite systems. Instead of performing the learning in real-time, we aim to consider operation-focused design and perform learning in advance. During actual operation, we plan to use the results obtained from the pre-training.

Acknowledgment

This work is partially supported by JSPS KAKENHI 22H00184. The authors would like to thank Prof. Masaki Yamagata in Kansai University for his useful comments concerning the battery model.

References

Aoki, T., Matsuyama, T., Miayata, K., Tsuruda, Y., & Yamagata, M. (2021). Internal resistance of commercial lithium-ion battery evaluated by current-rest-method and its application to on-orbit charge/discharge simulation for microsatellites, j191-13. *The JSME Annual Meeting 2021.* (in Japanese)

Hashiwaki, K., Iida, K., Kogiso, N., Nambu, Y., Higuchi, K., & Katsumata, N. (2019). Efforts for safety review of cubesat “hirogari”. *56th JSASS Kansai-Chubu Autumn Conference, A03.* (in Japanese)

Iida, K., Hashiwaki, K., Kogiso, N., Nambu, Y., Higuchi, K., & Katsumata, N. (2018). Development of cubesat ”hirogari”. *55th JSASS Kansai-Chubu Autumn Conference, B03.* (in Japanese)

Leveson, N. (2012). *Engineering a safer world.* MIT Press.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A., Veness, J., Bellemare, M., . . . Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533.

Nakase, H., Yamada, M., Maeda, Y., Nagasawa, K., Kano, T., & Kogiso, N. (2021). Lessons learned from the operation of the ultra-small satellite “hirogari”. *The 65th Space Science and Technology Conference*, 1I15. (in Japanese)

Osaka Prefecture University. (2021). *Report on nano-satellite “ hirogari ” operation.* https://www.osakafu-u.ac.jp/english-news/pr20211207e/. (Accessed on April 28, 2023)

Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., & Dormann, N. (2021). Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research, 22(268). 1-8.*

Yamada, M., et al. (2022). Resilient operation model of nanosatellite using stpa. *Aerospace Technology Japan*, *21*, 31-39. (in Japanese)