

Limits of Physics-Based GNSS Fault Detection in Multirotor UAV Simulation

Nitish Pendyala

South Brunswick High School
Monmouth Junction, New Jersey, USA
nitishpen007@gmail.com

Braiden Martin

South Brunswick High School
Dayton, New Jersey, USA
braidenmartin.r@gmail.com

Abstract—A false Global Navigation Satellite System (GNSS) trajectory does not always look physically impossible. A slow horizontal offset can preserve smooth position, velocity, acceleration, and jerk, making it difficult to reject with a limited onboard sensor set. This simulation study compares five causal monitors for a multirotor UAV: GNSS-acceleration disagreement, Kalman normalized innovation squared (NIS), GNSS-barometer altitude consistency, a fixed physics score, and a covariance-normalized physics score. Thresholds were fitted on 300 nominal and 300 uncertainty-sampled clean missions, then frozen before 7,480 evaluation missions. On a predefined 1,800-mission panel, covariance physics detected 82.9% and fixed physics 78.3%; on a separately generated 600-mission holdout, their rates fell to 41.5% and 34.3%, while Kalman NIS reached 42.0%. In 400 adaptive slow-horizontal missions, Kalman NIS detected 9 and the other four monitors detected none. Nominal physics calibration also produced 40.0–41.6% false-alarm rates under clean model mismatch. Broader calibration removed those false alarms but also removed the observed physics detections. The results show that detectability depends on the independent measurements available and on model uncertainty. No radio-frequency attack, receiver manipulation, or flight hardware is modeled.

Index Terms—GNSS integrity monitoring, multirotor UAV, fault detection, Kalman innovation, physics consistency, model mismatch

I. INTRODUCTION

GNSS supplies an absolute position and velocity reference for small uncrewed aircraft, but corrupted measurements can pull a navigation solution away from the true trajectory. Controlled experiments have shown that captured GNSS solutions can affect aircraft control [1]. Existing defenses operate at several levels. Receiver-level methods use signal structure, correlation, or other radio-frequency observables [2], [3]; navigation-level methods compare GNSS against inertial sensing, vehicle motion, carrier phase, or innovation statistics [4]–[11]. Additional external measurements can further improve integrity monitoring [12].

The difficult cases are not large jumps. A smoothly introduced horizontal offset may remain consistent with feasible motion, while a smaller vertical error can conflict with an independent barometer. Physics-based checks create a second problem: a clean flight can appear faulty when the assumed mass, drag, thrust scale, or wind differs from the vehicle being modeled.

This work asks a narrow question: under the simulated sensor and uncertainty models, which GNSS faults can the tested

monitors detect, and where do they fail? The contribution is not a new Kalman filter or a claim of universal spoofing detection. It is a controlled comparison under one causal detector interface, including a separate holdout, a slow-horizontal stress test, timing and benign controls, and explicit model-mismatch trials. Only synthetic GNSS position and velocity measurements are altered.

II. METHODS

A. Simulation and detector inputs

The simulator uses a local east–north–up frame and a 58-s mission sampled every 0.2 s. A smooth reference trajectory is perturbed by stochastic acceleration and integrated consistently,

$$\mathbf{v}_{k+1} = \mathbf{v}_k + \mathbf{a}_k \Delta t, \quad (1)$$

$$\mathbf{p}_{k+1} = \mathbf{p}_k + \mathbf{v}_k \Delta t + \frac{1}{2} \mathbf{a}_k \Delta t^2. \quad (2)$$

The nominal model uses a 2.2-kg vehicle, 28° maximum tilt, 42-N maximum thrust, 16-m/s horizontal speed limit, 4-m/s climb and descent limits, and 4-m/s³ jerk limit. Nominal sensor standard deviations are 0.55 m for GNSS position, 0.16 m/s for GNSS velocity, 0.08 m/s² for processed navigation-frame acceleration, and 0.35 m for barometric altitude. These are selected simulation parameters, not measurements from a specific aircraft.

A model-derived thrust proxy T and measured thrust-axis direction $\mathbf{b}_3$ give

$$\mathbf{a}_{\text{phys}} = -g\mathbf{e}_3 + \frac{T}{m}\mathbf{b}_3 - \mathbf{D}(\mathbf{v}_{\text{GNSS}} - \hat{\mathbf{w}}). \quad (3)$$

The model omits closed-loop six-degree-of-freedom dynamics, actuator transients, and rotor aerodynamics. Each GNSS update carries a physical sample time, reported timestamp, arrival time, validity, age, and source index. The detector receives only the timestamped channels shown in Fig. 1. Invalid samples cannot advance alarm persistence, and the Kalman filter rejects out-of-sequence updates.

B. Five detector scores

All derivative-based residuals use a causal first-order low-pass filter with time constant 0.8 s. Let $\bar{\mathbf{v}}_k^G$ be filtered

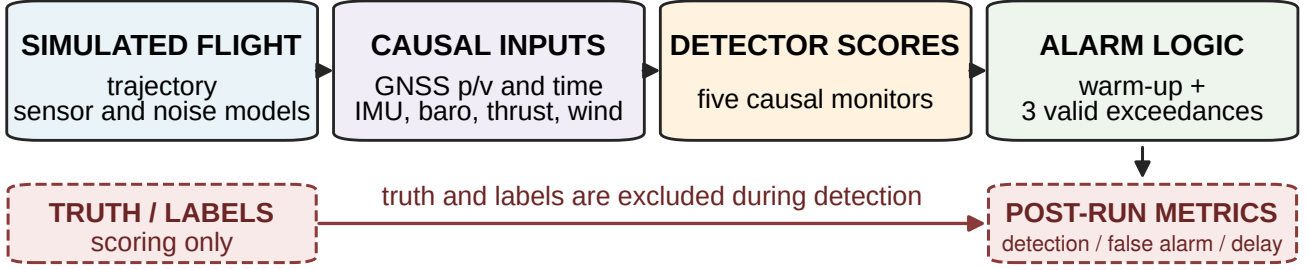


Fig. 1. Causal processing and scoring boundary. Truth states and fault labels are excluded from all detector inputs and are used only after a mission to label detections, false alarms, and delays.

GNSS velocity and $\bar{\mathbf{a}}_k^I$ filtered processed acceleration. The disagreement score is

$$r_d(k) = \left\| \frac{\bar{\mathbf{v}}_k^G - \bar{\mathbf{v}}_{k-1}^G}{t_k - t_{k-1}} - \bar{\mathbf{a}}_k^I \right\|_2. \quad (4)$$

This is a causal acceleration comparison; it is not an integrated-window residual.

The Kalman baseline has state $\mathbf{x} = [\mathbf{p}^T \mathbf{v}^T]^T$. For accepted measurement interval δt ,

$$\mathbf{F} = \begin{bmatrix} \mathbf{I} & \delta t \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \frac{1}{2} \delta t^2 \mathbf{I} \\ \delta t \mathbf{I} \end{bmatrix}, \quad (5)$$

with processed acceleration as the input. The initial covariance is $10^3 \mathbf{I}$. The process covariance uses a white-acceleration scale of 0.02 m/s^2 , and the measurement covariance is diagonal with the stated GNSS position and velocity variances. The score is $\text{NIS}_k = \boldsymbol{\nu}_k^T \mathbf{S}_k^{-1} \boldsymbol{\nu}_k$, evaluated with a Cholesky solve [13], [14]. The altitude score is $r_h = |z_{\text{GNSS}} - z_{\text{baro}}|/1.4$.

The fixed physics score is

$$r_f = \sqrt{(r_a/0.65)^2 + (r_z/1.4)^2 + r_e^2}, \quad (6)$$

where $r_a = \|\mathbf{a}_G - \mathbf{a}_{\text{phys}}\|_2$, r_z is the unscaled altitude difference, and r_e is the largest normalized violation of the speed, climb, acceleration, or jerk limits. The covariance score does not reuse r_e . It forms $\mathbf{z} = [r_a/\sigma_a \ r_z/\sigma_z]^T$ from first-order uncertainty estimates and evaluates

$$r_c = \sqrt{(\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu})}, \quad (7)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are fitted only from clean calibration residuals.

C. Calibration, faults, and analysis

Each calibration regime uses 300 clean missions. Nominal calibration keeps the baseline parameters fixed. Distributional calibration samples mass scale from 0.8–1.2, drag scale from 0.6–1.4, thrust scale from 0.85–1.15, wind-estimation bias with 0.8-m/s component standard deviation, wind variation from 0.2–2.5 m/s, attitude noise from 0.2° – 1.5° , GNSS position noise from 0.35–0.90 m, GNSS velocity noise from 0.10–0.28 m/s, acceleration noise from 0.04–0.16 m/s^2 , barometer noise from 0.20–0.70 m, latency from 0–0.4 s, packet loss from 0–8%, and timestamp jitter from 0–0.05 s. For each score, the

threshold is the pooled 0.997 quantile of valid post-warm-up clean samples. Fault outcomes are not used for threshold or covariance fitting.

A derivative-consistent fault with direction $\mathbf{d}$, magnitude M , start t_0 , and duration T uses $q(u) = 10u^3 - 15u^4 + 6u^5$ for $u \in [0, 1]$:

$$\Delta \mathbf{p} = M q(u) \mathbf{d}, \quad \Delta \mathbf{v} = \frac{M}{T} q'(u) \mathbf{d}. \quad (8)$$

Position-only and velocity-only variants suppress one of these two terms. The frozen campaign contained 600 calibration missions and 7,480 evaluation missions: 2,400 in the representative and holdout sets, 880 in the controlled and adaptive horizontal tests, 600 in the mechanism studies, and 3,600 in the benign, timing, and model-mismatch controls. The 8,080 mission seeds were unique, with no seed reuse across calibration or evaluation partitions.

After a 4-s warm-up, an alarm requires three consecutive valid score samples above the frozen threshold. The mission, not the time sample, is the statistical unit. Rates use pooled mission counts and 95% Wilson intervals [15]. Planned detector comparisons use mission-paired bootstrap intervals [16] and exact McNemar tests [17], with Holm adjustment over six comparisons [18]. Detection delay is summarized only when a mission is detected.

III. RESULTS

A. Representative panel and separate holdout

Figure 2 shows a large generalization gap. On the 1,800-mission representative panel, covariance physics detected 1,493 missions (82.9%), fixed physics 1,409 (78.3%), Kalman NIS 1,172 (65.1%), altitude-only 795 (44.2%), and disagreement 746 (41.4%). On the separately generated 600-mission holdout, the best result was Kalman NIS at 252/600 (42.0%), followed by covariance physics at 249/600 (41.5%). Fixed physics fell to 206/600 (34.3%), disagreement to 180/600 (30.0%), and altitude-only to 176/600 (29.3%). The holdout was not used to fit thresholds, covariance terms, or fault parameters.

On the representative panel, fixed physics exceeded disagreement by 36.83 percentage points (paired-bootstrap 95% interval 34.06–39.56), fixed physics exceeded Kalman NIS by 13.17 points (10.56–15.83), and covariance physics exceeded fixed

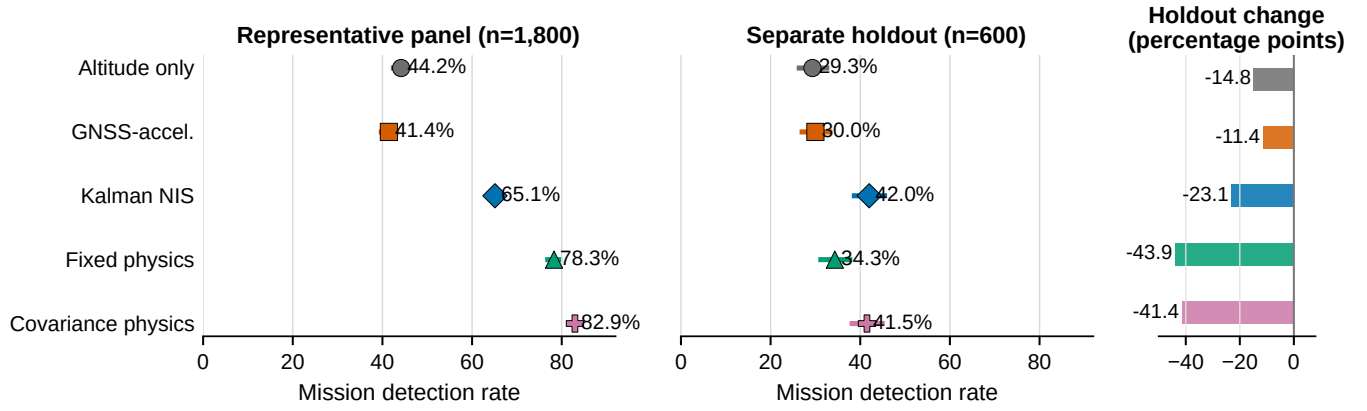


Fig. 2. Nominal-calibration detection rates on the predefined representative panel and a separately generated holdout. Points show pooled mission rates; horizontal bars show 95% Wilson intervals. The right panel reports the absolute change and does not imply paired missions between the two sets.

physics by 4.67 points (3.67–5.67). These within-panel gains do not resolve the weaker holdout performance.

B. Smooth horizontal faults

The controlled grid in Fig. 3 holds nuisance factors fixed while varying only final displacement and duration. Short, sharp offsets were consistently detected. Spreading the same displacement over 20–40 s usually reduced the added velocity and acceleration enough to avoid detection. Across all 480 grid missions, Kalman NIS detected 188 (39.2%), covariance physics 131 (27.3%), disagreement 131 (27.3%), and fixed physics 93 (19.4%); altitude-only detected none.

The adaptive slow-horizontal panel was harder. Kalman NIS detected 9/400 missions (2.25%, 95% CI 1.19–4.22%); the other four monitors detected 0/400, giving a 95% Wilson upper bound of 0.95% for each. This result is limited to the modeled sensor set and fault distribution. It does not establish that smooth horizontal corruption is impossible to detect with other sensors or architectures.

C. Independent constraints, controls, and mismatch

The fault families show where the evidence came from. All monitors except altitude-only detected all 400 sudden representative missions. In 400 vertical missions, altitude-only, fixed physics, and covariance physics each detected 400, Kalman NIS detected 159, and disagreement detected none. In 400 mixed missions, the corresponding counts were 395, 396, 399, 211, and 14. Across the combined 800 vertical and mixed missions, altitude-only detected 795 and fixed physics 796; most of the benefit in those cases therefore came from the independent barometer rather than the broader dynamics model.

No nominal false alarms occurred for disagreement, altitude-only, fixed physics, or covariance physics in 600 benign missions; Kalman NIS false-alarmed in 14 (2.33%). In 1,000 timing-control missions, Kalman NIS produced 10 nominal-calibration false alarms, while the other four monitors produced none.

Model mismatch was the main weakness of the physics scores (Fig. 4). Under nominal calibration, fixed and covariance physics false-alarmed in 400/1,000 and 416/1,000 clean mismatch missions, but detected only 3/1,000 and 45/1,000 separate faulted mismatch missions. Distributional calibration reduced the observed physics false-alarm and detection counts to zero. Kalman NIS moved from 7 false alarms and 156 detections under nominal calibration to 1 false alarm and 67 detections under distributional calibration. Wider calibration improved robustness but reduced sensitivity.

IV. DISCUSSION AND LIMITATIONS

No detector performed best across every test block. Physics scores were strong when the corrupted GNSS solution contradicted an independent altitude measurement or implied abrupt, infeasible motion. They were much less reliable when the horizontal corruption remained smooth or when the model parameters were wrong. Covariance normalization improved representative-panel detection by 4.67 percentage points over the fixed score, but it did not prevent a large holdout drop and did not solve mismatch sensitivity.

Kalman NIS generalized better to the holdout and retained some response under mismatch, but it produced more benign and timing false alarms. The altitude-only result is also important: much of the apparent physics advantage in vertical cases came from a simple independent measurement. Better horizontal observability would require information not controlled by the corrupted GNSS channel, such as airspeed, visual odometry, external ranging, authenticated signals, richer receiver observables, or independently validated propulsion and aerodynamic models.

This is a reduced-order, open-loop numerical study. It does not model receiver tracking loops, multipath, clock behavior, radio-frequency signals, closed-loop flight dynamics, rotor transients, vibration, or aerodynamic cross-coupling. Processed navigation acceleration is not raw accelerometer specific force, and thrust is represented by a model-derived proxy. The vehicle values and uncertainty ranges were chosen for simulation and were not identified from a real airframe. Every reported rate

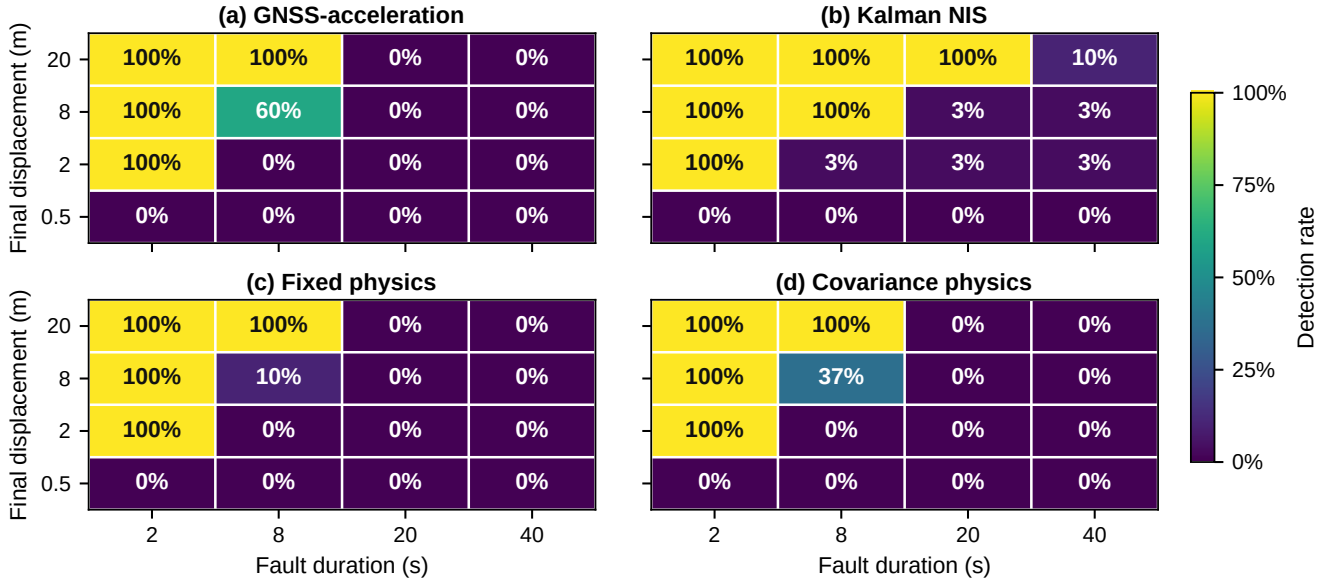


Fig. 3. Nominal-calibration detection rates for derivative-consistent along-track faults generated with a quintic smoothstep. Rows give final displacement, columns give fault duration, and each cell contains 30 missions under fixed nuisance factors. Altitude-only detected 0 of 480 missions and is omitted.

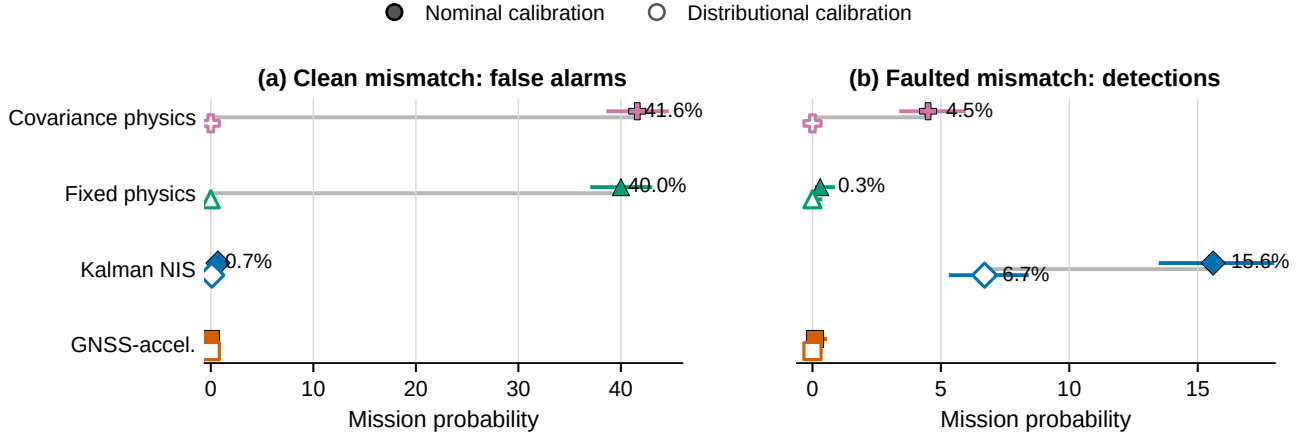


Fig. 4. Effect of calibration on model-mismatch trials. Filled markers are nominal thresholds and hollow markers are distributional thresholds; bars show 95% Wilson intervals. Clean and faulted panels use separate mission seeds. Lines connect calibration choices for the same detector, not clean-to-fault pairs.

is therefore conditional on the trajectory, sensor, fault, and uncertainty distributions used here.

A direct next step is playback or hardware-in-the-loop testing with synchronized flight-controller logs, independent position truth, GNSS, inertial and barometric data, attitude, rotor-speed or motor telemetry, and measured wind or airspeed. Model parameters and thresholds should be fitted on separate clean flights before the slow-horizontal and mismatch tests are repeated.

V. CONCLUSION

Reduced-order physics checks can detect synthetic GNSS faults when the corrupted motion conflicts with an independent barometer or exceeds clear kinematic limits. They should not be assumed to detect a smooth horizontal trajectory simply because that trajectory is false. Strong representative-panel results did not fully carry to a separate holdout, and nominal physics

thresholds were highly sensitive to model mismatch. Broader calibration removed the observed mismatch false alarms only by reducing fault sensitivity. In practice, the value of physics-based monitoring depends on which measurements remain independent of GNSS and how accurately vehicle uncertainty is represented.

DATA AND CODE AVAILABILITY

The frozen simulation code, calibration files, per-mission records, seed manifest, and figure-source tables used for this manuscript are available from the authors upon request. The complete package will be placed in a persistent public archive before final publication.

REFERENCES

- [1] A. J. Kerns, D. P. Shepard, J. A. Bhatti, and T. E. Humphreys, "Unmanned aircraft capture and control via GPS spoofing," *J. Field Robot.*, vol. 31, no. 4, pp. 617–636, 2014, doi: 10.1002/rob.21513.

- [2] M. L. Psiaki and T. E. Humphreys, "GNSS spoofing and detection," *Proc. IEEE*, vol. 104, no. 6, pp. 1258–1270, Jun. 2016, doi: 10.1109/JPROC.2016.2526658.
- [3] M. L. Psiaki, B. W. O'Hanlon, J. A. Bhatti, D. P. Shepard, and T. E. Humphreys, "GPS spoofing detection via dual-receiver correlation of military signals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 4, pp. 2250–2267, Oct. 2013, doi: 10.1109/TAES.2013.6621814.
- [4] S. Bhattacharyya and D. Gebre-Egziabher, "Kalman filter-based RAIM for GNSS receivers," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 3, pp. 2444–2459, Jul. 2015, doi: 10.1109/TAES.2015.130585.
- [5] C. Tanil, S. Khanafseh, and B. Pervan, "Detecting global navigation satellite system spoofing using inertial sensing of aircraft disturbance," *J. Guid., Control, Dyn.*, vol. 40, no. 8, pp. 2006–2016, 2017, doi: 10.2514/1.G002547.
- [6] C. Tanil, S. Khanafseh, M. Joerger, and B. Pervan, "An INS monitor to detect GNSS spoofers capable of tracking vehicle position," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 1, pp. 131–143, Feb. 2018, doi: 10.1109/TAES.2017.2739924.
- [7] Y. Liu, S. Li, Q. Fu, Z. Liu, and Q. Zhou, "Analysis of Kalman filter innovation-based GNSS spoofing detection method for INS/GNSS integrated navigation system," *IEEE Sensors J.*, vol. 19, no. 13, pp. 5167–5178, Jul. 2019, doi: 10.1109/JSEN.2019.2902178.
- [8] Z. Clements, J. E. Yoder, and T. E. Humphreys, "Carrier-phase and IMU based GNSS spoofing detection for ground vehicles," in *Proc. ION Int. Tech. Meeting*, 2022, pp. 83–95, doi: 10.33012/2022.18252.
- [9] Y. Ke, Z. Lv, C. Zhang, X. Deng, X. Shang, and W. Wu, "Tightly coupled GNSS/INS integration spoofing detection algorithm based on innovation rate optimization and robust estimation," *IEEE Access*, vol. 10, pp. 72444–72457, 2022, doi: 10.1109/ACCESS.2022.3186305.
- [10] Z. Zhou, H. Li, Z. Chen, and M. Lu, "Velocity consistency checking based GNSS spoofing detection method for vehicles," *IEEE Trans. Veh. Technol.*, vol. 73, no. 2, pp. 1974–1990, Feb. 2024, doi: 10.1109/TVT.2023.3321522.
- [11] C. Zhang, H. Wang, H. Hu, S. Wang, and H. Liu, "Acceleration consistency analysis based GNSS spoofing detection for autonomous vehicles," in *Proc. 3rd Int. Conf. Consumer Electron. Comput. Eng.*, 2023, pp. 66–72, doi: 10.1109/ICCECE58074.2023.10135427.
- [12] J. Khalife, M. Maaref, and Z. M. Kassas, "Opportunistic autonomous integrity monitoring for enhanced UAV safety," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 38, no. 5, pp. 34–44, May 2023, doi: 10.1109/MAES.2022.3178664.
- [13] A. Gelb, Ed., *Applied Optimal Estimation*. Cambridge, MA, USA: MIT Press, 1974.
- [14] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York, NY, USA: Wiley, 2001, doi: 10.1002/0471221279.
- [15] E. B. Wilson, "Probable inference, the law of succession, and statistical inference," *J. Amer. Stat. Assoc.*, vol. 22, no. 158, pp. 209–212, 1927, doi: 10.1080/01621459.1927.10502953.
- [16] B. Efron, "Bootstrap methods: Another look at the jackknife," *Ann. Stat.*, vol. 7, no. 1, pp. 1–26, 1979, doi: 10.1214/aos/11176344552.
- [17] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, pp. 153–157, 1947, doi: 10.1007/BF02295996.
- [18] S. Holm, "A simple sequentially rejective multiple test procedure," *Scand. J. Stat.*, vol. 6, no. 2, pp. 65–70, 1979.