Correlation-Guided Fuzz Testing for Aviation GNSS Receiver

Nina Haag
Collins Aerospace & Fédération ENAC ISAE-SUPAERO ONERA
Université de Toulouse, France

Daniel Prun
Fédération ENAC ISAE-SUPAERO ONERA
Université de Toulouse, France

Antoine Blais
Fédération ENAC ISAE-SUPAERO ONERA
Université de Toulouse, France

Christophe Ouzeau
Collins Aerospace
Blagnac, France

Abstract

Aviation GNSS receivers are critical for safety but challenging to test due to nonlinear behaviors, environmental variability, and timing-dependent faults. This paper presents a correlation-guided fuzz testing framework that integrates sensitivity analysis with hybrid search-based workflows, enabling systematic and interpretable robustness evaluation. Unlike prior approaches focusing on isolated modules or simulations, the framework targets system-level, real-time behavior of hardware-integrated receivers. Step changes, ramps, and transient peaks are prioritized as indicators of anomalies, including spoofing and jamming events. Initial results demonstrate the feasibility of focusing fuzz testing on high-impact inputs to improve efficiency and interpretability. Future work will validate the framework in larger-scale campaigns, including scenarios affected by the South Atlantic Anomaly and controlled jamming or spoofing, providing a scientifically rigorous yet practically usable tool for safety-critical GNSS systems.

CCS Concepts

• **Software and its engineering** → **Software testing and debugging**; • **Hardware** → *Embedded and cyber-physical systems*; • **General and reference** → *Evaluation.*

Keywords

Fuzz Testing (Fuzzing),Search-Based Software Testing (SBST), GNSS Receivers, Avionics,Safety-Critical Systems,Hardware-in-the-Loop (HIL) Testing, Sensitivity Analysis, Hybrid Testing Framework

ACM Reference Format:
Nina Haag, Daniel Prun, Antoine Blais, and Christophe Ouzeau. 2026. Correlation-Guided Fuzz Testing for Aviation GNSS Receiver. In *19th Search-Based and Fuzz Testing Workshop (SBFT '26), April 12–18, 2026, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3786155.3788583

1 Introduction

Modern aviation relies on Global Navigation Satellite System (GNSS) receivers, including GPS, Galileo, and BeiDou, for precise positioning, navigation, and timing. Despite their critical role, these systems remain vulnerable to jamming and spoofing,[1] threats that can compromise navigation safety and operational integrity. Efficient test of these receivers is challenging because faults often emerge only under rare or extreme input conditions, which are costly to identify and to simulate. An exhaustive exploration of the input domain remains impractical for real-time, hardware-in-the-loop systems

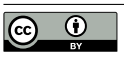
To address these challenges, Hybrid testing strategies have advanced state-of-the-art robustness evaluation. Search-Based Software Testing (SBST) explores high-dimensional input spaces using metaheuristic optimization, such as genetic algorithms or particle swarm optimization [2, 4]. Fuzz testing complements this approach by generating randomized or malformed inputs to uncover unexpected behaviors [8], with tools like AFL [7], libFuzzer [6], and GNSS-specific extensions such as FANDANGO [5]. However, a challenge of fuzz testing is to master the costs associated with generating and processing a large set of tests

This paper proposes a five stages conceptual framework mixing Fuzz testing and hybrid search-based testing into a unique workflow. By combining sensitivity-guided input prioritization with correlation informed fuzzing, it enables focused exploration of high-impact inputs while remaining feasible for routine engineering validation. Hybrid strategies such as SBST provide a natural foundation for Stage 1 and 2 exploratory fuzzing, guiding initial input sampling before correlation analysis. Specific output deviations—step changes, ramps, and transient peaks—are prioritized because they strongly indicate anomalies caused by GNSS spoofing or jamming [1]. Stages 3 and 4 cover targeted fuzzing and statistical change detection, focusing on high-impact inputs identified in earlier stages, while Stage 5 provides an iterative feedback loop to refine prioritization and steer subsequent fuzzing cycles.

The main contributions of this work are:

- A five-stage, correlation-guided fuzz testing workflow for aviation GNSS receivers.
- Integration of sensitivity analysis with hybrid search-based testing techniques.
- An initial hardware-in-the-loop demonstration of framework feasibility.

The proposed approach is evaluated through a pilot study using hardware-in-the-loop tests to illustrate feasibility and the practical application of sensitivity-informed fuzzing.



SBFT '26, Rio de Janeiro, Brazil

ACM ISBN 979-8-4007-2387-2/2026/04
https://doi.org/10.1145/3786155.3788583

[1] A *jamming attack* interferes with GNSS signals, preventing accurate reception, while a *spoofing attack* transmits counterfeit signals to mislead the receiver into computing incorrect positions or timing.

2 Challenges and Research Objectives

Aviation GNSS receivers process many inputs, including satellite geometry, signal distortions, atmospheric effects, multipath, interference, and internal settings. Testing these systems is challenging due to several factors:

- **Complex input–output behavior:** Small input changes can cause large or subtle variations, and stochastic internal processing may lead to non-deterministic outputs.
- **Time-consuming hardware-in-the-loop tests:** Each run requires significant setup time (15 minutes) and execution, limiting feasible experiments.
- **Rare or boundary-case scenarios:** Standard tests often miss extreme or low-probability input combinations that could trigger unsafe receiver behavior.

These challenges motivate the following **research questions**:

(1) Which input dimensions most strongly influence critical receiver outputs under nominal and perturbed conditions?
(2) How can anomalies caused by jamming or spoofing be systematically detected and quantified in high-dimensional, stochastic systems?
(3) How can testing remain feasible and interpretable in hardware-in-the-loop setups with large input spaces?

To address these questions, the overall research objective is to develop a structured, sensitivity-informed testing approach that systematically prioritizes influential inputs and detects critical output deviations, while remaining operationally feasible.

3 Correlation-Guided Fuzz Testing Workflow

This section presents a five-stage conceptual workflow for correlation-guided fuzz testing of aviation GNSS receivers (Figure 1), highlighting how high-impact inputs are identified, fuzzed, and systematically analyzed to detect anomalies. The framework builds upon prior empirical investigations [3] and defines a research agenda for hybrid, sensitivity-informed robustness testing rather than a fully implemented prototype.

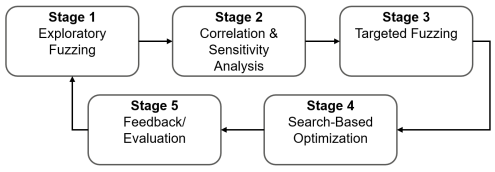


Figure 1: High-level conceptual workflow for correlation-guided fuzz testing.

Stage 1: Broad Fuzzing / Exploratory Sampling. A reference scenario is first obtained by running the receiver under nominal, unperturbed GNSS signals without any fuzzed or altered inputs. This reference run serves as the baseline for evaluating all subsequent tests. After establishing this baseline, a diverse set of test inputs is conceptually generated. In this work, inputs refer to controllable signal or configuration parameters—such as satellite geometry, signal distortions, atmospheric effects, multipath, interference power, and additional corrective message fields. The receiver responses to these inputs produce the monitored outputs, including position error, protection levels, timing deviations, signal integrity indicators, and navigation state changes. This exploratory stage maps how variations in the input parameters influence the receiver outputs and defines the input–output landscape for subsequent correlation analysis.

Stage 2: Correlation & Sensitivity Analysis. In this stage, relationships between input dimensions (controllable signal or configuration parameters) and output dimensions (observable receiver behaviors) are analyzed to identify high-impact inputs and determine which outputs warrant detailed statistical monitoring.

Let Y_i^{ref} denote the reference value of output dimension i under nominal (unperturbed) inputs, and $Y_i^{\text{test}(j)}$ denote the observed value of output dimension i when input dimension j is perturbed. Deviations are normalized by the corresponding reference value to prevent outputs with larger absolute magnitudes from dominating the sensitivity weight, and the absolute deviation between these values quantifies the effect of input j on output i:

$$\Delta_i^{(j)} = \frac{\left| Y_i^{\text{test}(j)} - Y_i^{\text{ref}} \right|}{|Y_i^{\text{ref}}| + \epsilon}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n \tag{1}$$

Here, $i = 1, \ldots, m$ indexes the monitored output dimensions, $j = 1, \ldots, n$ indexes input dimensions, and $\epsilon > 0$ is a small constant to avoid division by zero.

To summarize the influence of each input dimension, a *sensitivity weight* w_j is defined by aggregating the normalized deviations across outputs:

$$w_j = \frac{1}{m} \sum_{i=1}^{m} \Delta_i^{(j)}, \quad j = 1, \ldots, n \tag{2}$$

Sensitivity weights w_j quantify the influence of each input dimension on outputs: higher weights indicate stronger impact and guide Stage 3 targeted fuzzing, while low-weight dimensions may still contribute to rare anomalies. Occasional broader sampling can mitigate the risk of deprioritizing potentially critical inputs.

The sensitivity weights serve three purposes:

- Prioritizing high-impact input dimensions for targeted fuzzing (Stage 3),
- Defining monitoring priorities and thresholds for statistical change detection (Stage 4), and
- Supporting interpretability and expert-driven validation.

Weights can be determined either by expert judgment using visual analysis of input–output relationships or by formal correlation metrics (e.g., Pearson, Spearman, mutual information, or Shapley values) to systematically capture linear and nonlinear dependencies.

Stage 3: Targeted Fuzzing. Stage 3 refines the search space by concentrating exclusively on the high-impact inputs identified in Stage 2. Using the previously established reference scenario as a baseline, these inputs are systematically perturbed within their prioritized ranges, while all remaining parameters are held constant. This focused design ensures that deviations in the receiver's behaviour can be attributed to specific influential inputs, preventing

dilution effects from low-impact dimensions. The resulting output traces are then forwarded to Stage 4 for statistical anomaly detection.

Stage 4: Statistical Change Detection. Outputs generated during targeted fuzzing in Stage 3 are analyzed to identify statistically significant deviations from nominal behaviour. Since each high-impact input is perturbed individually, anomalies can be attributed to specific input dimensions. A dedicated detection algorithm extracts three characteristic classes of deviations:

- **Steps** Persistent shifts in the mean, detected using a CUSUM formulation $S_t = \max(0, S_{t-1} + x_t - \mu_0 - k)$, with μ_0 denoting the nominal mean and k the drift parameter.
- **Ramps** Gradual linear trends, identified via sliding-window gradient estimation under monotonicity constraints.
- **Peaks** Localised transient excursions, detected through z-score thresholding or percentile-based outlier detection.

Each detected event is time-stamped and forwarded to Stage 5 for interpretation.

Stage 5: Feedback & Evaluation. This stage evaluates the anomalies identified in Stage 4 against the reference scenario obtained in Stage 1. Because defining a concrete "bug" in GNSS receivers is non-trivial, deviations that exhibit steps, ramps, or peaks in the fuzzed run—but are absent in the reference—serve as heuristic indicators of potentially undesirable behaviour. Final assessment, however, requires expert judgment.

Fuzzed and reference output traces are first aligned via interpolation for point-wise comparability. Quantitative measures such as mean difference, standard deviation, maximum deviation, and correlation provide summary characterisations of divergence, while event-level analyses retain the semantics of detected ramps, steps, and peaks. This structured process ensures that anomalies are interpreted within the context of sensitivity-guided prioritisation and statistically grounded detection, enabling a systematic evaluation of fuzzing effects.

4 Test Setup and Early Results

To demonstrate the feasibility of the proposed framework, we conducted a preliminary hardware-in-the-loop (HIL) pilot study. The results presented here are illustrative and do not represent a full-scale evaluation. The aim is to verify the workflow, explore sensitivity trends, and highlight how targeted fuzzing and statistical analysis can detect high-impact input–output effects, which was demonstrated in a controlled HIL environment through a workflow capable of revealing sensitivity-driven dependencies.

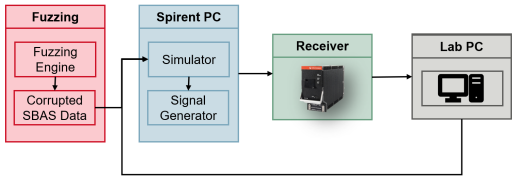


Figure 2: Hardware-in-the-loop (HIL) test environment for GNSS receiver fuzz testing.

The setup shown in (Figure 2) uses a Spirent GSS-9000 GNSS signal simulator with PosApp (v10), which produces realistic GPS constellations[2] augmented with SBAS (Satellite-Based Augmentation System)[3]. During testing, only SBAS message fields were varied, while GPS navigation and satellite geometry remained fixed. This controlled setup allowed precise observation of how individual inputs affect receiver behavior.

Testing followed a single-parameter-at-a-time methodology. Each SBAS field was individually varied across approximately 100–300 controlled input values, with 3–5 repetitions per input, corresponding to 27–70 hours of testing per field and roughly 900 hours in total. This *field-level fuzzing* approach, implemented via a custom in-house tool, ensures inputs remain syntactically valid and GNSS signals realistic, allowing observed output changes to be attributed directly to specific inputs. While simulation or record-and-playback environments can reduce logistical and environmental overhead compared to full HIL testing, they do not eliminate the receiver-internal initialization time (due to initial ephemeris decoding).

Table 1: Illustrative sensitivity scores for selected SBAS parameters (0 = No Effect, 10 = Very High Effect). Conceptual data adapted from [3].

SBAS Parameter	VPL/HPL	PDOP	NavState	w_j
MT9_URA	10	5	8	8
MT6_UDREI	9	2	7	6

Table 1 summarizes sensitivity scores for two selected SBAS parameters, based on previously measured GNSS receiver behavior [3]. These scores quantify how controllable signal, timing, and environmental inputs influence key outputs such as tracking stability, synchronization, and positioning error. In particular:

- **VPL and HPL (Vertical and Horizontal Protection Levels)** indicate the maximum expected positioning error.
- **PDOP (Position Dilution of Precision)** reflects the effect of satellite geometry on solution accuracy.
- **NavState** describes the receiver's navigation condition, including fix quality or signal lock.
- **Sensitivity weight** w_j (cf. Stage 2) identifies high-impact inputs for prioritization.

Among the tested inputs, MT9_URA[4] exhibited the highest sensitivity. This indicates that perturbations in this parameter produce the largest deviations in outputs, making it a prime candidate for Stage 3 targeted fuzzing. Internal validity was ensured by repeating each test and comparing deviations against baseline measurements.

Figure 3 illustrates the application of Stage 4 statistical change detection. Peaks, ramps, and step changes in outputs are automatically identified, providing a structured view of how fuzzed inputs affect receiver behavior. These detected events are then used in

[2]Up to 10 satellites broadcast GPS signals, including ranging signals (used to measure distances to the satellites) and navigation messages (used to calculate satellite positions and provide timing and status information).
[3]SBAS improves GNSS positioning accuracy and reliability by broadcasting differential correction messages that overlay the original GNSS signals.
[4]Message Type 9 User Range Accuracy (MT9_URA) represents an upper bound on satellite ranging error over the coverage area.

Stage 5 feedback, enabling systematic comparison of fuzzed scenarios against the reference and supporting interpretable, input-specific analysis.

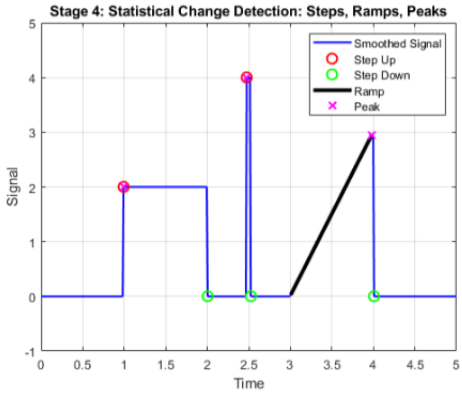


Figure 3: Example of statistical detection of anomalies in receiver outputs using Stage 4 methods.

5 Discussion, Benefits, and Challenges

The Correlation-Guided Fuzz Testing framework provides a structured approach to focus testing on the most influential input dimensions, improving efficiency, interpretability, and practical usability in safety-critical GNSS systems. By using correlation- and sensitivity-informed prioritization, the workflow reduces unnecessary test cases while remaining feasible for routine engineering validation. Its modular, staged design allows scalable exploration of complex input spaces and can incorporate advanced extensions, such as Shapley values or causal feature attribution. Practical implementation requires careful handling of hardware-in-the-loop setups and real-time constraints, as tests are time-consuming and outputs include both categorical and continuous measures. System stabilization delays further justify prioritizing high-impact inputs. Despite these challenges, targeted fuzzing and statistical monitoring have demonstrated relevance on individual cases, with broader empirical validation planned. While prior work has explored general-purpose fuzzing [6–8] and hybrid search-based testing [2], no existing approach combines sensitivity-guided input prioritization with structured, hardware-in-the-loop GNSS testing. The proposed workflow leverages sensitivity weights to systematically guide targeted fuzzing, improving anomaly detection efficiency, interpretability, and operational feasibility. This makes it relevant for safety-critical cyber-physical systems, including autonomous vehicles and industrial controllers, where complex input–output dependencies require thorough validation.

6 Planned Evaluation and Future Work

Future work will evaluate the framework in larger-scale testing campaigns to assess feasibility and practical utility, including hardware-in-the-loop setups to determine whether correlation-guided prioritization of high-impact inputs improves fuzz testing efficiency and reveals anomalies.

Scenarios will include effects of the South Atlantic Anomaly (SAA) on receiver behavior, as well as controlled real-world jamming and spoofing tests, to assess performance under realistic, safety-critical conditions and identify detectable anomalies.

The framework will also be extended with advanced sensitivity estimators and feature attribution methods to capture complex input–output dependencies. Expert assessments will evaluate interpretability and practical applicability. Together, these efforts aim to validate the framework as a scientifically rigorous yet practically usable tool for systematic, feedback-driven robustness evaluation in GNSS and other safety-critical systems.

7 Conclusion

This paper introduced a correlation-guided fuzz testing framework for aviation GNSS receivers that links sensitivity-informed input prioritization with search-based testing. By directing test effort toward statistically influential inputs and monitoring characteristic deviations such as steps, ramps, and transient peaks, the framework offers a structured and interpretable method for identifying anomalies relevant to spoofing, jamming, and integrity faults.

The approach aligns with the practical realities of GNSS validation—long stabilization times, hardware-in-the-loop constraints, and heterogeneous output types—making targeted prioritization both methodologically justified and operationally necessary. In doing so, the framework provides a principled pathway to more efficient and interpretable robustness testing, with particular relevance for safety-critical cyber-physical systems that exhibit complex input–output dependencies.

References

[1] Jiaqi Bi, Jiang Liu, Baigen Cai, and Jian Wang. 2024. Spoofing attack recognition for GNSS-based train positioning using a BO-LightGBM method. *Science Progress* 107, 4 (2024), 00368504241272731. arXiv:https://doi.org/10.1177/00368504241272731 doi:10.1177/00368504241272731 PMID: 39351631.

[2] Giovani Guizzo and Sebastiano Panichella. 2023. Fuzzing vs SBST: Intersections & Differences. *SIGSOFT Softw. Eng. Notes* 48, 1 (Jan. 2023), 105–107. doi:10.1145/3573074.3573102

[3] Nina Haag, Lotfi Fejri, Christophe Ouzeau, Daniel Prun, and Antoine Blais. 2025. Enhancing GNSS Receiver Resilience through Fuzz Testing: A Novel Approach to System Robustness in Avionics. In *Proceedings of the 44th Digital Avionics Systems Conference (DASC)*. IEEE/AIAA, Toulouse, France. Presented at the 44th IEEE/AIAA Digital Avionics Systems Conference (DASC), 2025.

[4] Mark Harman, S. Afshin Mansouri, and Yuanyuan Zhang. 2012. Search-based software engineering: Trends, techniques and applications. *ACM Comput. Surv.* 45, 1, Article 11 (Dec. 2012), 61 pages. doi:10.1145/2379776.2379787

[5] Stephan Neuhaus, José Antonio Zamudio Amaya, and Andreas Zeller. 2025. Personalized Fuzzing: A Case Study with the FANDANGO Fuzzer on a GNSS Module (Short Paper). In *Proceedings of the International Symposium on Software Testing and Analysis (ISSTA) — Fuzzing and Search-Based Testing Session.* CISPA Helmholtz Center for Information Security, Saarbrücken.

[6] Kostya Serebryany. 2015. Continuous Fuzzing with libFuzzer. In *Proceedings of the 2015 IEEE Cybersecurity Development Conference (SecDev)*. IEEE. doi:10.1109/SecDev.2016.10

[7] Michał Zalewski. 2014. American Fuzzy Lop (AFL). Online project documentation. https://lcamtuf.coredump.cx/afl/.

[8] Andreas Zeller, Rahul Gopinath, Marcel Böhme, Gordon Fraser, Christian Holler, Stefan P. Reiss, Rohan Padhye, and Abhik Roychoudhury. 2021. Fuzzing: Challenges and Reflections. *IEEE Software* 38, 3 (2021), 48–57. doi:10.1109/MS.2020.3045803