Experimental Study on the Impact of Long Communication Delays on Autonomous Decision-Making in Deep Space Habitats

Motahareh Mirfarah[a], Hyunjin Park[a,1], Zhiwei Chu[b], Sreehari Manikkan[a], Manuel Salmeron[c], Herta Montoya[d], Seungho Rhee[a], Christian E. Silva[a], Ilias Bilionis[a], Shirley J. Dyke[a,c,*]

[a]*School of Mechanical Engineering, Purdue University, 610 Purdue Mall, West Lafayette, 47906, IN, US*
[b]*School of Applied and Creative Computing, Purdue University, 610 Purdue Mall, West Lafayette, 47906, IN, US*
[c]*Lyles School of Civil and Construction Engineering, Purdue University, 610 Purdue Mall, West Lafayette, 47906, IN, US*
[d]*Department of Mechanical, Aerospace, and Industrial Engineering, The University of Texas at San Antonio, One UTSA Circle, San Antonio, 78249, TX, US*

Abstract

Space vehicles and habitats are complex systems with tightly coupled interdependencies. They operate under extreme, unforgiving conditions that challenge system performance and crew safety. Effective health management in such systems demands strong architectural and situational knowledge, leading to heavy reliance on ground control. As missions venture deeper into space, growing communication delays make such support impractical, prompting a paradigm shift toward greater on-board autonomy. However, computational constraints limit the migration of all health management components to the habitat. Running high-fidelity fault detection and diagnosis modules on the ground can conserve on-board resources but introduces delays in decision-making. This latency increases the need for effective decision-making policies, which may require interrupting agents to enable timely responses to higher-priority situations. In this paper, we study how communication delays and agent interruptibility impact decision-making in a habitat system, using the Human-centered Autonomous Resilient Space Habitat (HARSH), a reconfigurable cyber-physical testbed. A series of experiments is conducted to evaluate different design configurations under single and compound disruption scenarios. The ability of each configuration to respond and recover from faults is assessed using resilience metrics. Findings from the experiments inform recommendations to advance autonomous decision-making policies and enhance system resilience in future deep space habitats.

Keywords: deep space habitats, communication delay, decision-making policy, agent interruptibility, autonomy, system resilience, cyber-physical testing

1. Introduction

The pursuit of a sustained human presence beyond Earth is driving major efforts to establish a settlement on the Moon and Mars (Crusan et al., 2017; Kessler et al., 2022; Sun et al., 2024). Space habitats, a key element of these settlements, are intricate assemblies of interdependent subsystems where the behavior of each component can impact the behavior and performance of the entire system. In addition to their architectural and operational complexity, these systems must endure extreme and unforgiving conditions unique to space. Radiation, thermal fluctuations, seismic activity, dust contamination, and resource constraints are among the primary stressors, along with limited resupply

*Corresponding author. Donald A. and Patricia A. Coates Professor of Innovation in Mechanical Engineering and Professor of Civil and Construction Engineering, Purdue University, 610 Purdue Mall, West Lafayette, IN 47906, USA. Email: sdyke@purdue.edu

Email addresses: mmirfara@purdue.edu (Motahareh Mirfarah), park1721@purdue.edu (Hyunjin Park), chi32@purdue.edu (Zhiwei Chu), smanikka@purdue.edu (Sreehari Manikkan), salmeron@purdue.edu (Manuel Salmeron), herta.montoya@utsa.edu (Herta Montoya), rhee18@purdue.edu (Seungho Rhee), cesilva@purdue.edu (Christian E. Silva), ibilion@purdue.edu (Ilias Bilionis), sdyke@purdue.edu (Shirley J. Dyke)

[1]Present address: Assistant Professor, Department of Mechanical and Electric Automotive Engineering, Inje University, Gimhae, Gyeongnam 50834, Republic of Korea.

opportunities, inevitable equipment aging and deterioration, and unanticipated anomalies. Together, these factors present substantial challenges to ensuring crew safety and maintaining habitat functionality over long missions and extended system lifespans (Kitching et al., 2021).

In light of these challenges, resilience emerges as an overarching goal for the design and operation of future space habitats (Dyke et al., 2021; Maghareh et al., 2021). Resilience refers to the ability of a system to react to, survive, and recover from a wide range of disruptions (Uday and Marais, 2015). Achieving resilience requires continuous system monitoring, situational awareness, and timely execution of control and recovery actions. Given the complex interdependencies among habitat subsystems, delivering this capability demands substantial expertise across domains and a deep understanding of system architecture.

For example, on the International Space Station (ISS), fault detection and decision-making is largely managed by ground-based mission control centers (Dempsey, 2017). These centers are staffed by large specialized teams working around the clock. They monitor telemetry from thousands of on-board sensors, interpret trends, and respond appropriately to ensure the continued health and functionality of the station. To maintain safe and efficient operation, they provide solutions to safety-critical issues and routinely oversee the vehicle by sending hundreds of commands daily (Badger et al., 2020).

Relying on ground-based control is effective in low Earth orbit, where latency is minimal (generally less than 1 second). However, as missions extend farther from Earth, this approach becomes increasingly impractical due to limited bandwidth, potential signal blackouts, and escalating communication delays, ranging from a few seconds to the Moon to between 4 and 22 minutes one-way to Mars (Dempsey et al., 2024). Such extended delays can critically impair timely support from Earth, prompting a shift toward greater reliance upon on-board autonomy for system health management (Parisi et al., 2024; Wu et al., 2021; Pillai, 2024; Badger et al., 2019; Valinia et al., 2022).

NASA defines autonomy as "the ability of a system to achieve goals while operating independently of external control" (NASA, 2015). To compensate for the loss of constant ground support in deep space missions, this capability must expand considerably. However, several factors make on-board autonomy challenging to achieve (Valinia et al., 2022). Whether dormant or minimally staffed, the habitat requires an on-board intelligent health management system (HMS) capable of handling faults and making decisions to recover independently (Wu and Vera, 2019). Yet, limited on-board computational capacity constraints relocating all the HMS functionalities to space habitats. Space-rated processors, designed for and proven to have high reliability in harsh environments, lag behind state-of-the-art terrestrial hardware in performance (Oche et al., 2021). Moreover, strict constraints on size, mass, power, and cooling further limit the deployment of demanding analytics or even AI algorithms on-board (Lim et al., 2020; Diana and Dini, 2024).

To guide system design under such constraints, the Mars Campaign Office (MCO) of NASA's Moon to Mars program promotes the principle of *minimum Earth independence*, advocating that any functionality that can reasonably remain on Earth should do so (Maddox et al., 2024). This approach helps offload computational demands from the habitat while leveraging potential domain expert knowledge and superior computational resources on Earth. Similarly, NASA's human system research for Mars missions clarifies that full autonomy is not the objective. Instead, Earth-based systems are expected to retain as many responsibilities as possible, especially for tasks that tolerate communication delays (Abercromby et al., 2025). Accordingly, only a subset of mission control center capabilities can be transferred to deep space habitats. System designers must therefore identify which capabilities are critical to perform locally and how to distribute HMS responsibilities between ground and on-board systems, particularly in the presence of communication delays (Wu and Vera, 2020; Frank et al., 2013; Wu and Vera, 2019).

Fault detection is a key function of the HMS and is typically performed by embedded Fault Detection and Diagnosis (FDD) modules. Although simple on-board FDDs have historically been deployed in early missions, such as Sputnik and Apollo (Ben-Itzhak, 2024), the increasing complexity and interdependence of deep habitat systems demand more advanced approaches. High-fidelity FDDs, including AI-based techniques, offer improved diagnostic capabilities but are, as of now, too resource-intensive for on-board deployment (Diana and Dini, 2024; Ben-Itzhak, 2024; Meß et al., 2019). Offloading FDD computations to ground systems can help overcome these limitations, though it also introduces communication delays that may compromise timely fault response, hinder recovery, and reduce overall system resilience.

Latency in fault detection can inevitably affect downstream functions and expose the habitat to events with shorter time-to-effect. NASA's Autonomous Mission Operations (AMO) project (Frank et al., 2013) highlights the importance of integrating adaptive decision-making policies into on-board autonomy as a mitigation strategy for operations under communication delays. To support timely responses and maintain habitat resilience, one viable option is for

the autonomous HMS to interrupt the ongoing tasks of robotic agents to address higher-priority situations. Drawing on subject-matter expertise, Rollock and Klaus (2025) identified task interruption and reallocation, *e.g.*, through procedural breakpoints, as a potential yet underdeveloped mechanism to enhance the responsiveness and self-reliance of autonomous coordination in deep space habitats.

At present, there is no direct mission evidence, either within or outside of NASA, demonstrating how a small crew on an extended mission, relying on the continuous functioning of complex systems, can survive and operate without constant support from Earth (Valinia et al., 2022; Parisi et al., 2023). As such, simulation-based studies offer a viable means to verify system capabilities and lay the groundwork for developing countermeasures that mitigate associated risks. Landon et al. (2025) reviewed two decades of research on communication delays in the context of missions beyond low Earth orbit. Although their primary focus was on crew autonomy and mission control interactions, their findings remain broadly relevant to autonomous systems. Notably, they emphasized the difficulty of isolating the effects of delay from other variables and the need for more controlled studies under varied communication delay profiles and experimental settings, such as isolated habitats and laboratory simulations.

The purpose of this paper is to study how communication delay affects autonomous decision-making and system resilience in a space habitat when FDD computations are performed on the ground. Experiments are designed and conducted using the Human-centered Autonomous Resilient Space Habitat (HARSH) testbed (Silva et al., 2025) to evaluate the impact of such delays. We further explore how the outcomes vary under different decision-making policies, focusing on agent interruptibility — a design factor whose influence emerged as important during earlier experimental iterations.

HARSH is a cyber-physical testbed designed to support experimental research in autonomy, featuring a reconfigurable HMS that enables a wide range of investigations relevant to deep space habitats. We use a cyber-physical testbed as it is impossible to reproduce the Lunar or Martian environment in an experimental laboratory setup. Conversely, purely numerical simulation lacks real-world uncertainties and modeling errors, resulting in a simplification of situations. Cyber-physical testing bridges this gap by capturing emergent behavior that arises from hidden interdependencies and real-world uncertainties. As a result, HARSH provides a structured platform to evaluate system design features, including decision-making policies, and learn how to deal with the complexity of decisions relevant to future space habitats. We conduct a series of experiments involving disruption scenarios with both single and confounding faults. Each configuration is evaluated based on its ability to detect, isolate, respond to, and recover from faults. System behavior is analyzed in terms of resilience in recovery from off-nominal conditions.

The specific research questions we aim to address through these experiments include:

- How does communication delay impact habitat resilience?
- What is the added value of running FDD modules on-board?
- Does allowing HMS to interrupt the agent enhance resilience?

The paper is structured as follows: Section 2 introduces the technical approach, including the choices for setting up the testbed and the selected resilience metrics used for evaluation. Section 3 describes the specific experimental setup used for this study along with a description of a sample event sequence. Section 4 presents the results of the experiments, using paired comparisons designed to isolate the effects of communication delays and agent policies. Section 5 highlights broader insights and key lessons learned from the results. Finally, Section 6 concludes with the implications of this study on future habitat design.

2. Technical Approach

The HARSH cyber-physical testbed, shown at the center of Figure 1, is one realization of a smart space habitat (Silva et al., 2025). The space habitat is partitioned into virtual and physical components, enabling rapid reconfigurability, low-cost testing, and the ability to inject controlled uncertainties into a test. HARSH includes an HMS designed to be modular and enable integration of a variety of fault detection methods, diagnostic tools, and decision-making approaches (Chu et al., 2025). The HMS continuously monitors relevant sensor and telemetry data to infer the health status and condition of the habitat. Based on this assessment, it isolates faults from a pre-defined library, determines the appropriate response, and sends commands to control testbed components and direct the agent

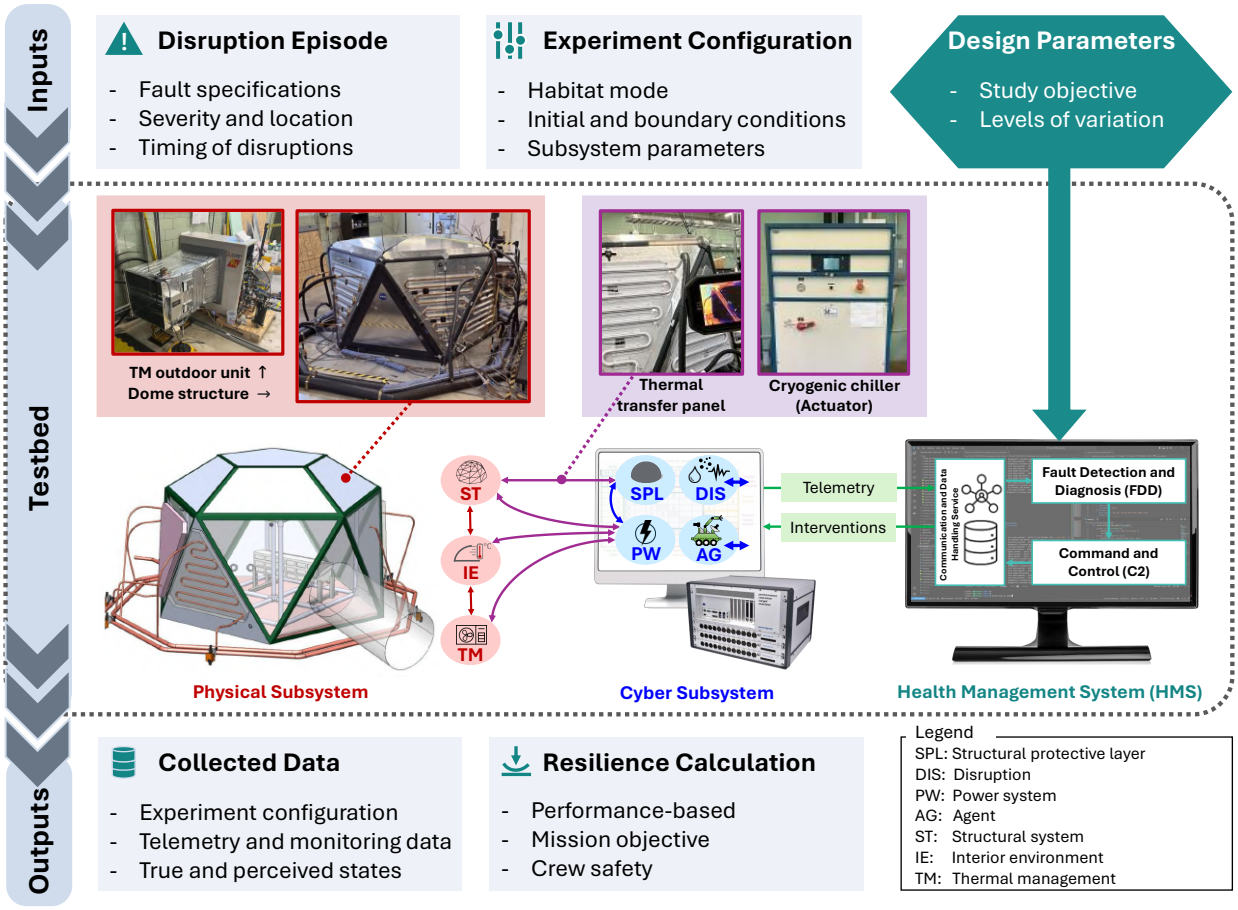


Figure 1: Experimental approach used for evaluating design features and management policies.

when intervention is required. A variety of algorithms and methods may be examined using this testbed, and new capabilities will be included moving forward.

To perform an experiment using HARSH, we follow the procedure illustrated in Figure 1. Each experiment requires three types of inputs: (i) design parameters, which define what we are interested in studying and what range to consider (*e.g.*, duration of the communication delay); (ii) habitat setup configuration, which describes the assumptions and initial conditions for a given test; and (iii) disruption episode, which specifies the sequence and details of the disruption events the habitat is subjected to. The output of each individual experiment includes: the sensor measurements (including both telemetry data and data needed to monitor and document each experiment), a log of perceived states over time and the commands sent and actions taken, and a set of resilience metrics derived from telemetry data and ground truth, as explained in Section 2.6.

The primary physical components of the habitat include a scaled structural dome, referred to as the structural system (ST); a lightly pressurized air volume, known as the interior environment (IE), which is enclosed by the ST; and a thermal management (TM) system that regulates the temperature of the air volume. Thermal boundary conditions are enforced through nine lateral heat transfer panels to emulate the effects of the exterior temperature. These panels circulate a heat transfer fluid — supplied by a cryogenic chiller — across the ST surface and are instrumented to monitor and control relevant temperatures. Thus, the ST experiences the temperature and heat flux as if there is a physical regolith layer surrounding it that is exposed to the exterior environment. A subset of the nine panels may be dynamically changed by means of a controller when it is necessary to emulate the thermal effects of a localized disruption. The specifics of this real-time thermo-mechanical cyber-physical testing method are discussed in Montoya

(2024); Montoya et al. (2025); Salmeron et al. (2024).

The cyber (numerical) components of the habitat are mainly modeled in MATLAB/Simulink (R2020b) (The MathWorks, Inc., 2020) and deployed on a Speedgoat Performance real-time target machine for execution. These components are adapted from HabSim (Azimi et al., 2025), a virtual habitat simulator framework. Among them is the robotic agent model (AG), responsible for conducting inspections and repairs using an on-board inventory of tools and consumables. The inventory model is currently assumed to be perfect, though future studies could incorporate modeling errors to reflect more realistic conditions. The numerical partition also incorporates a power system (PW), responsible for supplying energy to the habitat subsystems, including the TM, the agent, and a lumped load representing smaller power-consuming devices such as sensors. In addition, a finite element model of the Structural Protective Layer (SPL) is included, providing thermal and mechanical protection to the ST. The disruptions (DIS) are triggered through the numerical model, as described in Section 2.3. While further details and specifications on the HARSH testbed are available in Silva et al. (2025), the remainder of this section focuses on the adjustable testbed components relevant to this study.

2.1. Design Parameters

To configure an experiment, we first define the design parameters to be varied across the set of tests so that they effectively address the target questions. For instance, the location of the computational resources and details of the agent-interruptibility policy settings are defined. We determine the specific discrete levels of variation to consider. These parameter sets are then deployed to the relevant subsystems in the testbed to support the study objectives.

2.2. Experiment Configuration

The testbed flexibility allows a wide range of options to set up a scenario of interest. Accordingly, the variables to be modified and the main assumptions that influence the outcomes of the experiments conducted need to be specified to set up the experiment configuration. This configuration includes considerations such as habitat operational mode (crewed, dormant, or transitional), external conditions, time of (lunar) day, TM operating mode, IE temperature setpoint, available sources of power generation, nominal component loads, agent action preparation and actuation durations, available tools and consumables in the inventory, and recovery strategy assumptions.

Next, we establish initial and boundary conditions of each subsystem in the numerical components to reflect the habitat starting state and projected operational trajectory. These initial conditions need to be compatible with the physical components, so it is typically necessary to run the experiment for a period of time to allow the entire system to reach a steady state. By deliberately setting limited levels for parameters such as initial battery storage and the availability of the agent and inventory, we simulate scenarios involving decision-making under constrained resources. To ensure experimental consistency and repeatability, these initial conditions are prescribed uniformly across all experiments considered in this paper.

2.3. Disruption Options

We assess each design by setting the corresponding configuration in HARSH and subjecting the habitat to a complete fault-estimation-decision-recovery cycle. Each cycle is a path in the plausible sequence of events, emerging from a *disruption episode*, which is specifically designed to be informative in the exploration of the design parameters. A disruption episode is defined by the disruption type(s), its severity, location, and timing. With this testbed, we can generate numerous variations of physical and cyber disruptions within the habitat system. In this study, we target two specific faults, including the SPL damage due to the micrometeorite strike (in the numerical subsystem) and the TM refrigerant leak (which is a physical fault).

SPL damage (loss of thermal insulation) caused by the micrometeorite impact is realized as damage to the numerical part of the habitat, whose boundary condition with the habitat structure is emulated interactively using the thermal transfer system (Montoya, 2024; Montoya et al., 2025). The intensity of this damage is set to 75% reduction in SPL thickness for three out of nine lateral panels, achieved by active temperature control of those panels with the cryogenic chiller, as illustrated in Figure 1. The damaged panels are adjacent and referred to as panels 7–9. Reducing the thickness of these panels from their nominal value of 0.2 m to 0.05 m primarily increases thermal loss to the external environment. This, in turn, results in more frequent TM cycles and higher power consumption to maintain the target IE conditions.

The TM refrigerant leak is induced through a fixed opening in the refrigerant circuit, targeted to maintain a relatively constant leak rate of approximately 0.05 g/sec. This rate is determined based on separate test results to ensure a gradual and monotonic change in refrigerant properties, avoiding abrupt disturbances or any mechanical malfunctions in TM operation. The leaked refrigerant is captured by a secondary loop equipped with a recovery pump and storage tank, effectively isolating it from the primary vapor compression cycle (Silva et al., 2025). The system initially contains a full refrigerant charge of approximately 1560 g. As the refrigerant charge gradually declines, the TM response curve shifts, altering the temperature, pressure, and power consumption profiles.

Disruption episodes are defined by either a single fault or a compound scenario combining both faults, while each is introduced at a predetermined time in this particular study to allow direct comparisons to be made between different experiments. To ensure a fair and consistent comparison across designs, each disruption episode follows a deterministic fault schedule under identical habitat initial conditions and experiment configurations.

2.4. Detection, Isolation and Decisions

The fault detection, isolation, and decision-making capacity is enabled through the developed autonomous HMS, as depicted in Figure 1. The modular architecture incorporates an embedded communication network emulator and a data handling service used for communication between the habitat subsystems, including the HMS (Chu et al., 2025). The communication network also emulates the transmission of data to ground control, on Earth, and imposes the associated communication delays when they are included for a particular experiment.

Three FDD modules are used in this study: one for detecting a leak in the TM system and two for detecting damage to the SPL layer. Each FDD module monitors selected telemetry data from the habitat and outputs whether the system is healthy or faulty. This binary status is then sent to the Command and Control (C2) component of the HMS, as shown in Figure 1. The FDD modules offer the capability to either run on-board the habitat or on-ground, meaning both the telemetry data used by FDDs and their status received by C2 are delayed. For different fault detection purposes, a variety of algorithms, ranging from the extended Kalman filter (Mirfarah et al., 2025) to information field theory (Hao and Bilionis, 2024) are developed and integrated into the HMS. A description of the three relevant FDD algorithms employed here is summarized in Table 1. The choice of FDD algorithms is not the focus of the study; rather, they serve to enable the execution of the experiments and the study to be carried out.

We include two FDD algorithms for detecting damage to the SPL layer: one operates when the TM is inactive (TM off), and the other when it is active (TM on). The SPL (TM off) FDD detects SPL damage using a lumped-parameter resistance-capacitance (RC) network model that captures the thermal behavior of the interior air volume (Bergman et al., 2011). This algorithm processes temperature measurements from the interior environment and applies Bayesian inference to estimate the probability of damage. The SPL (TM on) FDD activates when the TM is operating, which is the case in the default condition of the habitat. This FDD tracks average TM power consumption and triggers a detection alarm if consumption exceeds a preset threshold. The TM Leak (TM on) FDD estimates refrigerant loss via a Bayesian switching model based on the inlet–outlet temperature differential of the TM indoor unit. When the posterior probability indicates that the exponential decay model explains the data better than the steady-state model, and the residual passes a certain threshold, a leak detection is reported.

Table 1: Specifications of the employed FDD algorithms.

FDD Purpose	**SPL Damage**	**SPL Damage**	**TM Leak**
Operating Condition	TM off	TM on	TM on
Observation Data	IE temperature	TM power consumption	TM indoor unit inlet & outlet temperatures
Algorithm	Physics-based	Statistics-based	Statistics & Physics-based
UQ* Capability	Yes	No	Yes
Observation Batch [min]	40	15	10
Computational Time [min]	3-7	<1	2-3
Inference Frequency [min]	12.5	3.0	5.0

* UQ: Uncertainty Quantification.

The C2 module receives all of the FDD outputs, agent-inspection results, and resource status (*e.g.*, available energy, inventory) to perceive the health state of the habitat and make recovery decisions. The C2 logic relies on a causal model of habitat functionality (Murali Krishnan et al., 2025; Rajasekharan Pillai and Bilionis, 2025) and an associated diagnostic reasoner (DR). The DR systematically isolates faults using a dependency matrix (D-matrix), mapping built-in test outcomes (FDDs and inspections) to potential faults (Aaseng, 2015; Aaseng et al., 2023).

The DR classifies the status of each fault into one of four categories: good, unknown, suspect, or bad. A fault is marked unknown when no relevant test output is available. If a FDD module runs but does not indicate a fault, the DR labels the fault as good. When an FDD test reports a fault, the DR suspects all implicated faults. If this fault is subsequently confirmed through the agent inspection, the DR isolates the fault and classifies it as bad.

Given the set of potential faults $X_{1:N}$ in the system, the C2 system applies the following heuristic decision policy:

- Standby: If all faults $X_{1:N}$ are either in good or unknown states.
- Inspect X: If fault X is in suspect state and inspection resources are available.
- Repair X: If fault X is in bad state and repair resources are available.
- Turn off the TM: If the available energy is constrained or a leak is confirmed (bad state), and it is safe to do so.
- Turn on the TM: If available energy is not constrained and no leak is present (good state).
- Alarm: If no safe actions remain.

Within the scope of the disruptions being investigated in this paper — SPL damage and TM leak — recovery actions include toggling the TM system (on/off) and dispatching the agent for the inspection or repair of each component. The decision policy is fully configurable and easily replaceable. Here, we introduce an alternative decision-making policy that prioritizes TM leaks, interrupting and rerouting the agent to address this critical fault first. Further details are provided in Section 3.1.

2.5. Instrumentation and Monitoring

The testbed is equipped with a total of 120 sensors, including thermocouples, pressure transducers, mass flow meters, and power meters. Among them, 21 are used for FDD purposes and are sent to the HMS. The remaining sensors are used for executing, monitoring, and analyzing the experiment, including 26 that are required to enforce realistic cyber-physical interactions. The data are acquired at a sampling rate of 1 Hz and transmitted to the HMS computer via Application Programming Interfaces (API) implemented using User Datagram Protocol. They are then distributed to the relevant modules through the embedded communication network emulator and data handling service discussed in Chu et al. (2025). The HMS also logs each FDD result, the isolated faults identified by the DR, and the corresponding decisions, ensuring comprehensive documentation and supporting result explainability. Further details on the data acquisition system can be found in Silva et al. (2025).

2.6. Resilience Metrics

Finally, resilience metrics provide a quantitative evaluation of the habitat design configurations. No single metric is entirely adequate on its own. To ensure a meaningful comparison of resilience across the experiments conducted, we employ three widely-used resilience metrics from the literature (Jain et al., 2024; Cheng et al., 2022; Dehghani et al., 2021), as summarized in Table 2. Each metric is computed based on a system performance curve $P(t)$, which represents performance as a function of time, as illustrated in Figure 2. Relevant parameters include the initial performance (P_i), recovered performance (P_r), and recovery time ($t_r - t_i$) as indicated.

The HRM (Height Ratio Metric) evaluates the ability of a system to recover by comparing the recovered performance level (P_r) to the initial performance level (P_i), capturing how much of the original functionality is regained after the disruption. The DMN (Duration Metric Normalized) quantifies the speed of recovery by finding the inverse of the time the system takes to recover ($t_r - t_i$); a shorter recovery time results in a higher resilience score. Finally, the ABM (Area-Based Metric) measures the total retained system performance during the disruption and recovery phases. It is calculated by integrating the normalized actual performance over a fixed evaluation time window $[t_i, T_{max}]$, where

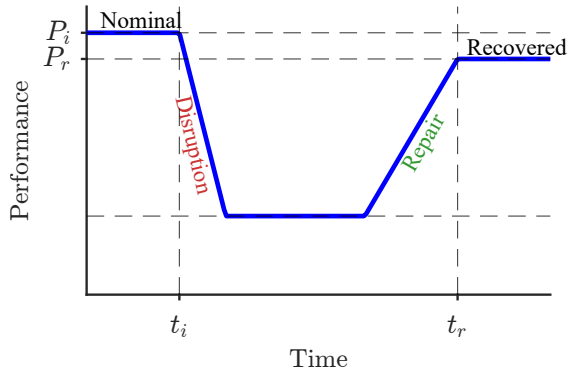


Figure 2: General performance curve for illustrating resilience.

T_{max} represents the maximum recovery time observed across all experimental scenarios to ensure comparability. A higher ABM value indicates greater retained performance and higher resilience. Together, these metrics offer complementary perspectives on resilience, focusing on the extent, speed, and cumulative impact of recovery. For presentation of the results, HRM and ABM are reported as percentages, while DMN is expressed in $10^{-2}/min$ units.

Table 2: Resilience metrics adopted in this study.

Resilience Metric	**HRM**	**DMN**	**ABM**
Definition	P_r/P_i	$1/(t_r - t_i)$	$1/(T_{max} - t_i) \int_{t_i}^{T_{max}} (P(t)/P_i)\, dt$
Range	$[0, \infty)$	$(0, \infty)$	$[0, \infty)$

We utilize both component-level and system-level performance curves. At the component level, the performance curves are based on the actual health states of damageable components observed in the episodes. Specifically, we derive separate curves for the SPL thickness and the TM refrigerant level. At the system level, the performance curve representing the overall functionality of the habitat is defined by the mission objectives: ensuring crew safety and enabling scientific research. To capture this, we consider two key parameters — interior temperature and power demand — as indicators of crew thermal safety and research capacity, respectively.

Let $\mathcal{T}$ and $\mathcal{R}$ represent the sets of N_T possible thermal safety and N_R possible research ability states

$$\mathcal{T} = \{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_{N_T}\}, \quad \mathcal{R} = \{\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_{N_R}\}. \tag{1}$$

At each time t, $\mathcal{T}_i \times \mathcal{R}_j \in \mathcal{T} \times \mathcal{R}$ is a possible set of states in which the system may be. We assign a performance value V_{ij} to each possible set of states, based on how favorable the condition is. The system-level value per unit time at time t for a system with given state of (T_t, R_t) is then calculated by:

$$P(t) = \sum_{i=1}^{N_T} \sum_{j=1}^{N_R} \mathbb{1}_{\mathcal{T}_i \times \mathcal{R}_j}(T_t, R_t) \cdot V_{ij}, \tag{2}$$

where $\mathbb{1}_{\mathcal{T}_i \times \mathcal{R}_j}(T_t, R_t)$ is the indicator function defined as

$$\mathbb{1}_{\mathcal{T}_i \times \mathcal{R}_j}(T_t, R_t) = \begin{cases} 1, & \text{if } (T_t, R_t) \in \mathcal{T}_i \times \mathcal{R}_j, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Based on the range of variability exhibited across the experimental configurations, the interior environment temperature and power demand are discretized into intervals representing safe, hazard, and accident states. Each combination of these state intervals defines an entry in the performance matrix $V \in \mathbb{R}^{N_T \times N_R}$, presented in Table 3, where higher values indicate more desirable operating conditions. The nominal initial condition used in the experiments corresponds to $V_{31} = 1.00$. The baseline region corresponding to temperature level $\mathcal{T}_3$ and power demand level $\mathcal{R}_1$

represents the nominal operating envelope observed during habitat experiments. Conversely, the extreme operating regions correspond to the worst plausible conditions observed during disruption experiments. Only the intermediate transition boundaries (between $\mathcal{T}_1$-$\mathcal{T}_2$ and $\mathcal{R}_2$-$\mathcal{R}_3$) permit some flexibility in threshold selection. The sensitivity of system-level resilience outcomes to these threshold definitions and performance value assignments is examined in Appendix A.

Table 3: Performance value matrix V for possible combinations of state sets.

	Power Demand (kW)		
Temperature (°C)	$\mathcal{R}_1 = [0, 0.9)$	$\mathcal{R}_2 = [0.9, 1.44)$	$\mathcal{R}_3 = [1.44, 3.6]$
$\mathcal{T}_1 = [14.0, 21.5)$	0.30	0.20	0.10
$\mathcal{T}_2 = [21.5, 24.1)$	0.80	0.60	0.50
$\mathcal{T}_3 = [24.1, 27.1)$	1.00	0.85	0.70
$\mathcal{T}_4 = [27.1, 30.0]$	0.80	0.60	0.50

3. Experimental Setup

In this section, we walk through the procedure described in the previous section to configure the specific design parameters, experiment parameterization, and disruption episodes considered in these experiments. We then clarify the event sequence through a sample case of the SPL damage episode.

3.1. Design Parameters and Disruption Episodes

Experiment design parameters are selected to explicitly introduce the variations that are of interest to meet the objectives of a study. This particular investigation examines the effect of significant communication delays on the ability of the habitat to diagnose and recover. The FDD algorithms may be executed on-board the habitat, or at mission control, designated here as on-ground. Consistent FDD implementations are used across configurations so that observed differences primarily reflect communication latency rather than variations in FDD capability or FDD computational complexity. Regardless of the FDD location, all decision-making by HMS occurs on-board via a non-blocking, asynchronous architecture that remains active even when the robotic recovery agent is occupied. Decisions result in commands sent to the robotic agent, and the habitat manager may activate the policy to interrupt the agent and reassign it to a new task based on priority. Accordingly, the two design parameters are designated here as "location of FDD" and "agent interruptibility." In Table 4, we specify the possible options for each design parameter as the level. Both of these design parameters have two mutually exclusive options to consider.

Table 4: Design options considered in this study.

Parameter		**Agent Interruptibility**	
	Level (Index)	Activated (A)	Deactivated (D)
FDD Location	On-Board (1)	(1, A)	(1, D)
	On-Ground (2)	(2, A)	(2, D)

When the FDD modules execute on-board the habitat, as in designs (1,A) and (1,D), there is no communication delay. However, this setup incurs a computational cost on the habitat energy storage. In contrast, executing the FDD modules on-ground requires no additional on-board power; rather, it introduces a two-way communication delay that depends on the assumed distance between the habitat and ground control. Accordingly, executing the FDDs on the ground can be implemented with any reasonable communication delay. In this study, we focus on a few representative values to examine Mars-scale communication under ideal conditions with no data loss. Specifically, we fix the round-trip delay at 10, 20, or 30 minutes in the on-ground designs (2,A) and (2,D), respectively. This delay (defined as twice the one-way latency) is treated as a constant, since the communication latency does not vary significantly over the duration of the investigated events. These delay values are significant enough to influence the recovery and thus allow us to evaluate the implications of communication delay on the system performance and resilience.

The agent interruptibility feature addresses the varying urgency of responding to different faults. In this context, interruptibility refers to the allocation of tasks to the robotic agent responsible for performing maintenance and repair actions. The interruptibility policy is intentionally introduced as a simple priority-based preemptive rule to allow the effect of communication delay and task interruption to be examined in isolation. This formulation helps maintain a consistent basis for interpretation across experimental configurations. It should therefore be interpreted as an experimentally controlled intervention rather than a claim of optimal autonomous scheduling. A refrigerant leak in the TM system is generally perceived as posing a greater risk than a loss of insulation in the SPL. Therefore, recovery actions related to TM failures are prioritized. This task prioritization is grounded in criticality classification practices used in NASA spaceflight and support systems. NASA guidance defines five levels of functional criticality, categorizing failures according to consequence severity, ranging from inconsequential effects to loss of mission, loss of vehicle, or loss of crew (NASA Goddard Space Flight Center, 2024). Under this classification, the criticality of a function is determined by the consequences associated with its loss. Failures affecting the active thermal control system, including refrigerant leaks, are typically designated as critical faults due to their vital role in mission continuity and crew safety (Rollock and Klaus, 2026).

When interruptibility is enabled, as in designs (1,A) and (2,A), the agent can be redirected from a lower-priority task to a higher-priority task if such a condition arises. When this happens, the ongoing agent task is interrupted, the agent is redirected to another task, and the unfinished task resumes after the higher-priority task is completed. Conversely, when this feature is disabled, tasks are executed strictly in the order they are assigned. This feature is most relevant in compound fault scenarios, where multiple faults occur in close succession.

To formalize this policy, we define a priority function $\pi : \mathcal{F} \to \mathbb{R}$ over fault modes $\mathcal{F} = \{\text{SPL}, \text{TM}\}$ such that $\pi(\text{TM}) > \pi(\text{SPL})$. Let $a_c(t)$ denote the current agent task at time t and $a_n(t)$ a newly issued task associated with a fault mode through a mapping $g : \mathcal{A} \to \mathcal{F}$. When interruptibility is activated, the C2 module applies the following preemptive priority rule:

$$\text{Policy}(a_n(t), a_c(t)) = \begin{cases} \text{Interrupt}, & \text{if } \pi(g(a_n(t))) > \pi(g(a_c(t))), \\ \text{Queue}, & \text{otherwise.} \end{cases} \tag{4}$$

The habitat initial condition and configuration for the experiments are as follows: (i) the exterior environmental condition is −269 °C (lunar night), and the interior environment is regulated with the TM in heating mode using a set point of 22 °C; (ii) a single robotic agent is available on-board to perform inspection and repair actions and is in standby mode by default; (iii) no source of power generation is available, so the habitat relies on its battery storage under a resource-constrained operating condition; and, (iv) the habitat initially has sufficient energy storage (70 kWh) and a full inventory of required tools and consumables for the agent to perform actions. This storage level sustains the 10-hour experimental duration under nominal conditions and corresponds to approximately 4 days of operation at the baseline power demand, representing a plausible emergency battery-only configuration. With this initial state, decisions are not affected by resource limitations such as energy and inventory items, ensuring that any commanded recovery actions may be completed. Note that if we had allowed this range of options in the decision-making, the decisions would have relied much more on the capacity to predict future trajectories and the depth of the resource models. As a result, the complexity of the problem would grow substantially.

The recovery phase requires agent actions to address suspected or confirmed system faults. Given the relevance of agent interruptibility policy to this study, Table 5 outlines the key attributes of agent actions, including the time and power required to carry them out. These values are deterministic in this study; however, we can readily incorporate severity-dependent values or even random variables in scenarios where such variations are of interest. Each action has a preparation period before actuation begins, and then an actuation time to perform the action. Among the various actions included, inspections require less time and energy as compared to repairs. Additionally, SPL-related actions are assumed to be more demanding than those associated with TM leaks due to additional maneuvering required to acquire sufficient diagnostic data. Repairing SPL damage involves sintering regolith (Choi et al., 2020), a process that demands a substantial amount of energy.

Focusing on the SPL damage and TM refrigerant leak disruptions, we define three episodes outlined in Table 6. Episode 0 represents the nominal condition, which serves as a reference. Episode 1 involves a single fault, SPL damage, which serves as our baseline disruption due to its cascading effects on multiple habitat subsystems. Episode 2 is a compound fault scenario, building upon Episode 1, where SPL damage is accompanied by a TM refrigerant leak

Table 5: Specifications of the agent actions.

Action	Preparation Duration [min]	Actuation Duration [min]	Power Demand [kW]
Inspect SPL	5.0	7.5	0.050
Repair SPL	6.0	20.0	1.507
Inspect TM Leak	1.0	3.0	0.07
Repair TM Leak	3.0	4.0	0.60

as a secondary effect of a micrometeorite strike. This episode necessitates the agent interruptibility feature to play a role in managing confounding failure modes. We recognized this feature as essential to the safe execution of our earlier experiments, which motivated us to include it as a design option in this study.

A separate TM-leak-only episode is excluded because, with the SPL intact, the change in thermal load remains negligible, and the leak would likely go undetected. In these experiments, the SPL damage is initiated at 60 minutes, followed by the TM leak at 90 minutes in compound episodes. The selected values support a meaningful investigation of both the agent interruption feature and communication delay effects. Fault injections follow a fixed timing schedule to ensure consistent comparison across configurations and allow direct interpretation of delay and interruptibility effects.

Given the above episode descriptions, each experiment follows a 10-hour protocol comprising initialization, fault injection, detection and isolation, repair, and recovery to nominal conditions. Trials are conducted under the design configurations in Table 4, with constant two-way communication delay samples of 10, 20, and 30 minutes applied to the on-ground FDD cases (2,A) and (2,D). Selected compound fault episodes, particularly those with the longest communication delay in the on-ground configuration (20 minutes) and the corresponding on-board configuration (0 minutes), are repeated to account for inherent experimental variability. These experiments are intended to reflect representative system behavior rather than to serve as a large-sample statistical study. The 30 minutes delay case is included as a single realization due to experimental limitations. Table 6 summarizes the 15 experiments conducted.

Table 6: Complete list of experiments conducted for this study.

Design Index	FDD Location	Agent Interruptibility	#Repetition
Episode 0: Nominal			
(1,A)	On-Board (0 min delay)	Activated	1
Episode 1: SPL Damage			
(1,A)	On-Board (0 min delay)	Activated	1
(2,A)	On-Ground (10 min delay)	Activated	1
(2,A)	On-Ground (20 min delay)	Activated	1
Episode 2: SPL Damage and TM Leak			
(1,A)	On-Board (0 min delay)	Activated	2
(2,A)	On-Ground (10 min delay)	Activated	1
(2,A)	On-Ground (20 min delay)	Activated	2
(1,D)	On-Board (0 min delay)	Deactivated	2
(2,D)	On-Ground (10 min delay)	Deactivated	1
(2,D)	On-Ground (20 min delay)	Deactivated	2
(2,D)	On-Ground (30 min delay)	Deactivated	1

Because the repeated experiments are limited in number, we complement the descriptive resilience comparisons with Bayesian hypothesis testing to quantify epistemic uncertainty in a small-sample setting. For each episode and resilience metric, repeated observations across delay settings are modeled using a Gaussian likelihood with a weak proper conjugate prior, and posterior distributions of the pairwise delay contrasts are computed analytically. In this Bayesian analysis, the pairwise delay contrasts are the primary quantities of interest, as they directly capture the effect

of communication delay. The corresponding posterior summaries are reported in Section 5, while additional details regarding the prior specification and full posterior results are provided in Appendix B.

3.2. Sample Illustration of Sequence of Events in Episode 1: SPL Damage

Next, we walk through an illustrative trajectory from the baseline episode to explain the sequence of events following the single SPL damage fault. Figure 3 illustrates how the states of the habitat evolve throughout this episode.

Each snapshot highlights a key event that alters the habitat state, with a legend included in the bottom right. For each event, we display the corresponding time, an annotation describing the event, a conceptual image of the habitat true state at that moment, and a frame representing the habitat state as perceived by C2. The three icons at the bottom of the perceived states indicate the interior set-point temperature and resource availability (inventory and stored energy) at that instant. Finally, the decision made by C2 is shown, which in turn influences the subsequent sequence of events. Event numbers and times correspond to the experiment conducted under design configuration (1,A), with no communication delay and agent interruptibility activated. The actual experimental data of this case is further discussed in Section 4.1.

The perceived state of the habitat is informed by: general information (*e.g.*, setpoints, TM system status), telemetry data (*e.g.*, agent availability), FDD inferences, DR outcomes for fault isolation, and the current availability of resources such as energy and inventory. The logic embedded in the HMS, described in Section 2.4, enables this awareness of the habitat state. While many trajectories are possible, this figure explains one realization in which sufficient stored energy and inventory are available (set in the initial configuration). Although the initial configuration provides sufficient resources, C2 performs ongoing evaluations during the experiment, as indicated by the consistent icons beneath the perceived states.

The experiment begins with a crewed habitat operating under nominal conditions during the lunar night. With the FDDs on-board, this experiment involves no communication delay. The TM system is active, and C2 perceives that all components are in good condition. The decision is standby. At event 1, a micrometeorite hits the habitat, damaging the SPL. The TM cycles increase in frequency and consume more power to compensate for the loss of insulation, as shown in event 2. After some time, the corresponding FDD module uses the telemetry data and detects the SPL damage, and at event 3, this fault is immediately reported to C2 with no communication delay. Since the habitat resources are found to be sufficient, C2 dispatches the agent to inspect the habitat protective panels.

After a period of preparation, at event 4, the agent inspects the panels and informs C2 of the damage to panels 7–9. With the SPL fault isolated, C2 commands the agent to initiate repair of the identified damaged panels. At event 5, the agent collects the regolith and begins repairing the damaged SPL. Upon completing the repair at event 6, the habitat stabilizes and returns to an overall healthy condition by event 7. The episode concludes with continued monitoring until the 600 minutes mark to confirm recovery and to enable comparisons across all experiments conducted. While the experimental data for this illustration is shown and discussed in the next section, the corresponding event sequence diagrams for Episode 2 are included in Appendix D, following the same format.

4. Experimental Results

In this section, we discuss the experimental results, examining Episodes 1 and 2 separately. For each episode, we begin by summarizing the key events that occurred during the test. We then systematically analyze how different design features — communication delay and agent interruptibility — influence the outcomes. This side-by-side analysis yields insight into how specific design configurations impact habitat performance and safety. Quantitative results are used to compare all cases collectively in Section 5.

4.1. Episode 1: SPL Damage

In Episode 1 (SPL damage), the presence of a single fault renders the agent interruptibility feature irrelevant, making communication delay the sole variable, as noted in Table 6. Three experiments are performed in Episode 1 with communication delays of 0, 10, and 20 minutes. For this discussion, we focus on the extreme cases, 0 and 20 minutes, corresponding to the plots shown in Figure 4 (a-d).

Figure 4 (b) displays the true health states of the habitat system, reflected by the severity of the imposed damage. This information is hidden from the HMS and is used only to analyze the results. Figure 4 (c) presents a selection

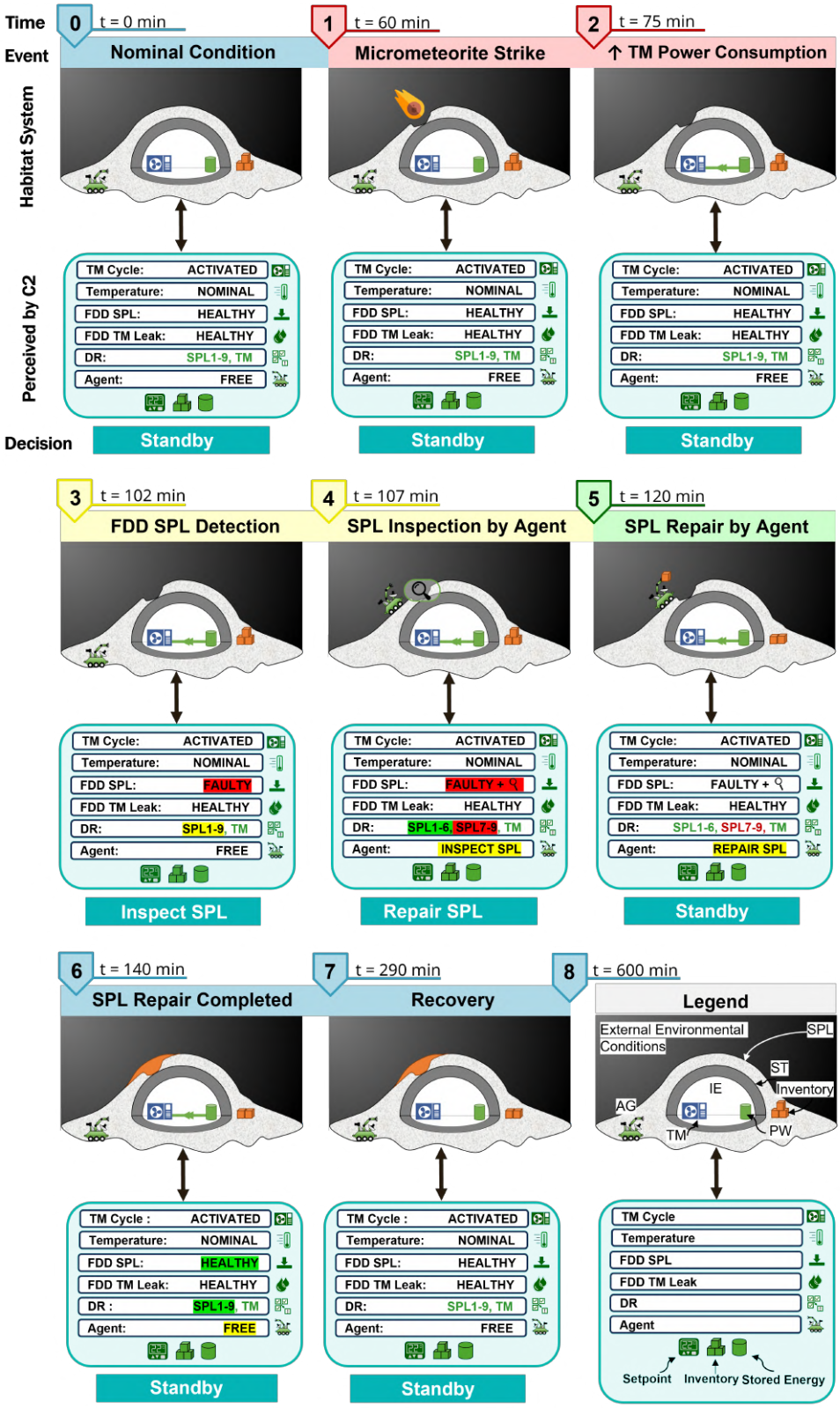
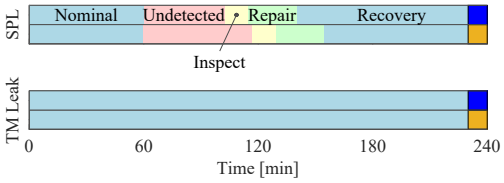
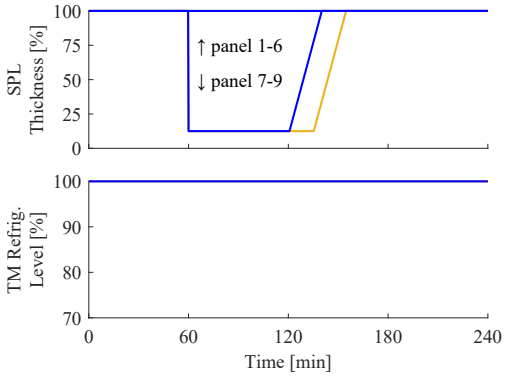


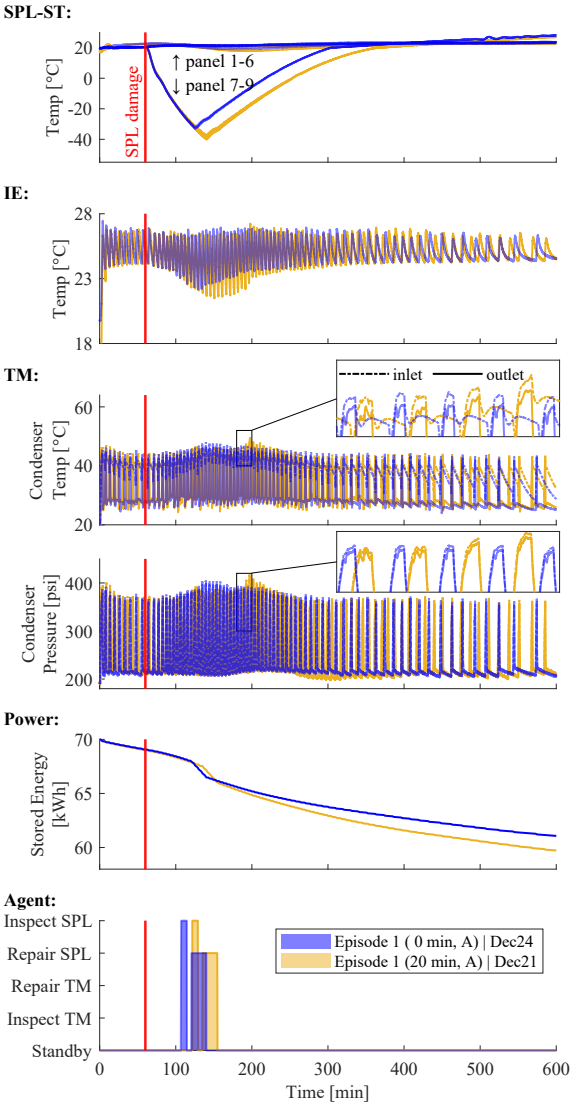
Figure 3: Sample illustration of sequence of events in Episode 1 (SPL damage) corresponding to (0 min, A).



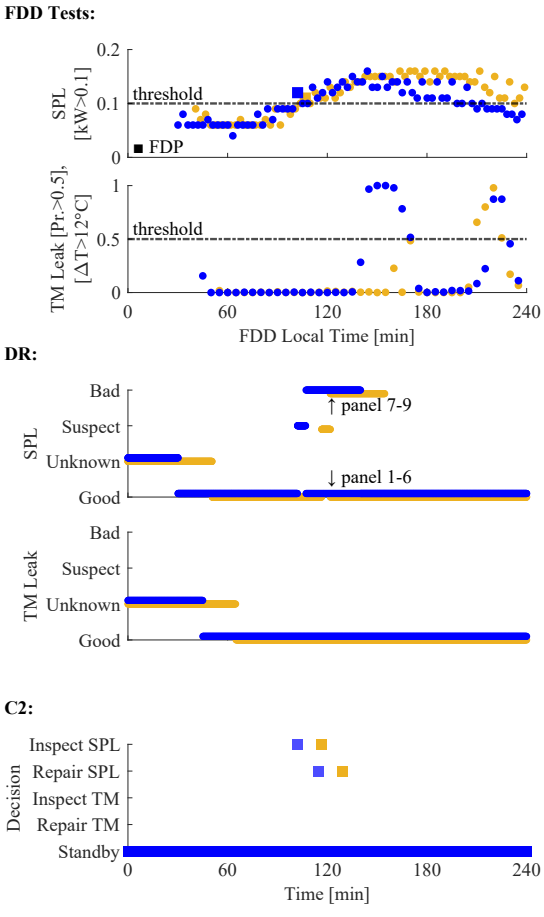
(a) Timeline showing disruptions status



(b) True health states



(c) Sensor measurements from habitat



(d) Perceived health states

Figure 4: Episode 1 (SPL damage): visualization of how communication delay affects perceived system health, decision-making, and recovery.

of sensor measurements, the observable information about the habitat. These measurements include temperatures at panels 1–9 along the SPL-ST interface, the IE temperature, and the refrigerant temperatures and pressures at the inlet and outlet of the TM condenser (indoor unit). Telemetry acquired from the virtual components includes stored energy level and agent status. Note that the time shown in all the plots corresponds to the local time on the habitat, unless otherwise specified.

Figure 4 (d) shows the states of the habitat as perceived by the HMS. These health states include the FDD outputs for both SPL and TM faults, timestamped based on their computation location. The First Detection Point (FDP) indicates the first fault detection reported by the FDD. This value corresponds to the detection time at the ground computation location, including the time required to accumulate the telemetry batch, the communication latency associated with telemetry transmission, and the inference time of the FDD module. The effective detection delay at the habitat is then obtained by adding the one-way communication latency required to transmit the detection result back to the habitat. As described in Section 2.4, the TM leak FDD reports a fault when the data better fits the degraded model and the differential temperature exceeds a threshold. For visualization purposes, only the former criterion is shown in the plot, while the corresponding FDP reflects both conditions. The health states also include the DR output as we isolate the fault based on the corresponding tests. Note that some time is required at the beginning of the experiment to collect enough data for reliably estimating the habitat state. Based on the available telemetry data and this perception of the states, C2 issues a decision at each time step, as illustrated in the figure, triggering a series of commands to facilitate habitat recovery.

Figure 4 (a) presents a timeline summarizing all phases of Episode 1, as laid out in the sequence of events described in Section 3.2. Based on the telemetry data, both experiments in this episode exhibit a similar pattern during the initialization and fault injection phases. In these two experiments, the FDD detects the SPL fault at local times of 102.1 and 96.7 minutes, corresponding to the FDD tests FDPs shown in Figure 4 (d), for the on-board (0 min, D) and on-ground (20 min, D) designs, respectively. Although the on-ground detection occurs slightly faster, the 20 minute communication delay results in a longer effective detection time (102.1 minutes versus 116.7 minutes). Consequently, the latency is effectively increasing the detection time. If the additional computational resources available on the ground do not significantly enhance fault detection (assuming comparable levels of information and sufficient bandwidth to transmit the telemetry data), communication delays will negatively affect the habitat performance and resilience. For the SPL damage scenario considered here, using identical FDD algorithms, the observed difference between raw FDD detection times due to experimental uncertainties is 5.4 minutes, substantially less than the impact of the communication delay itself.

A direct consequence of the delay in communication with FDD is that the inspection and repair actions for the SPL occur later in the experiment. This delay in addressing the fault, as reflected by the agent status responsible for recovery actions, leads to a deteriorated habitat condition. For instance, as shown in the IE and power plots in Figure 4 (c), the internal temperature drops further due to the prolonged damage, and more energy is consumed. In this case, the state of the habitat is considerably farther from nominal when the C2 sends a command to take action.

The resilience of the damaged component (SPL) and the overall habitat performance are quantified using the resilience metrics DMN, HRM, and ABM introduced in Section 2.6. The component-level performance curves are directly accessible from the true states. The system-level performance curve — capturing crew safety and research capability — is calculated using temperature and power demand levels defined in Table 3. Figure 5 presents this curve along with the corresponding system states relative to these defined levels. The plot indicates that communication delay lowers the overall system-level value throughout the recovery phase. The dependence of the observed system-level resilience trends on the adopted performance-level discretization and assigned performance values is examined through sensitivity analysis. Consistent resilience behavior is observed across these variations, as detailed in Appendix A. The quantitative implications are further discussed in Section 5.

4.2. Episode 2: SPL Damage and TM Leak

Episode 2 involves a compound fault scenario, with initial SPL damage followed by a TM refrigerant leak. Here we examine the effects of two design parameters: communication delay and agent interruptibility. To isolate and illustrate the individual impact of each parameter, we discuss illustrative pairs of experiments with similar FDD detection times while keeping the other design parameter constant. These paired comparisons are structured to distinguish resilience degradation associated with communication delay from that associated with the interruption policy by varying

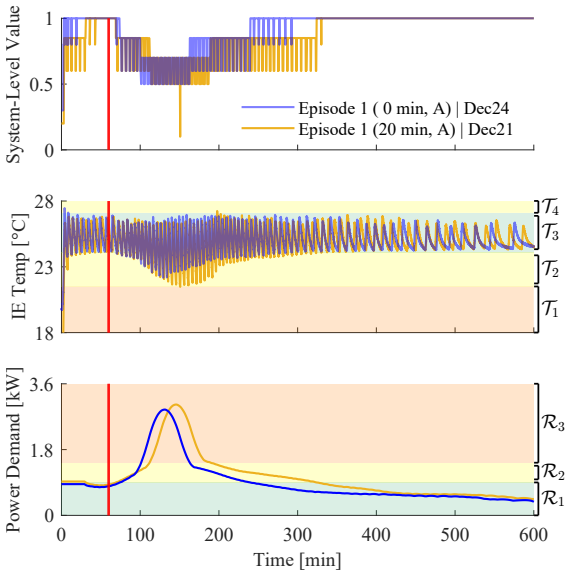


Figure 5: Episode 1 (SPL damage): determination of system-level performance curve based on the sensor measurements from the habitat.

one design parameter at a time while holding the other fixed. Detailed comparisons for two such pairs are provided in Subsections 4.2.1 and 4.2.2.

The timing of fault events in these scenarios is deliberately configured to potentially induce conflicts in the HMS decision-making, thereby highlighting the relevance of agent interruptibility. Accordingly, across all experiments, the SPL damage FDD consistently detects the fault first, followed by the TM leak FDD.

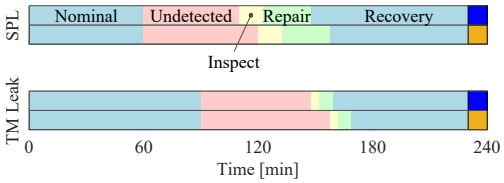
4.2.1. Effect of Communication Delay with Deactivated Agent Interruptibility (10 min, D) vs. (20 min, D)

This comparison aims to isolate the consequences of communication delays ranging from 10 to 20 minutes, under compound fault conditions involving both SPL damage and a TM leak. In both experiments, agent interruptibility is disabled. As shown in the FDD local time plots in Figure 6 (d), the FDPs from both experiments indicate nearly identical detection times, offering a fair basis for evaluating the impact of communication delay. As in Episode 1, a longer communication delay directly affects the timing of C2 decisions. However, the consequences are more extensive here: in addition to delaying the SPL inspection and repair, there are further delays in the inspection, repair, and shutdown of the TM system. These delays prolong the habitat exposure to a hazardous state, increasing both the duration and severity of the hazardous state.

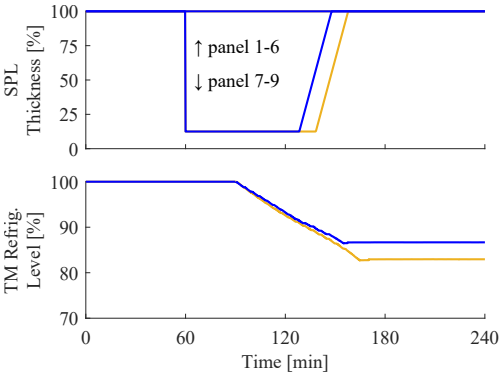
Compared to Episode 1, the leak in the TM causes a noticeable deviation in the inlet and outlet temperatures of the TM condenser, which is not the case when only the SPL is damaged. This telemetry data is used by the TM leak FDD to test for the possibility of a leak, prompting the DR to suspect a leak in the TM system. Note that this event occurs some time after the SPL damage is detected (during the SPL repair), as illustrated in the FDD plots.

Since agent interruptibility is disabled, the first fault detected is also inspected first and repaired first. Consequently, the SPL fault, detected before the TM leak, is addressed first. Here, the C2 is designed to disregard the potential criticality of the suspicion raised by the TM leak FDD and simply commands a TM inspection to be scheduled after completing the ongoing action. This order of actions is evident from the agent status shown in Figure 6 (c).

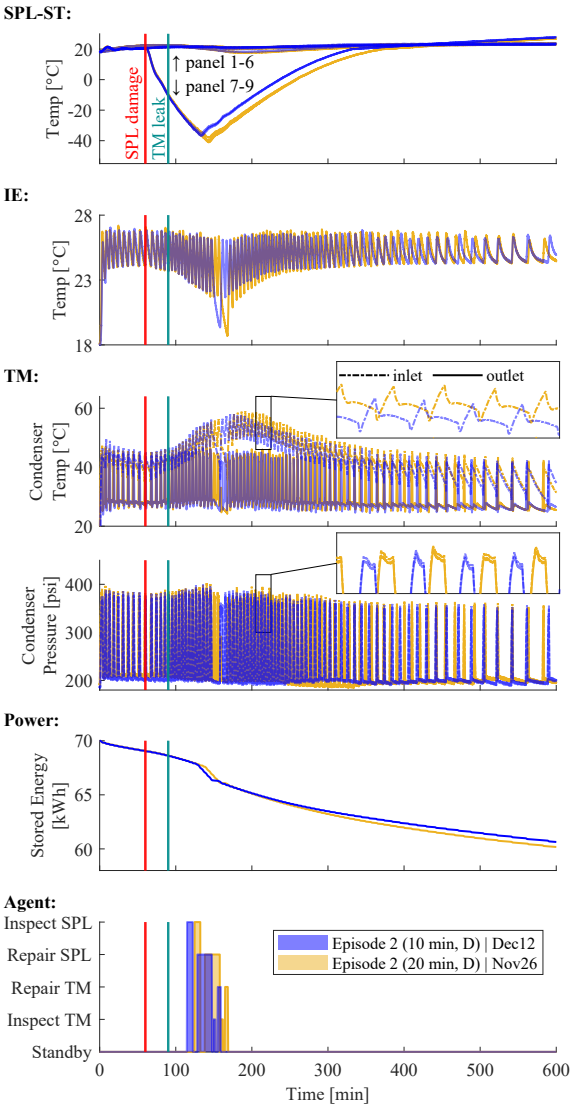
Upon inspection to confirm the TM refrigerant leak status as bad, the TM system is shut down by the C2. Note that the TM is restarted once the agent completes the repair to stop the leak. The TM shutdown duration is visible in the data by the presence of several missing operational cycles in condenser temperature and pressure measurements, as shown in TM plots in Figure 6 (c). Due to this shutdown, the IE temperature experiences a more severe decline, as depicted in the IE plot in Figure 6 (c), compared to that in Episode 1. Also, the same communication delay in Episode 2 results in a prolonged exposure of the habitat to lower IE temperatures as compared to Episode 1.



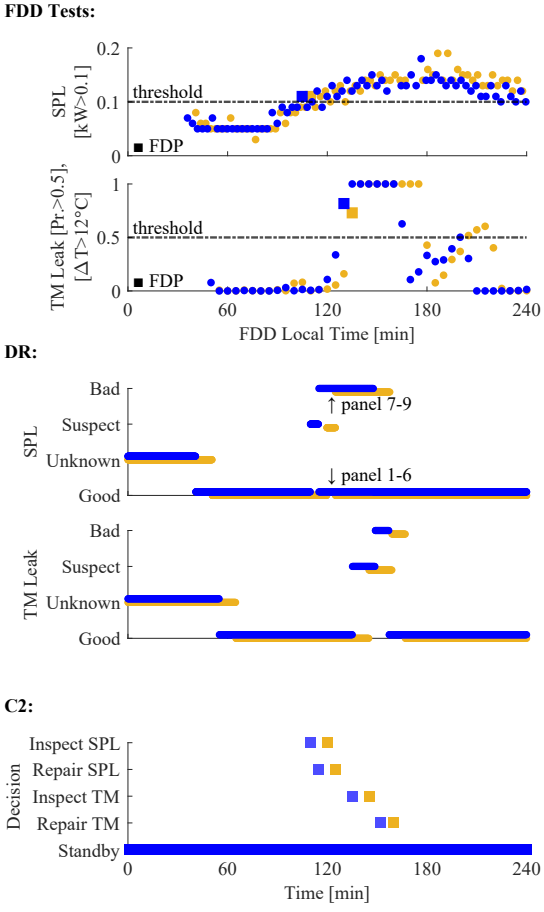
(a) Timeline showing disruptions status



(b) True health states



(c) Sensor measurements from habitat



(d) Perceived health states

Figure 6: Episode 2 (SPL damage and TM leak): visualization of how communication delay affects perceived system health, decision-making, and recovery, with agent interruptibility deactivated.

Figure 7 illustrates the system-level performance calculation for these two experiments. A quantitative assessment of the results shows that a 10-minute delay, with disabled agent interruptibility, reduces resilience by 3.7% in the HRM metric (TM leak). In terms of the DMN metric (reported in 10^{-2}/min), resilience decreases by 11.61 for the SPL and by 5.26 at the system level. The ABM metric, which reflects retained performance, shows corresponding reductions of 1.92% and 2.04%, respectively. A more comprehensive analysis of these findings is presented in Section 5.

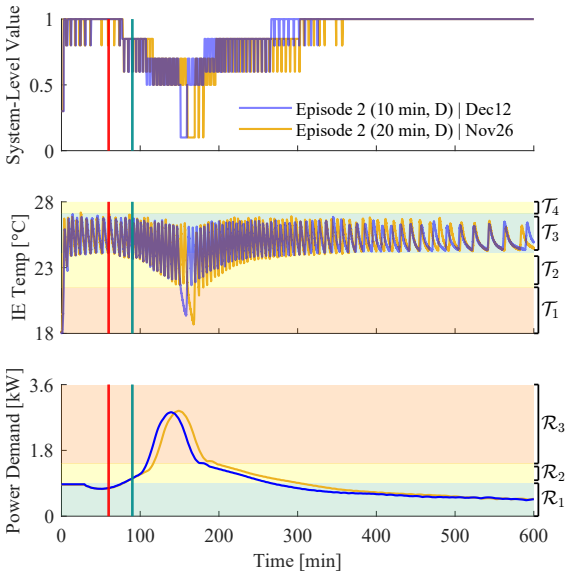


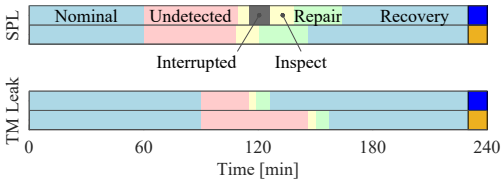
Figure 7: Episode 2 (SPL damage and TM leak): determination of system-level performance curve based on the sensor measurements from the habitat, with agent interruptibility deactivated.

4.2.2. Effect of Agent Interruptibility with On-Board FDDs (0 min, D) vs. (0 min, A)

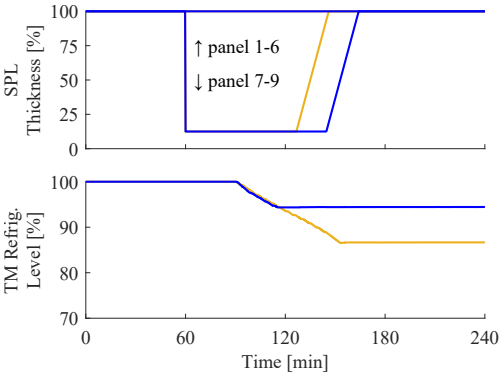
This one-to-one comparison considers the question of how the agent interruptibility feature affects decision-making and, consequently, habitat recovery. The analysis specifically focuses on a comparable pair of experiments with 0-minute communication delay, which is equivalent to the on-board FDD configuration. Both experiments share similar fault detection timing. The TM leak is identified and communicated to C2 while the agent is still in the early stages of inspecting the SPL, approximately at minute 115, as shown in the FDD plots in Figure 8 (d). A visual summary of the corresponding event sequences for these two experiments is presented in Figures D.18–D.21 (Appendix D), highlighting how each configuration unfolds over time.

The effect of interruption is primarily reflected in the true health states of SPL and the TM refrigerant level. When interruptibility is activated, the TM leak is prioritized, leading the agent to stop the SPL inspection task to address the TM fault immediately. This interruption appears in the SPL recovery trajectory shown on the timeline in Figure 8 (a). Only after resolving the TM leak does the agent return to complete the SPL inspection and repair. Consequently, enabling agent interruptibility in this case mitigates the consequences of the TM leak. This prioritization maintains the TM refrigerant at a higher charge level, although it also prolongs the SPL damage and delays its return to nominal conditions.

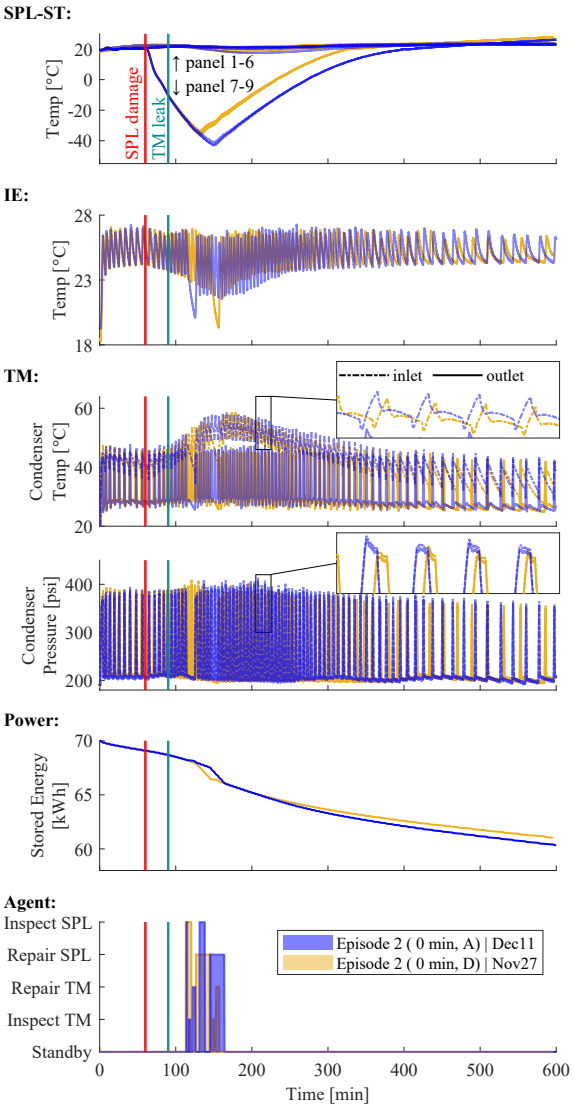
The choice of the agent policy noticeably influences the habitat performance, as the recovery trajectories diverge early in the inspection phase. Addressing the TM leak first (0 min, A) leads to an earlier shutdown of the TM system, while thermal conditions at the SPL-ST interface have not yet deteriorated significantly. As a result, in case (0 min, A), the habitat experiences a less severe drop in IE temperature, as shown in the IE plot in Figure 8 (c). However, because the SPL remains damaged longer, restoring the interior temperature to nominal after restarting the TM happens more slowly. Furthermore, enabling agent interruptibility in this scenario leads to a reduction in energy efficiency, as indicated by the lower amount of stored energy observed at the end of the experiment. The larger amount of energy consumed in the agent interruptible case can be attributed to two factors. First, additional energy is required



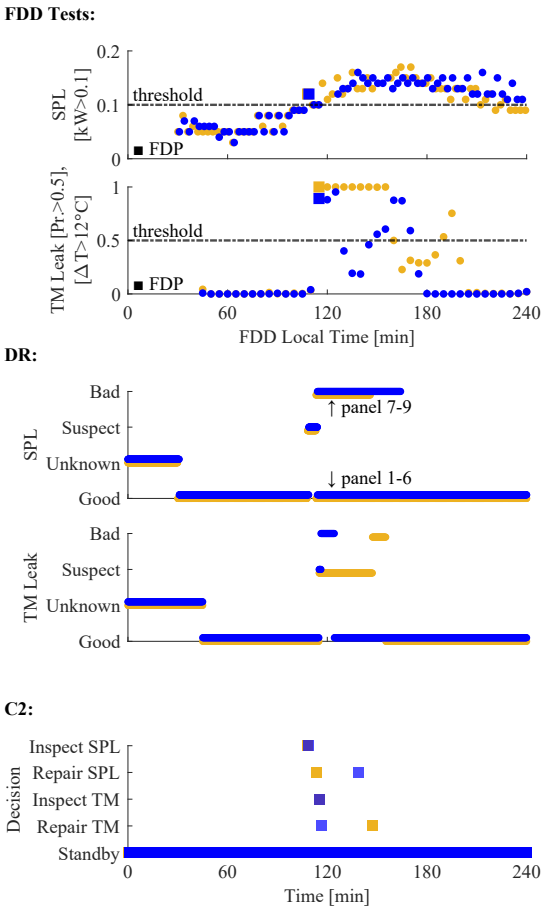
(a) Timeline showing disruptions status



(b) True health states



(c) Sensor measurements from habitat



(d) Perceived health states

Figure 8: Episode 2 (SPL damage and TM leak): visualization of how agent interruptibility affects perceived system health, decision-making, and recovery, with on-board FDD computation.

to resume an unfinished action. Second, the interior temperature must be maintained by the TM over a longer period with insulation loss. This observation is evident from the increase in the rate of TM cycles.

The system-level value is computed for these cases and shown in Figure 9. The ABM resilience metric based on these performance curves shows a 1.9% reduction (from 88.09 to 86.19) when the agent interruption policy is in use. Postponing the SPL repair to prioritize the TM leak does not improve the system-level performance values, as it causes less favorable combinations of IE temperature and power consumption. This outcome may be attributed to the greater severity of the SPL damage compared to the TM leak. A more detailed exploration of these observations is provided in Section 5.

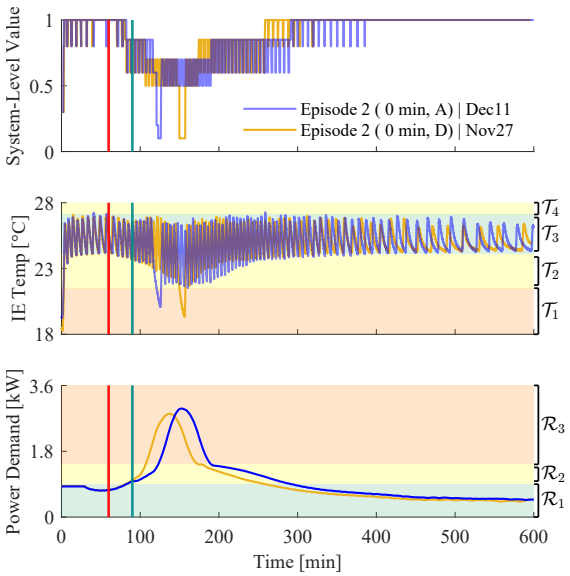


Figure 9: Episode 2 (SPL damage and TM leak): determination of system-level performance curve based on the sensor measurements from the habitat, with on-board FDD computations.

5. Discussion

To further examine the design configurations in Episode 1 (SPL damage only) and Episode 2 (SPL damage and TM leak), Table 7 and Figure 10, respectively, present the marginalized resilience distributions across the experimental realizations. Figure 10 summarizes these results using box plots that illustrate the median response and variability across design configurations. Because the configuration with a 30-minute delay and deactivated agent interruptibility lacks a comparable counterpart, it is excluded from the marginalization of the agent interruptibility design feature (activated vs. deactivated) to ensure a fair basis for comparison. To complement these descriptive comparisons, Table 8 reports posterior summaries from a Bayesian hypothesis testing analysis of the repeated communication delay experiments. The primary quantities of interest are the pairwise delay contrasts, which directly measure the change in resilience associated with increasing communication delay. For each contrast, the table reports the posterior median and the 95% central credible interval, thereby characterizing epistemic uncertainty under the limited number of repeated trials. The posterior median gives the most representative estimate of the delay-induced resilience change, while the 95% credible interval shows the plausible range supported by the conducted experiments. The table further provides a basis for assessing whether the inferred delay contrasts are credibly separated from zero. This supplementary analysis provides additional statistical context for the design-oriented interpretation of the experiments, while Appendix B presents the full posterior results. The analysis and findings should be interpreted for the designed disruption episodes considered in this study, where deterministic fault-injection settings and operational assumptions are prescribed to enable fair comparisons across experiments.

The resilience trends across both episodes clearly indicate that communication delay leads to a decline in all relevant metrics, including the recovered state in the HRM metric, recovery time in the DMN metric, and the overall

Table 7: Episode 1 (SPL damage): evaluation of resilience metrics across design configurations.

Component/Habitat	Design Configuration	Metric DMN [$10^{-2}/min$]	Metric ABM [%]
SPL	(0 min, -)	124.82	86.46
	(10 min, -)	114.00	85.00
	(20 min, -)	105.56	83.65
System-Level Value	(0 min, -)	44.98	89.82
	(10 min, -)	39.67	87.52
	(20 min, -)	38.29	85.13

performance loss indicated by the ABM metric. This pattern is consistent with the anticipated consequences of delayed fault detection and recovery, a point also reflected in the predominantly negative posterior delay contrasts reported in Table 8. From a mission-planning perspective, more negative posterior medians, together with credible intervals that place more of their range below zero, help identify cases in which delayed ground-based FDD imposes a clearer resilience penalty and on-board FDD capability has greater operational value. Additionally, the resilience obtained in Episode 1 (Table 7) is generally higher than that in Episode 2 (Figure 10) across almost all resilience metrics. This result is expected, as Episode 2 builds on the disruption from Episode 1 by incorporating the additional TM leak fault, which further challenges habitat recovery. To further examine whether the observed resilience changes are associated with communication delay, agent interruptibility, or their interaction, Appendix C provides a Bayesian GLM-ARD regression analysis for Episode 2 across the resilience metrics.

The higher DMN resilience metric for the SPL component, relative to the system-level value, suggests that habitat global recovery takes longer than local recovery of the SPL subsystem. This highlights the fact that the effects of communication delay are more persistent at the system level than in isolated components. Consequently, resilience assessments should account for system-wide (global) dynamics rather than relying solely on localized behavior, especially when estimating recovery times and determining appropriate criteria for resetting FDDs to reflect restored habitat functionality.

Table 8: Posterior delay-contrast summaries for Episodes 1 and 2. Values are reported as posterior median [95% central credible interval], with interval endpoints given by the 2.5th and 97.5th percentiles. The weak-prior reference is shown for comparison.

Component/Habitat	Contrasts	Metric HRM [%]	Metric DMN [10^{-2}/min]	Metric ABM [%]
	Δ_{prior}	0.00 [−28.32, 28.32]	0.00 [−51.11, 51.11]	0.00 [−15.08, 15.08]
Episode 1: SPL damage				
SPL	Δ_{10-0}	–	−8.30 [−28.88, 12.15]	−1.11 [−5.88, 3.65]
	Δ_{20-10}		−6.50 [−27.00, 14.00]	−1.03 [−5.81, 3.73]
System-Level	Δ_{10-0}	–	−4.07 [−19.65, 11.58]	−1.75 [−6.84, 3.35]
	Δ_{20-10}		−1.04 [−16.64, 14.57]	−1.83 [−6.90, 3.24]
Episode 2: SPL damage and TM leak				
SPL	Δ_{10-0}	–	−4.94 [−21.77, 11.84]	−1.04 [−4.49, 2.43]
	Δ_{20-10}		−10.36 [−27.16, 6.45]	−1.89 [−5.36, 1.56]
TM Leak	Δ_{10-0}	−1.01 [−6.62, 4.60]	–	–
	Δ_{20-10}	−2.39 [−8.01, 3.21]		
System-Level	Δ_{10-0}	–	−0.22 [−10.55, 10.11]	−1.85 [−6.35, 2.66]
	Δ_{20-10}		−5.23 [−15.59, 5.08]	−3.69 [−8.18, 0.83]

Focusing on the impact of communication delay in Episode 2, both the ABM and DMN metrics show that the performance degradation of the local SPL component is roughly linear with latency. The ABM metric specifically follows a gentler decline compared to the system-level loss, a trend consistent with observations from Episode 1. How-

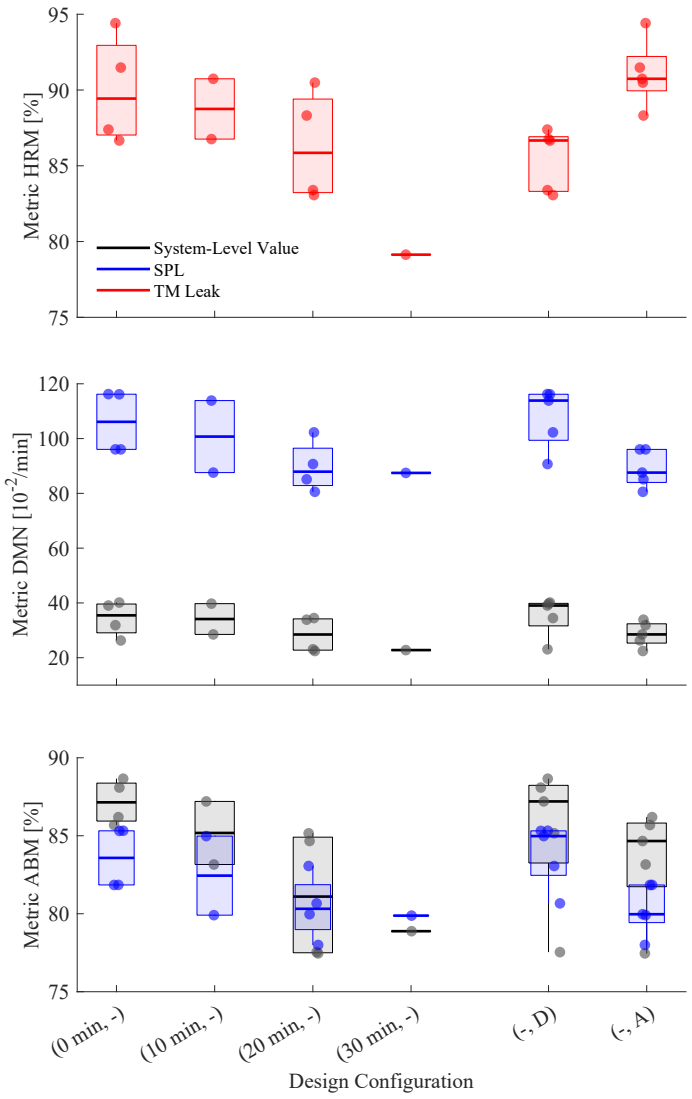


Figure 10: Episode 2 (SPL damage and TM leak): comparison of resilience metrics for different design configurations.

ever, the linear trend in resilience loss does not hold for the calculated HRM metric for the TM leak. Furthermore, the system-level habitat performance in the compound-fault scenario of Episode 2 displays a distinctly non-linear response, exhibiting disproportionately greater losses as communication delay increases. The observed nonlinear trend remains consistent under variations in temperature and power demand thresholds, as well as assigned performance values used to define the system-level performance curve, as confirmed by curvature-based sensitivity analysis in Appendix A. This suggests that conclusions drawn from short communication delay scenarios may not generalize to longer delays, potentially leading to unpredictable decision-making unless sufficient data are available for scenarios involving extended delays. Such effects are expected to be even more pronounced in complex habitat systems with greater interdependencies. The posterior analysis of delay contrasts in Table 8 further supports this interpretation. In Episode 1, the posterior delay contrasts remain relatively similar across the compared delay increments, indicating only modest separation in resilience between successive delay settings. In Episode 2, by contrast, the larger negative contrasts indicate that the resilience loss becomes more pronounced as communication delay increases. This pattern remains detectable despite the limited number of repeated experiments, suggesting that the observed differences are not fully explained by small-sample variability alone.

Another noteworthy observation in the compound scenario of Episode 2 is that, as communication delay increases, the variation in the ABM resilience metric remains relatively stable for the local SPL component, while it increases noticeably at the system level. For the SPL component, this stable variation is likely due to the damage having already reached its worst-case intensity, rendering further delay-induced uncertainty less influential and even diminishing its effect. In contrast, the increased variability in the system-level performance can be attributed to heightened sensitivity of the system resilience to the chosen strategy for managing the faults, specifically, the activation or deactivation of agent interruptibility. This outcome suggests that an improper interruption policy is increasingly detrimental under longer communication delays.

The effect of agent interruptibility, which is relevant only in Episode 2, exhibits a mixed impact. Activating agent interruptibility, which prioritizes TM leak recovery actions over SPL damage, results in a better recovered state for the refrigerant leak, as reflected in the HRM metric. However, this prioritization negatively affects the SPL recovery time and associated performance loss, as captured by the DMN and ABM metrics. Consequently, the system-level resilience, computed based on the habitat's global performance, declines with the activation of agent interruptibility. This behavior suggests that mission success is more strongly influenced by SPL-related losses than by the TM leak. Therefore, prioritizing TM leak mitigation does not necessarily improve the system-level resilience of the habitat. This conclusion, however, may differ in scenarios involving repeated cycles of disruption and recovery, and could also vary depending on the severity of the faults.

As shown in Figure 10, the agent interruptibility feature also influences the variability of resilience across experimental realizations. When this feature is activated, the spread of resilience values appears more symmetric than in the deactivated case, indicating that deviations are roughly equally likely on either side of the median, and the system behaves more consistently. In contrast, when agent interruptibility is deactivated, the distribution skews toward lower values — most runs achieve relatively high resilience, but a long one-sided tail emerges, reflecting rare but severe performance degradations. This asymmetry suggests the presence of specific, uncommon scenarios that can significantly impair the system resilience in this design.

It is informative to examine the specific details that resulted in the outliers. One such outlier is an experiment conducted with deactivated agent interruptibility and a 20-minute communication delay (20min, D). In this realization, a rare sequence of events occurred: the TM leak was detected while the agent was in the middle of a 5-minute preparation phase for SPL inspection. Due to the deactivated agent interruptibility and a first-come, first-served action execution policy, the agent proceeded with SPL inspection, followed by TM leak inspection, TM shutdown, SPL repair, TM leak repair, and finally TM restart. As the TM was turned off immediately after the inspection confirmed the leak, it remained off throughout the lengthy SPL repair, causing a sharp decline in system-level resilience. In contrast, the case with agent interruptibility—although not necessarily more efficient—exhibited more uniform and controllable behavior under similar conditions.

Based on the impact of agent interruptibility on overall system behavior, particularly as reflected in the ABM metric, we argue that the adverse effects of communication delay can be mitigated through a more efficient and effective safety control strategy. In this context, deactivating agent interruptibility emerges as a more favorable approach. Interrupting the agent is not necessarily effective when done without an awareness of the remaining time needed to complete the current task and without evaluating the time-to-criticality of the ongoing failure modes using real-time data. Simply preempting the current task to address the TM leak (which is treated as higher priority) may delay the ongoing action, potentially near completion, and require restarting it from the beginning, ultimately worsening the situation. These experiments illustrate a relatively simple example of the deep informational dependencies present in complex systems like deep space habitats (Murali Krishnan et al., 2025).

In future C2 design, this dependency can be represented through a cost-aware preemption condition, rather than a priority-based interruption rule, as in Eq. (4). Let $r_c(t)$ denote the remaining time to complete the current task, $r_n(t)$ the time required to execute the newly issued task, $r_{\text{resume}}(t)$ the additional time required to resume or restart the interrupted task after $a_n(t)$ is completed, and $S_{c\to n}(t)$ the penalty associated with switching from $a_c(t)$ to $a_n(t)$. If $C_i(t, \Delta t)$ denotes the expected system-level cost of leaving the fault associated with task i unresolved for an additional duration Δt, which can be informed by its time-to-criticality, then interruption could be justified when $C_n(t, r_c(t)) > S_{c\to n}(t) + C_c(t, r_n(t) + r_{\text{resume}}(t))$. Here, the left-hand side represents the cost of allowing the newly detected fault to wait until the current task is completed, while the right-hand side represents the switching penalty and the cost of delaying completion of the current task.

5.1. Lessons Learned and Implications for Mission Management

The findings from this series of experiments provide insight into how communication delays and agent interruptibility affect habitat resilience and recovery dynamics. Here, we summarize the key lessons learned and their implications for mission management in support of future autonomous health management systems for deep space habitats:

- Communication delays substantially degrade habitat resilience by postponing fault detection, inspection, and recovery phases.
- In compound fault episodes, shorter communication delays do not extrapolate well to extended-delay conditions, implying that as system complexity and interdependencies grow, the adverse and unpredictable impacts of communication delays intensify.
- Ground-based computational resources might be exploited to improve FDD results (by reducing uncertainty and false outcomes, or by adding more information such as damage localization), but they only make sense if the resulting gains in time savings and safety outweigh the added cost and complexity. Assigning intensive tasks to ground control while keeping lighter tasks local creates a trade-off, and finding an optimal balance and employing digital twins of varying fidelities is an interesting aspect for future exploration. However, if communication blackouts are expected, this may not be a productive path to pursue.
- System-level recovery is more sensitive to the impact of communication delay than local repair of the faults. Individual components can be restored quickly, but bringing the entire habitat back to full functionality typically takes much longer.
- Fault-priority rules and static criticality rankings provide a useful baseline for identifying high-consequence faults. However, in highly coupled, resource-constrained habitat systems, using these static risk-assessment analyses as the sole basis for real-time task prioritization can lead to recovery decisions that appear locally urgent but are unfavorable at the system level. As observed in our experiments, prematurely interrupting a near-complete task to address a TM leak can force a full restart and worsen overall recovery. Future autonomous health management systems should therefore shift toward dynamic, state-aware policies in which preemption decisions account for fault urgency, time-to-criticality, and the system-level cost of interrupting an ongoing task.
- Agent interruptibility reflects a management trade-off between predictable performance and average resilience in the design of autonomous safety protocols. Enabling interruption, although it does not improve system-level resilience in these experiments, yields more uniform and predictable outcomes, whereas disabling it can produce higher average resilience but with greater variability and rare severe degradation. For safety-critical applications such as deep space habitats, predictable recovery may be preferable to higher average resilience with high-consequence outliers.

6. Conclusions

This study examines how communication delays associated with FDD execution affect decision-making and ultimately the resilience of an autonomous space habitat system. Recognizing that the decision-making logic itself could significantly influence the outcomes, we also explore a policy that allows the agent to be interrupted in the middle of a task and redirected to a higher-priority action. Using the HARSH testbed, we experimentally evaluate the habitat behavior under single and confounding disruption episodes involving damage to the structural protective layer as well as a thermal management refrigerant leak. We assess the emergent behaviors associated with typical decision-making policies in certain circumstances, providing insight into how various design configurations may impact the overall habitat performance.

Given the range of experiments conducted, we conclude that communication delays, particularly those of Mars-scale duration, consistently degrade resilience across all evaluated metrics. These effects are more lasting at the system level than at the individual component level, underscoring the need to evaluate habitat designs through a global lens rather than relying solely on isolated subsystem testing. Moreover, in compound disruption episodes, the impact

of communication delay grows disproportionately with its duration, highlighting the risks of extrapolating system behavior from short-delay experiments to longer-delay conditions. An agent interruptibility policy introduces an additional layer of complexity into the decision-making process. While it enables timely responses to emergent faults by preempting ongoing tasks, our results show that this strategy does not always enhance resilience. In scenarios where fault detections occurred in close succession, preemptive task switching resulted in delayed recovery and increased energy consumption, ultimately reducing system-level performance. These findings emphasize that, although agent interruptibility can be beneficial in principle, its effectiveness depends on careful consideration of task progress and fault criticality.

Beyond the specific findings, this work demonstrates the value of cyber-physical testing in exploring design trade studies for future deep space habitats. Unlike purely numerical simulations, our approach captures hardware constraints, real-world uncertainties, and the compound effects of system interdependencies. As a result, it offers a structured basis for assessing autonomy-enabling policies under realistic operational conditions.

Future studies may explore embedding adaptive or context-aware decision policies that account for fault criticality, task progress, available resources, and predicted consequences. We hope that insights from this work can inform the development of robust autonomous health management strategies for deep space habitats where communication with ground support is constrained.

CRediT authorship contribution statement

Motahareh Mirfarah: Conceptualization, Data curation, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Hyunjin Park:** Conceptualization, Data curation, Investigation, Software, Writing – review & editing. **Zhiwei Chu:** Conceptualization, Data curation, Investigation, Methodology, Software, Writing – review & editing. **Sreehari Manikkan:** Conceptualization, Investigation, Methodology, Software, Writing – review & editing. **Manuel Salmeron:** Investigation, Methodology, Software, Writing – review & editing. **Herta Montoya:** Data curation, Investigation, Methodology, Software, Writing – review & editing. **Seungho Rhee:** Investigation, Methodology, Software, Writing – review & editing. **Christian E. Silva:** Data curation, Investigation, Methodology, Writing – review & editing. **Ilias Bilionis:** Conceptualization, Funding acquisition, Investigation, Writing – review & editing. **Shirley J. Dyke:** Conceptualization, Funding acquisition, Investigation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by a Space Technology Research Institute grant (number 80NSSC19K1076) from NASA's Space Technology Research Grants Program. We acknowledge the contributions of RETHi team members to this work, including Murali Krishnan, Kairui Hao, and Dawn Whitaker.

Appendix A. Sensitivity of System-Level Resilience to Threshold and Performance Value Selection

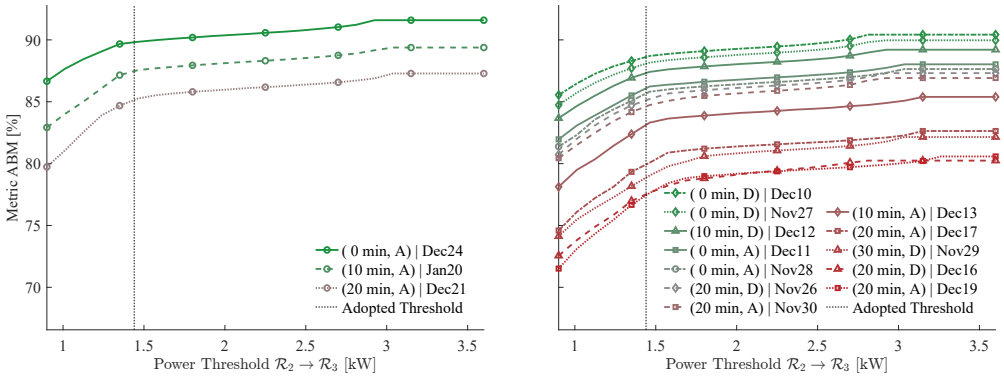
The system-level resilience metrics rely on mission-driven objectives used to discretize habitat performance states through performance-region thresholds and the associated performance value matrix. To ensure that the reported conclusions are not artifacts of these modeling assumptions, sensitivity analyses are conducted with respect to variations in both threshold definitions and assigned performance values.

Appendix A.1. Sensitivity to Performance Region Thresholds

System-level resilience evaluation classifies habitat operation according to discretized performance levels defined by temperature ($\mathcal{T}_1$–$\mathcal{T}_4$) and power demand ($\mathcal{R}_1$–$\mathcal{R}_3$) intervals (Table 3). The nominal operating region and extreme degraded regions are fixed based on experimentally observed habitat behavior under baseline and disruption conditions, while only intermediate transition boundaries (between $\mathcal{T}_1$-$\mathcal{T}_2$ and $\mathcal{R}_2$-$\mathcal{R}_3$) introduce some flexibility in threshold selection used for performance evaluation.

Because this discretization introduces modeling choices in defining performance regions, we examine whether the resulting resilience conclusions depend on the selected threshold boundaries. Among the utilized resilience metrics, the system-level DMN metric is largely insensitive to threshold selection since it depends only on the time required for recovery to nominal operation, whereas threshold definitions primarily influence the system-level ABM metric. Accordingly, variations in the ABM resilience metric are evaluated and presented separately for Episodes 1 and 2.

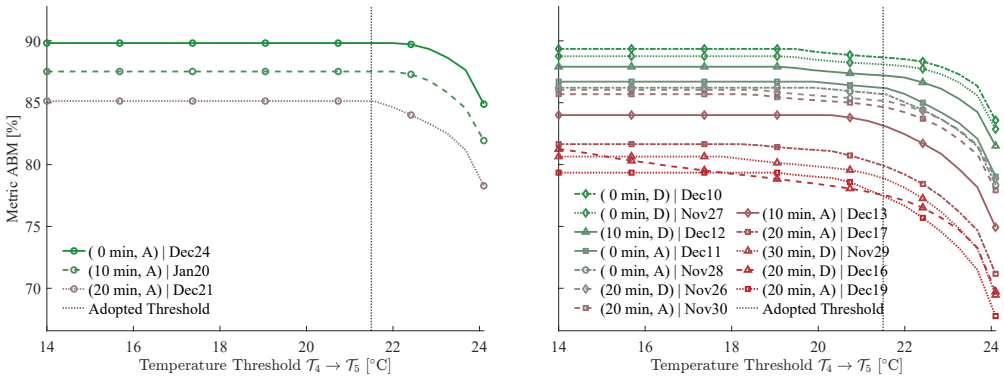
Figures A.11 and A.12 present one-at-a-time sensitivity analyses obtained by independently varying power demand and temperature thresholds across feasible ranges. Each curve corresponds to an experimental configuration under different communication delay conditions. Across all experiments, the qualitative ordering and trends of system-level resilience remain unchanged, indicating that the performance evaluation is not strongly affected by reasonable variations in threshold definitions.



(a) Episode 1: SPL damage

(b) Episode 2: SPL damage and TM leak

Figure A.11: Sensitivity of system-level ABM resilience to variations in the power demand threshold defining the transition between performance regions ($\mathcal{R}_2 \rightarrow \mathcal{R}_3$).

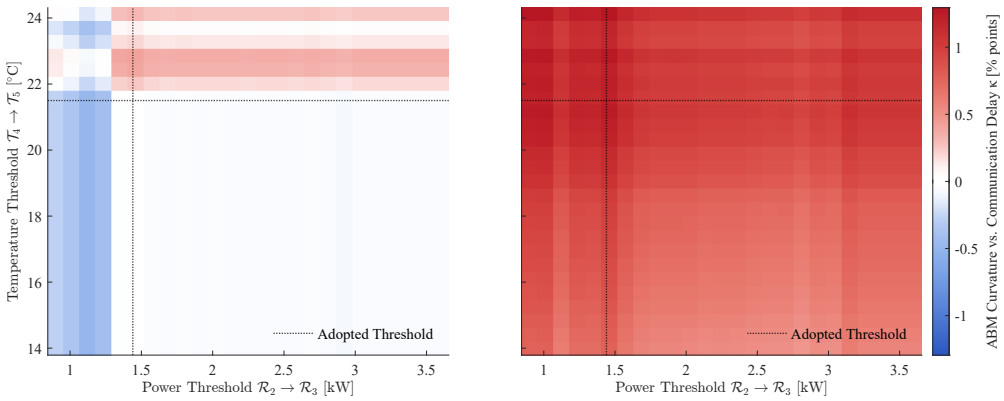


(a) Episode 1: SPL damage

(b) Episode 2: SPL damage and TM leak

Figure A.12: Sensitivity of system-level ABM resilience to variations in the temperature threshold defining the transition between performance regions ($\mathcal{T}_4 \rightarrow \mathcal{T}_5$).

Second, a joint two-parameter sensitivity analysis simultaneously varies both temperature and power thresholds to construct a two-dimensional response surface. The curvature of this surface (quantified through a quadratic fit



(a) Episode 1: SPL damage

(b) Episode 2: SPL damage and TM leak

Figure A.13: Sensitivity of κ, the curvature of system-level ABM resilience versus communication delay, to joint variation of temperature and power demand thresholds defining performance regions.

measuring deviation from linear behavior with respect to communication delay) is used to assess whether the observed resilience trends depend on specific threshold selections. The results demonstrate that in Episode 1 (SPL damage only) the curvature remains close to zero across the full parameter sweep, indicating that the delay–resilience relationship is well approximated by a linear trend and remains insensitive to reasonable threshold variations. In contrast, in Episode 2 (SPL damage and TM leak), persistent non-zero curvature is observed, demonstrating a nonlinear dependence of resilience on communication delay that remains present regardless of threshold selection. To quantify deviations from linear communication delay–resilience behavior, a curvature metric is defined based on system-level ABM resilience evaluated at the experimentally tested communication delays (0, 10, and 20 min). Denoting by R_d the mean system-level ABM resilience corresponding to delay d, curvature is computed as $\kappa = R_{10} - \frac{1}{2}(R_0 + R_{20})$. Values near zero indicate approximately linear dependence of resilience on communication delay, whereas persistent non-zero values indicate nonlinear delay sensitivity.

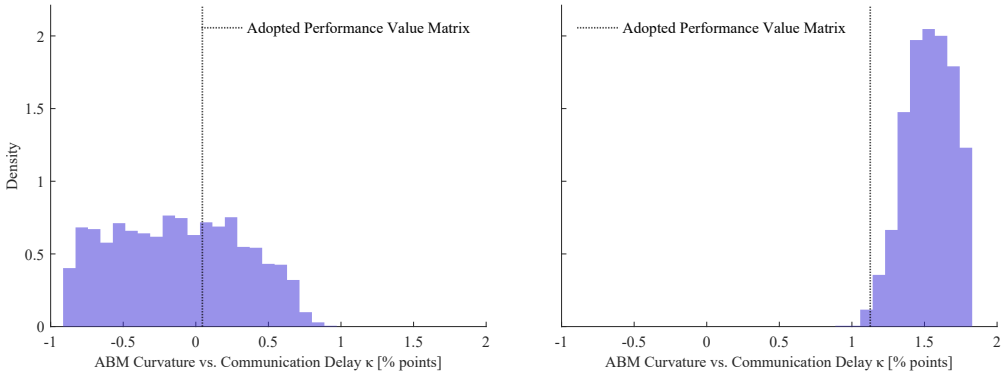
A two-dimensional sensitivity analysis of system-level ABM resilience curvature is performed by jointly varying temperature and power demand thresholds, shown in Figure A.13. Panel (a) corresponds to Episode 1 (SPL damage only), while panel (b) represents Episode 2 (SPL damage and TM leak). Across the explored threshold space, Episode 1 exhibits curvature values close to zero, indicating that the delay–resilience relationship remains well approximated by a linear trend and is largely insensitive to reasonable variations in performance-level definitions. In contrast, Episode 2 displays persistent non-zero curvature throughout the threshold sweep, demonstrating a nonlinear dependence of system-level resilience on communication delay that remains present regardless of threshold selection. These results indicate that the reported resilience trends are not artifacts of a particular discretization of temperature or power states. Rather, they arise from the underlying system dynamics and decision-making interactions. Accordingly, the conclusions of the study remain qualitatively unchanged under reasonable variations of threshold definitions.

Appendix A.2. Sensitivity to Performance Value Matrix Assignments

In addition to threshold selection, the system-level ABM metric depends on the numerical values assigned in the performance value matrix (Table 3). To assess robustness of the delay–resilience conclusions to the assigned performance values, we generate an ensemble of alternative performance value matrices through constrained sampling that preserves the same mission-driven precedents. Specifically, sampled matrices preserve: (i) the nominal operating region as the highest-utility state ($V_{31} = 1.0$); (ii) a monotonic decrease in utility as temperature or power conditions degrade; and (iii) a higher mission-objective value associated with maintaining crew thermal safety relative to power demand deviations.

For each sampled performance value matrix, we recompute the ABM resilience metrics for different communication delays (0, 10, and 20 min) and evaluate the corresponding curvature κ. Figure A.14 summarizes the resulting distribution of κ across 2000 sampled matrices for Episodes 1 and 2. The results indicate that Episode 1 remains concentrated near zero curvature, corresponding to an approximately linear delay–resilience relationship, whereas

Episode 2 exhibits persistent non-zero curvature, indicating nonlinear delay sensitivity. This demonstrates that the observed trends in resilience behaviors are not artifacts of a particular selection of performance values in the matrix. Collectively, these analyses demonstrate that the observed relationship between communication delay and system-level ABM resilience remains robust to reasonable variations in both performance region definitions and assigned performance values.



(a) Episode 1: SPL damage

(b) Episode 2: SPL damage and TM leak

Figure A.14: Sensitivity of κ, the curvature of system-level ABM resilience versus communication delay, to variations in performance value matrix assignments. Distributions are obtained from randomly sampled matrices that preserve the mission-driven performance value structure.

Appendix B. Bayesian Hypothesis Testing under Limited Experimental Repetitions

To assess epistemic uncertainty given the limited number of repeated delay experiments, we use a Bayesian approach to infer posterior distributions of delay contrasts for each resilience metric. Let $y_{i,d}$ denote the observed value of a given metric in repetition i at communication delay d. For each episode and metric, the repeated observations are modeled as $y_{i,d} \sim \mathcal{N}(\mu_d, \sigma^2)$, where μ_d is the mean associated with delay setting d and σ^2 is a variance shared across delay settings within that episode and metric. We place a weak proper Normal–Inverse–Gamma prior on the unknown parameters, specifically $\sigma^2 \sim \text{InvGamma}(\alpha_0, \beta_0)$ and $\mu_d \mid \sigma^2 \sim \mathcal{N}(m_0, \sigma^2/\kappa_0)$, where the prior location m_0 is centered on the corresponding metric scale. In the present implementation, we set $\alpha_0 = 1.5$, while κ_0 and β_0 are calibrated from the empirical scale of each metric so that the resulting prior remains weak and broad relative to the observed resilience variations. In standardized terms, this corresponds to an approximate prior coverage of ± 5 standard deviations.

The primary quantities of interest are the pairwise delay contrasts $\Delta_{10-0} = \mu_{10} - \mu_0$ and $\Delta_{20-10} = \mu_{20} - \mu_{10}$. For metrics appearing in both Episode 1 and Episode 2, the same prior scale is used in both episodes, while posterior inference is carried out separately using the observations from each episode. The posterior distributions for Episode 1 and Episode 2 are shown in Figures B.15 and B.16, respectively, along with their 95% credible intervals. Each posterior is compared with the corresponding weak prior to show how the repeated observations update the prior and reduce epistemic uncertainty. In general, Episode 2 exhibits narrower posterior bounds than Episode 1, consistent with the relatively greater amount of repeated experimental information available for that episode.

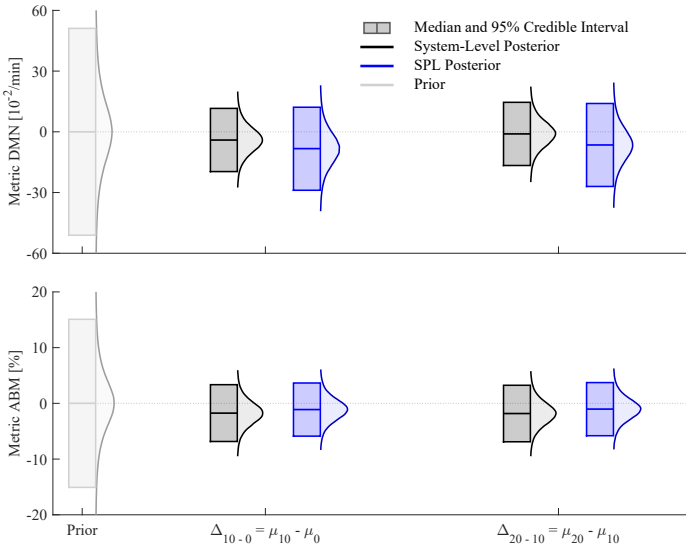


Figure B.15: Episode 1 (SPL damage): posterior distributions of the pairwise delay contrasts (Δ_{d2-d1}) from repeated evaluations of the resilience metrics, shown with their median and 95% credible intervals.

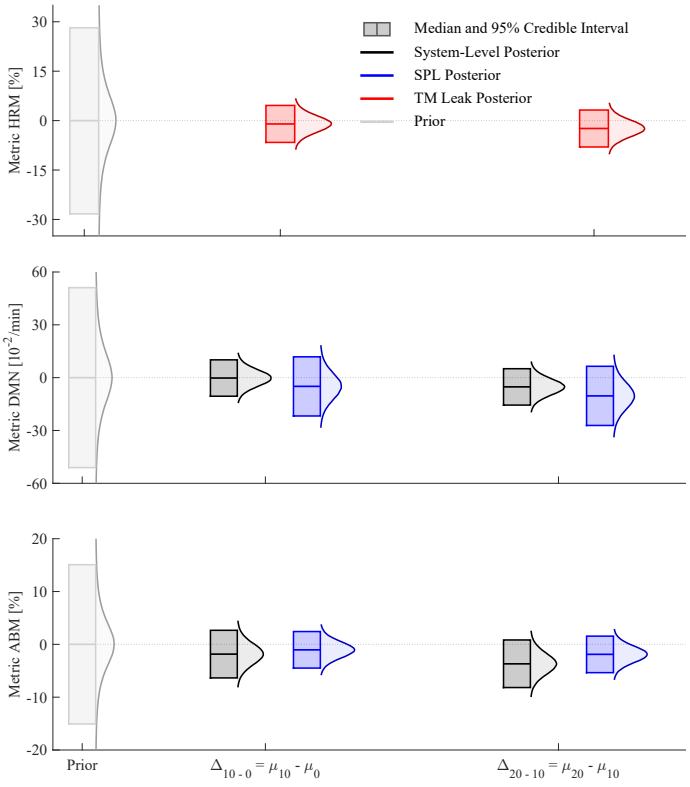


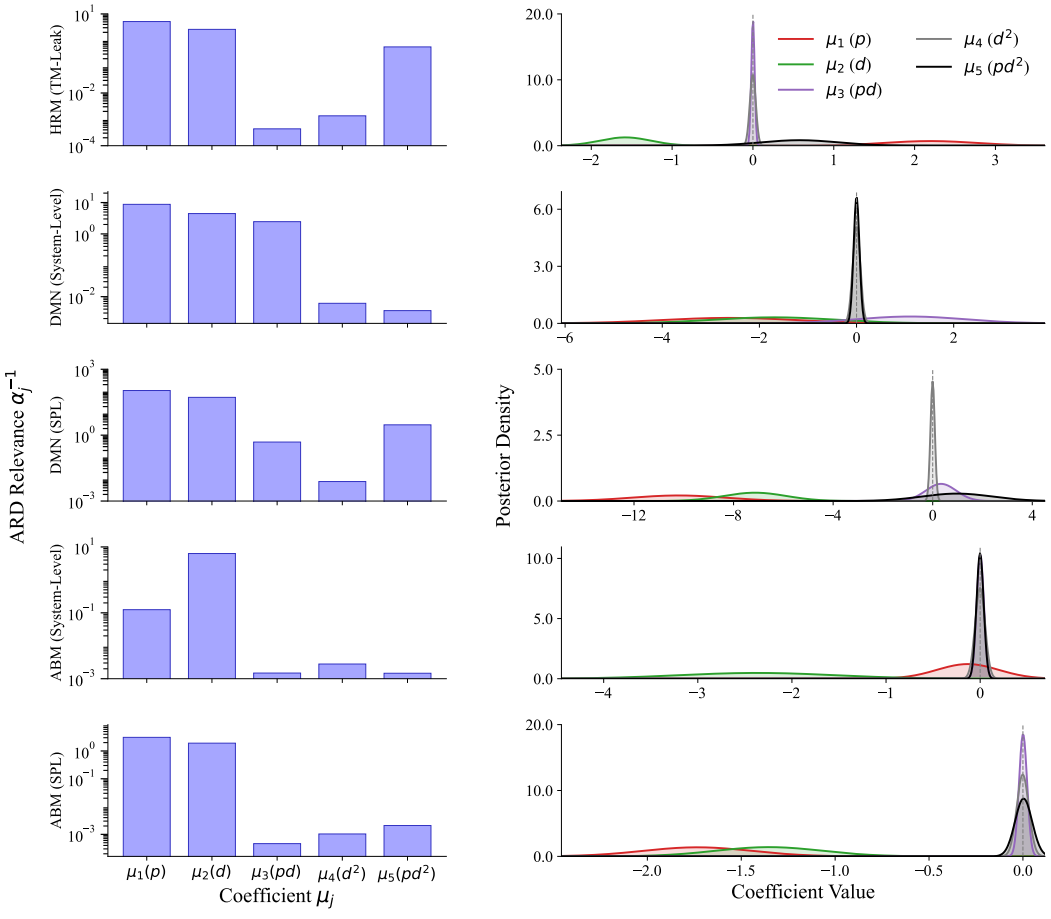
Figure B.16: Episode 2 (SPL damage and TM leak): posterior distributions of the pairwise delay contrasts (Δ_{d2-d1}) from repeated evaluations of the resilience metrics, shown with their median and 95% credible intervals.

Appendix C. Bayesian GLM-ARD Regression Analysis of Delay and Policy Effects

To examine the relative roles of communication delay, agent interruptibility, and their interaction in the Episode 2 resilience responses, we employ a Bayesian generalized linear model (GLM) with automatic relevance determination (ARD). This regression-based attribution analysis expresses the centered resilience response as a function of communication delay, the agent-interruptibility policy, their interaction, and nonlinear delay effects. Let y denote the centered resilience value for the metric under consideration, $p \in \{0, 1\}$ the agent interruptibility indicator (1 = activated, 0 = deactivated), and d the communication delay. The model includes the design terms p, d, pd, d^2, and pd^2, with each non-intercept column of the design matrix standardized before ARD estimation. The fitted Gaussian GLM is

$$y = \mu_0 + \mu_1 p + \mu_2 d + \mu_3 pd + \mu_4 d^2 + \mu_5 pd^2 + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2). \tag{C.1}$$

In ARD, each coefficient is assigned its own prior precision parameter α_j, which controls how strongly that coefficient is shrunk toward zero. Terms with large α_j values are strongly shrunk and are therefore weakly supported by the data. We report the inverse precision, $1/\alpha_j$, as an ARD relevance score, so that larger values indicate terms that are more strongly retained by the model. Figure C.17 summarizes the inferred ARD relevance scores $1/\alpha_j$, together with the posterior coefficient densities for the computed resilience metrics. The relevance scores provide an exploratory comparison of interruptibility, communication delay, interaction, and nonlinear delay terms, while the coefficient densities show the direction and uncertainty of their inferred effects.



(a) ARD relevance scores

(b) Posterior coefficient densities

Figure C.17: Bayesian GLM-ARD for Episode 2 (SPL damage and TM leak). The left column shows the ARD relevance $1/\alpha_j$ for each model coefficient, while the right column shows the corresponding posterior coefficient densities.

Figure C.17 shows that the communication-delay term $\mu_2(d)$ has relatively large ARD relevance scores across the resilience metrics, with posterior coefficient densities mostly shifted toward negative values. This provides additional statistical support for the interpretation that increasing communication delay is associated with reduced resilience after accounting for the tested interruptibility policy. The policy term $\mu_1(p)$ also shows noticeable relevance for several metrics, consistent with the discussion of agent interruptibility effects in Sections 4.2.2 and 5. By comparison, the interaction and nonlinear delay terms show lower ARD relevance for most resilience metrics, suggesting that the main delay and policy terms play a more prominent role in this regression representation.

Appendix D. Sample Illustration of Sequence of Events in Episode 2: SPL Damage and TM Leak

As with Episode 1 (Figure 3), we present two illustrative samples of the event sequence for Episode 2, corresponding to configurations (0min, A) and (0min, D). Figures D.18 and D.19 together depict the full sequence for (0min, A), while Figures D.20 and D.21 show the sequence for (0min, D). The event sequences shown here reflect the experimental data discussed in Section 4.2.2.

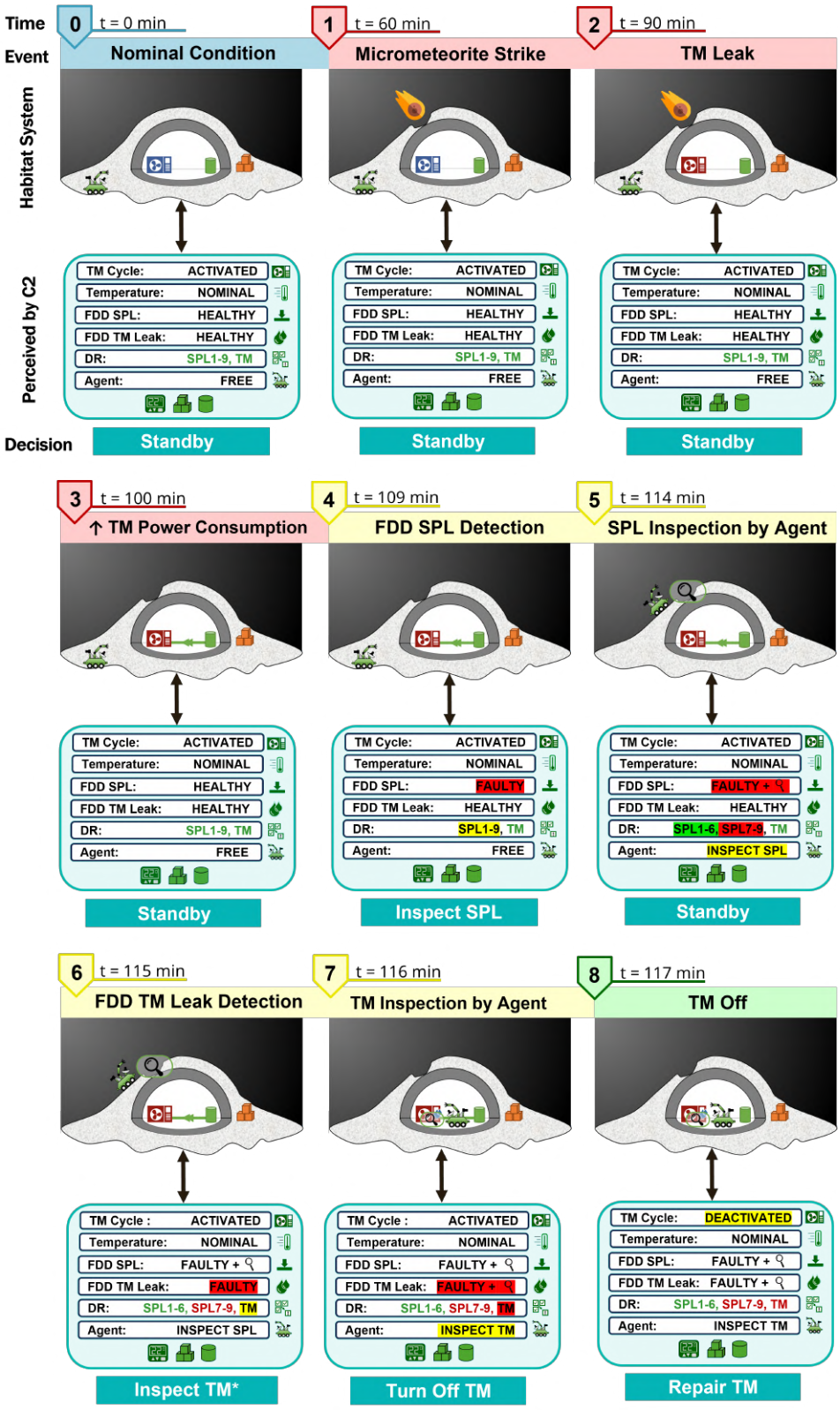


Figure D.18: Sample illustration of sequence of events in Episode 2 (SPL damage and TM leak) corresponding to (0 min, A) — Part 1.

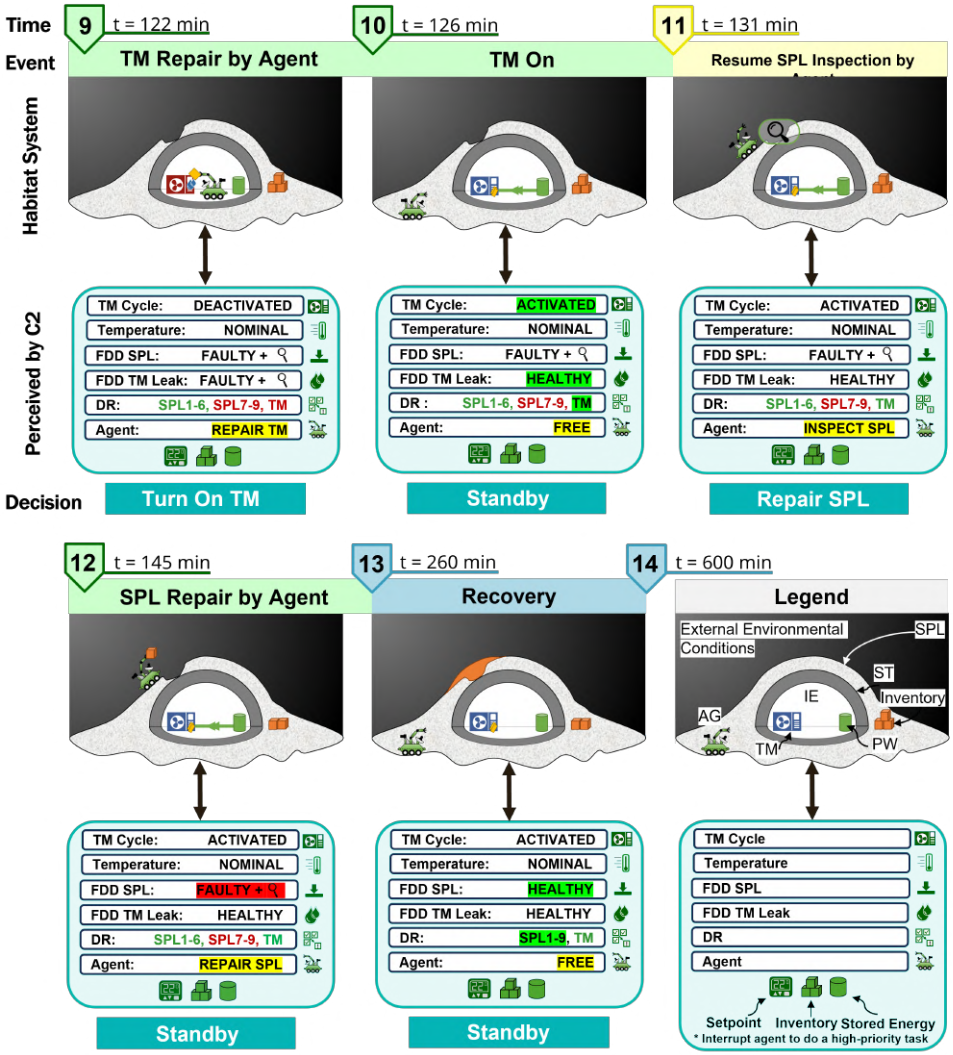


Figure D.19: Sample illustration of sequence of events in Episode 2 (SPL damage and TM leak) corresponding to (0 min, A) — Part 2.

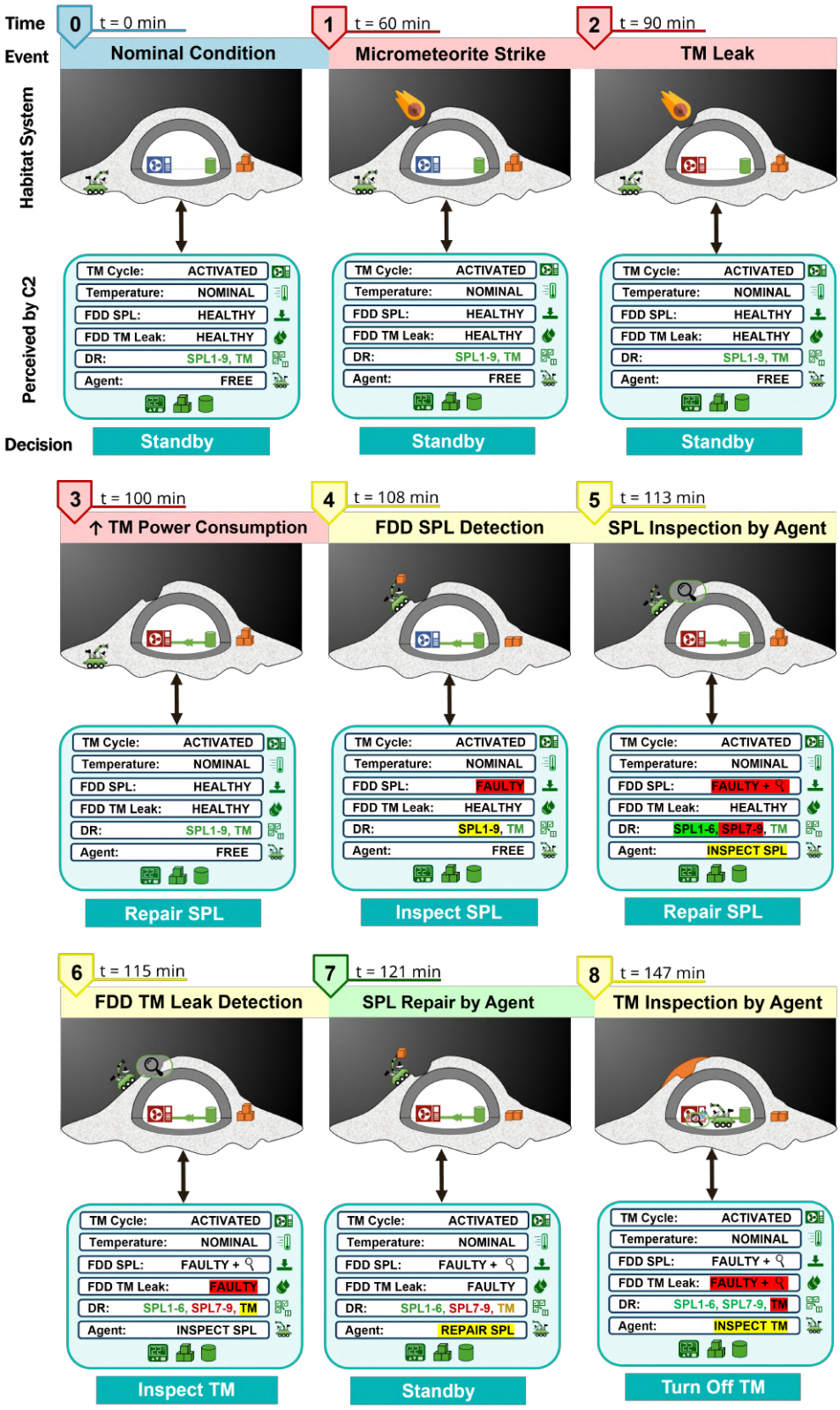


Figure D.20: Sample illustration of sequence of events in Episode 2 (SPL damage and TM leak) corresponding to (0 min, D) — Part 1.

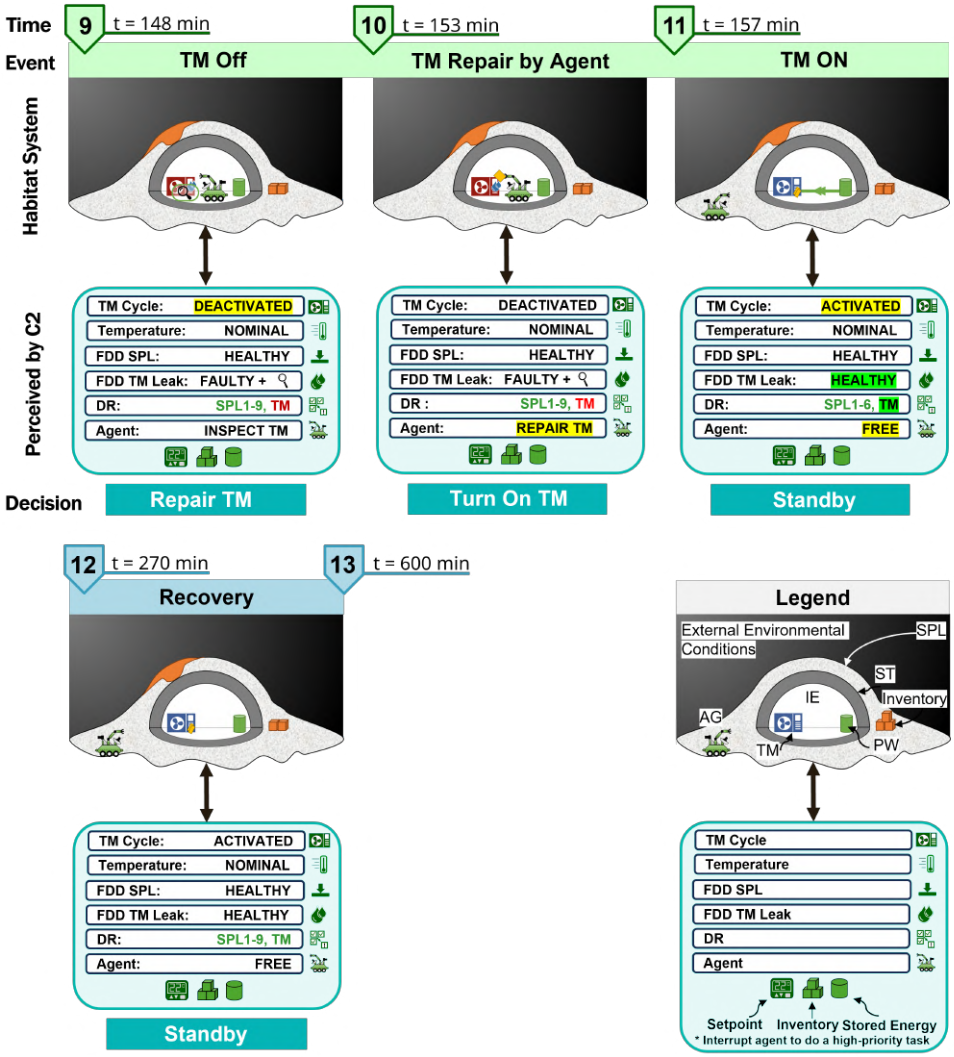


Figure D.21: Sample illustration of sequence of events in Episode 2 (SPL damage and TM leak) corresponding to (0 min, D) — Part 2.

Data Availability

The data supporting the findings of this study will be made publicly available upon publication. A DOI will be provided at that time.

References

Aaseng, G., Do, M., Frank, J., Fry, C., Sweet, A., 2023. Integrating planning, diagnosis and execution for vehicle systems management, in: 33rd International Conference on Automated Planning and Execution, pp. 1–9.

Aaseng, G.B., 2015. Scaling up model-based diagnostic and fault effects reasoning for spacecraft, in: AIAA Space 2015 Conference and Exposition, p. 4465.

Abercromby, A., Baumann, D., Berdich, D., Broyan, J., McCoy, T., Watkins, S., Zapp, N., 2025. Nasa's top human system research and technology needs for mars. Acta Astronautica 228, 931–939.

Azimi, M., Lund, A., Fu, Y., Montoya, H., Vaccino, L., Murali Krishnan, R., Rhee, S., Chebbo, L., Shahriar, A., Wang, Z., et al., 2025. Habsim: A modular-coupled virtual testbed for simulating extraterrestrial habitat systems. AIAA Journal 63, 1–13. doi:10.2514/1.J063845.

Badger, J., Hollaway, D., Taylor, E., 2020. When the Eyes Don't Have It: Autonomous Control of Deep Space Vehicles for Human Spaceflight, in: ASCEND 2020. Aerospace Research Central. volume Nov. 2020, p. 4163. doi:10.2514/6.2020-4163.

Badger, J.M., Strawser, P., Claunch, C., 2019. A distributed hierarchical framework for autonomous spacecraft control, in: 2019 IEEE Aerospace Conference, IEEE. pp. 1–8.

Ben-Itzhak, S., 2024. The use of artificial intelligence in outer space capabilities. SAIS Review of International Affairs 44, 5–30.

Bergman, T.L., Lavine, A.S., Incropera, F.P., DeWitt, D.P., 2011. Introduction to heat transfer. John Wiley & Sons.

Cheng, Y., Elsayed, E.A., Huang, Z., 2022. Systems resilience assessments: a review, framework and metrics. International Journal of Production Research 60, 595–622.

Choi, S.H., Moses, R.W., Park, C., Fay, C.C., 2020. Multipurpose cassegrain system, in: Earth and Space 2021. American Society of Civil Engineers, pp. 1156–1165.

Chu, Z., Hacker, T.J., Mirfarah, M., Manikkan, S., Bilionis, I., Dyke, S.J., Silva, C.E., Krishnan, M., Zhang, Z., Chi, M., 2025. "lessons learned from developing a health management system for the human-centered autonomous resilient space habitat". Manuscript under review.

Crusan, J.C., Craig, D.A., Herrmann, N.B., 2017. Nasa's deep space habitation strategy, in: 2017 IEEE aerospace conference, IEEE. pp. 1–11.

Dehghani, A., Sedighizadeh, M., Haghjoo, F., 2021. An overview of the assessment metrics of the concept of resilience in electrical grids. International Transactions on Electrical Energy Systems 31, e13159.

Dempsey, D., Sargent, R., Hobbs, A., Latorella, K., Huang, Z., Niemann, C., Chaikin, A., Lewis, R., Robinson, P., Karasinski, J., et al., 2024. Trade Space Analyses: Balancing Crew and Mission Design Parameters. Technical Report NASA/TM-20240011265. NASA.

Dempsey, R.C., 2017. The International Space Station: Operating an outpost in the new frontier. Government Printing Office.

Diana, L., Dini, P., 2024. Review on hardware devices and software techniques enabling neural network inference onboard satellites. Remote Sensing 16, 3957.

Dyke, S.J., Marais, K., Bilionis, I., Werfel, J., Malla, R., 2021. Strategies for the design and operation of resilient extraterrestrial habitats, in: Sensors and smart structures technologies for civil, mechanical, and aerospace systems 2021, SPIE. p. 1159105.

Frank, J., Spirkovska, L., McCann, R., Wang, L., Pohlkamp, K., Morin, L., 2013. Autonomous mission operations, in: 2013 IEEE Aerospace Conference, IEEE. pp. 1–20.

Hao, K., Bilionis, I., 2024. An information field theory approach to bayesian state and parameter estimation in dynamical systems. Journal of Computational Physics 512, 113139.

Jain, R., Cilento, M., Ulmer, J., Marais, K., 2024. Control effectiveness: Metric development and application to resilient lunar habitat design. AIAA Journal 62, 3208–3224. doi:10.2514/1.J063849.

Kessler, P., Prater, T., Nickens, T., Harris, D., 2022. Artemis deep space habitation: Enabling a sustained human presence on the moon and beyond, in: 2022 IEEE Aerospace Conference (AERO), IEEE. pp. 01–12.

Kitching, R., Mattingly, H., Williams, D., Marais, K., 2021. Resilient space habitat design using safety controls, in: Earth and Space 2021. ASCE, pp. 992–1003. doi:10.1061/9780784483374.091.

Landon, L.B., Karasinski, J.A., Morissette, L.G., McTigue, K.R., Parisi, M.E., Wu, S.c., Panontin, T.L., 2025. Assessment of the State of Communication Delay Research in Preparation for Missions Beyond Low Earth Orbit. Technical Report NASA/TM-20250003885. NASA.

Lim, M., Mousavi, P., Sirovljevic, J., You, H., 2020. Onboard artificial intelligence for space situational awareness with low-power gpus, in: Advanced Maui Optical and Space Surveillance Technologies Conference.

Maddox, I.D., Vera, A., Martinez, A., 2024. Earth independent operations development for nasa's mars campaign office, in: Accelerating Space Commerce, Exploration, and New Discovery (ASCEND).

Maghareh, A., Lenjani, A., Krishnan, M., Dyke, S., Bilionis, I., 2021. Role of cyber-physical testing in developing resilient extraterrestrial habitats, in: Earth and Space 2021. American Society of Civil Engineers, pp. 1059–1068.

Meß, J.G., Dannemann, F., Greif, F., 2019. Techniques of artificial intelligence for space applications-a survey, in: European workshop on on-board data processing (OBDP2019), European Space Agency.

Mirfarah, M., Lund, A., Dyke, S.J., 2025. Estimation of pressure leakage severity for space habitats using extended kalman filter. AIAA Journal 63, 1615–1628.

Montoya, H., 2024. Thermomechanical Real-Time Hybrid Simulation: Development and Execution for Lunar Habitats. Ph.d. dissertation. Purdue University. West Lafayette, IN. doi:10.25394/PGS.27957042.v1.

Montoya, H., Salmeron, M., Silva, C.E., Mirfarah, M., Dyke, S.J., 2025. Experimental validation of thermomechanical cyber-physical testing for the assessment of lunar habitats. Acta Astronautica 234, 644–659. doi:10.1016/j.actaastro.2025.04.053.

Murali Krishnan, R., Zhang, Z., Hao, K., Manikkan, S., Parsons, P., Dyke, S.J., Bilionis, I., Wang, J., Xue, C., Han, S., et al., 2025. Habsim-hms: A systems testbed to investigate situational awareness for extraterrestrial habitation. AIAA Journal 63, 389–403.

NASA, 2015. NASA Technology Roadmap, TA 4: Robotics and Autonomous Systems. Technical Report Rep. No. 492. NASA. Washington, DC. URL: https://www.nasa.gov/wp-content/uploads/2016/08/2015_nasa_technology_roadmaps_ta_4_robotics_and_autonomous_systems_final.pdf.

NASA Goddard Space Flight Center, 2024. Guideline for Failure Modes and Effects Analysis and Risk Assessment. Technical Report GSFC-HDBK-8004. NASA Goddard Space Flight Center. URL: https://standards.nasa.gov/sites/default/files/standards/GSFC/Baseline/0/GSFC-HDBK-8004_Approved_1.pdf.

Oche, P.A., Ewa, G.A., Ibekwe, N., 2021. Applications and challenges of artificial intelligence in space missions. IEEE Access 12, 44481–44509.

Parisi, M., McTigue, K., Wu, S.C., Karasinski, J., Panontin, T., Landon, L., Vera, A., 2024. The impact of delayed communication on nasa's human-systems operations: Preliminary results of a systematic. Advances in Human Factors of Transportation , 49.

Parisi, M., Panontin, T., Wu, S.C., McTigue, K., Vera, A., 2023. Effects of communication delay on human spaceflight missions, in: 14th International Conference on Applied Human Factors and Ergonomics (AHFE), pp. 64–73.

Pillai, M.K.R., 2024. Systems Health Management for Resilient Extraterrestrial Habitation. Ph.D. thesis. Purdue University Graduate School.

Rajasekharan Pillai, M.K., Bilionis, I., 2025. Control-oriented dynamic computational modeling. Available at SSRN 5223568 .

Rollock, A.E., Klaus, D.M., 2025. Characterizing the impact of emergent technologies on earth communications reliance for crewed deep space missions. Acta Astronautica 226, 803–813.

Rollock, A.E., Klaus, D.M., 2026. Functional prioritization strategies and tradespace considerations for a self-reliant deep space habitat. Acta Astronautica .

Salmeron, M., Montoya, H., Silva, C., Dyke, S., 2024. Real-time temperature field state observer and digital replica to support cyber-physical testing. Experimental Techniques , 1–15.

Silva, C.E., Montoya, H., Dyke, S.J., Rhee, S., Mirfarah, M., Salmeron, M., Park, H., Chu, Z., Ma, J., Rajasekharan Pillai, M.K., et al., 2025. Development of a cyber–physical testbed for smart and resilient space habitats. ASCE OPEN: Multidisciplinary Journal of Civil Engineering 3, 04025007.

Sun, H., Duan, M., Wu, Y., Zeng, Y., Zhao, H., Wu, S., Lin, B., Yang, R., Tan, G., 2024. Designing sustainable built environments for mars habitation: integrating innovations in architecture, systems, and human well-being. Nexus .

The MathWorks, Inc., 2020. MATLAB and Simulink (R2020b). MathWorks. Natick, Massachusetts. URL: `https://www.mathworks.com/products/simulink.html`.

Uday, P., Marais, K., 2015. Designing resilient systems-of-systems: A survey of metrics, methods, and challenges. Systems Engineering 18, 491–510.

Valinia, A., Allen, J.R., Francisco, D.R., Minow, J.I., Pellish, J.A., Vera, A.H., 2022. Safe human expeditions beyond low Earth orbit (LEO). Technical Report NASA/TM-20220002905. NASA.

Wu, S.C., Parisi, M., McTigue, K., Panontin, T., Vera, A., 2021. Toward enabling safe earth-independent mission operations. International Association for the Advancement of Space Science .

Wu, S.C., Vera, A.H., 2019. Supporting Crew Autonomy in Deep Space Exploration: Preliminary Onboard Capability Requirements and Proposed Research Questions. Technical Report of the Autonomous Crew Operations Technical Interchange Meeting. Technical Report NASA/TM-2019-220345. NASA.

Wu, S.C., Vera, A.H., 2020. Capability considerations for enhancing safety on long duration crewed missions: Insights from a technical interchange meeting on autonomous crew operations. Journal of Space Safety Engineering 7, 78–82.