EPFL

National Aeronautics and Space Administration

ICRA 2026 – Workshop on Intelligent Space Robotics

# Gremlin

## AI-Based Adversarial Stress-Testing of Autonomous Systems

Marie Ethvignot[1,2], Hiro Ono[2], Richard Rieber[2]

[1]Swiss Federal Institute of Technology in Lausanne (EPFL)
[2]Jet Propulsion Laboratory, California Institute of Technology

## The Challenge

**As robotic space exploration extends to the outer Solar System, environments are poorly characterized before arrival and modern spacecraft are increasingly autonomous.**

**Current verification and validation (V&V) methods still depends on engineers designing handcrafted test scenarios from requirements. This approach is limited by human imagination and cannot anticipate the full range of possible behaviors.**

**As autonomy grows, the number of possible system states increases exponentially, making exhaustive testing infeasible. We need AI-driven methods capable of automatically exploring this vast state space and identifying the regions of highest uncertainty and risk.**

## The Adversarial Gremlin

Gremlin helps exposes **off-nominal behaviors** in closed-loop simulations. It acts as an opponent inside the simulator, reacting to the autonomy under test at each decision point.

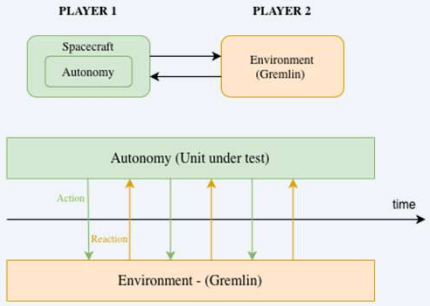


*Fig. 1 — Interaction loop between Gremlin and the Autonomy under Test*

Gremlin uses **Monte Carlo Tree Search** (MCTS) to efficiently explore the space of possible disturbances, guiding the search toward scenarios that degrade mission performance.

## Contributions

**1. GP-correlated continuous disturbances**

**2. The "insanity budget"**

## Case Study: UOP Descent

**Uranus Orbiter & Probe (UOP) [1]** is the next NASA Flagship mission (2023-2032 Decadal Survey). Evaluated using **MuSCAT [2]**, JPL's open-source mission simulator. Descent event triggers depend directly on atmospheric density.

## Disturbance Model & Sampling

**Atmospheric density** is perturbed multiplicatively in log-space via a **Gaussian-process prior**. The GP encodes correlation across altitudes. The **insanity budget** keeps the search inside a constant-probability ellipsoid.

$$\rho(h) = \rho_{nom}(h)\cdot\exp(g(\tau)), \quad g(\tau) \sim GP(0, \sigma^2 k(\tau,\tau')), \quad d^T K^{-1} d \le C$$

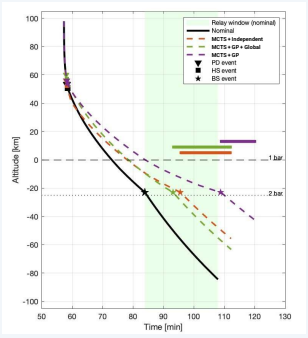


*Fig. 2 — Worst-case descent trajectory per sampling strategy, with relay windows shown as horizontal bars.*

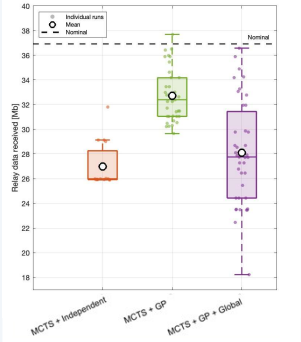


*Fig. 3 — Relay data received across sampling strategies*

**Three sampling strategies :**

**MCTS + Indep.**
**−27.0% (26.97 Mb)**
Independent sampling per event → no correlation. Adversarial but physically unrealistic.

**MCTS + GP**
**−25.1% (27.68 Mb)**
GP prior adds inter-event correlation. No global constraint → can drift to implausible sequences.

**MCTS + GP + Global ★ proposed**
**−11.9% (32.56 Mb)**
**GP + insanity budget. Adversarial AND statistically plausible under the prior.**

## Mission Impact

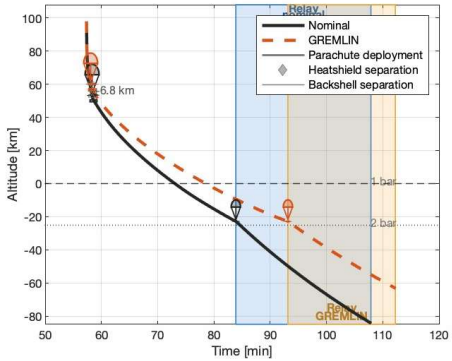


*Fig. 4 — Nominal vs. Gremlin descent trajectory (solid vs. dashed, relay windows shaded)*

**Key mission metrics (MCTS + GP + Global, N_iter = 20)**

| Metric | Nominal | Gremlin |
|---|---|---|
| Relay data [Mb] | 36.94 | 29.65 (−19.7%) |
| Relay duration [min] | 24.0 | 19.1 (−4.9 min) |

## Conclusion & Future Work

Gremlin applies to any autonomous system that can be simulated forward under uncertainty — not specific to atmospheric entry. The insanity budget is a tunable knob: tighten for statistically credible scenarios, loosen for deeper adversarial exploration.

**Limitations:** each rollout requires a full simulation propagation (30 s nominal → ~90 min for N_iter=20). The 19.7% data loss is also a lower bound, a higher-fidelity link model would amplify the impact.

**Future work:** surrogate modeling to reduce compute cost, extensions to multiple simultaneous uncertainty sources, stochastic autonomy policies, learnable GP hyperparameters.

**References:**
[1] UOP Concept Study, https://science.nasa.gov/wp-content/uploads/2023/10/uranus-orbiter-and-probe.pdf
[2] Bandyopadhyay et al., MuSCAT, AIAA ASCEND 2024