GNSS Spoofing Detection through Metric Combinations: Calibration and Application of a general Framework

Fabian Rothmaier, Leila Taleghani, Yu-Hsuan Chen, Sherman Lo, Eric Phelts, Todd Walter, *Stanford University*

BIOGRAPHY

Fabian Rothmaier is a PhD candidate at the GPS Laboratory at Stanford University. He received his B. Engr. degree from the University of Applied Sciences Bremen, Germany in 2015 and his M. Sc. degree from Stanford University in 2017.

Sherman Lo is a senior research engineer at the Stanford GPS Laboratory. He received his Ph.D. in Aeronautics and Astronautics from Stanford University in 2002. He has and continues to work on navigation robustness and safety, often supporting the FAA. He has conducted research on Loran, alternative navigation, SBAS, ARAIM, GNSS for railways and automobile. He also works on spoof and interference mitigation for navigation. He has published over 100 research papers and articles.

Eric Phelts is a research engineer in the Department of Aeronautics and Astronautics at Stanford University. He received a Ph.D. in mechanical engineering from Stanford University in 2001. His research involves signal monitoring techniques and analysis for SBAS, GBAS, and ARAIM.

Todd Walter received his Ph.D. in Applied Physics from Stanford University in 1993. He is a Research Professor in the Department of Aeronautics and Astronautics at Stanford University. His research focuses on implementing high-integrity air navigation systems. He has received the ION Thurlow and Kepler awards. He is also a fellow of the ION and has served as its president.

ABSTRACT

This paper is an application example of a general framework to combine an arbitrary number of metrics for GNSS spoofing detection using the Generalized Likelihood Ratio Test (GLRT). The framework maximizes robustness to wide ranges of attack modes while guaranteeing a false alert probability under the Neyman-Pearson paradigm. In this paper we summarize the previously published GLRT framework and analyze its performance against conventional means to combine metrics through logical gates. For a combination of power and signal distortion metrics the worst case P_{MD} is reduced by 60%. We then calibrate measurement models for a receiver model based on flight data and validate compliance with the false alert guarantee for power and a signal distortion metric as well as pseudorange residuals. We apply the framework using the calibrated models to spoofing scenarios from the TEXBAT dataset processed through the same receiver type. We demonstrate highly robust detection behavior by leveraging the framework to combine the three metrics.

I. INTRODUCTION

GNSS spoofing is the altering of a receiver's navigation solution by GNSS-like signal radiated from a source different from the navigation satellites. It poses a dangerous threat, as it can - if undetected - introduce an arbitrarily large error to the victim's position, velocity and time (pvt) solution without the user's knowledge. If the user is an aircraft landing in instrument meteorological conditions, an autonomous vehicle or the time synchronization of the electrical grid, the implications of a wrong pvt solution can potentially be catastrophic.

Several incidents of spoofing attacks outside of an academic setting have been reported in recent years. Russia has for example been accused of developing significant GPS spoofing capabilities as part of its cyberwarfare strategy and as standard means of VIP and location protection [1], and boats in the world's busiest port in Shanghai have reported erroneous positions through their Automated Identification System (AIS) [2].

GNSS spoofing, its detection and mitigation has been researched in the public literature for at least two decades since the Volpe report [3] in 2001. Recent overviews of attack modes and common defense strategies are given by [4], [5] and [6]. Numerous approaches have been proposed to detect spoofing attacks. Every approach has its strengths and weaknesses, and no single observable alone can guarantee robust detection. Several techniques however show complementary characteristics, making a combination of defenses attractive.

This paper presents an application example of the general framework combining an arbitrary number of metrics for an optimal detection detailed in [7]. The cited paper contains several details we have to gloss over in this manuscript for brevity and is recommended to be read first or jointly. The framework focuses on techniques that can be implemented by the user segment and rely on GNSS signals alone. It excludes encryption-based defenses like the one presented in [8] and defenses based on drift monitoring with the help of inertial sensors such as in [9]. The goal is to support the most educated decision on the presence of a spoofing attack by the GNSS receiver itself and without changes to the GNSS signal structure.

For brevity, we will limit the analysis in this paper further to combinations of Signal Quality Monitoring (SQM) metrics like the ones presented in [10], [11], [12], [13], received power monitoring as in [14], [15], and pseudorange residual checks as employed in Receiver Autonomous Integrity Monitoring (RAIM) and originally suggested by [16]. [17, 18] leverage the complementary characteristic of power and SQM metrics in a Bayesian monitor to discern between nominal conditions, multipath, jamming and spoofing, [19] offers a combined monitor where the detection threshold is set empirically based on data recorded in flight and by WAAS stations. A more general approach to combine techniques is followed in [20] through the use of belief functions. [21] presents the compelling general concept of a PNT trust-inference engine, but a lot of work remains to be done until its final implementation and it is unclear how similar performance guarantees can be given as with the approach followed in this paper.

Specifically the contributions of this paper beyond the cited literature and the work in [7] are:

- A calibration of measurement models based on flight data to confirm the low false alarm probability guaranteed by the detection test. We present Gaussian over bounds for the various observables in a noise environment encountered in flight.
- An application of these exact measurement models to spoofing scenarios processed through the same receiver type, for a realistic analysis of the detection capability of the considered defenses. We achieve successful detection of all considered attacks before any navigation error is introduced.
- A deeper discussion and quantitative performance analysis of various defense combinations. We reiterate on the argument that the optimal defense choice depends on the expected attack parameters and present a robust and versatile combination. We point out challenges and future work to calibrate defenses.

The remainder of the paper is divided into three sections and a summary and conclusion. In Section II we summarize the detection framework presented in [7]. In Section III we analyze flight data to calibrate measurement models for the used receiver. In Section IV we finally apply the detection framework to spoofing scenarios using the previously calibrated models.

II. THE GLRT FRAMEWORK

The framework introduced in [7] and followed in this paper is based on the Generalized Likelihood Ratio Test (GLRT). This section summarizes how we use the GLRT for spoofing detection and gives a generic means of combining detection monitors through logical gates. The interested reader is referred to [7] for details. We will then take a look at a first performance comparison for a discussion of which combination is optimal.

1. The Solution to an Optimization Problem

We use the GLRT to solve a detection problem under the Neyman-Pearson paradigm [22]. The trademark of a Neyman-Pearson approach is that it is independent of prior probabilities. Specifically we want to minimize the probability of missed detections P_{MD} while satisfying a constraint on the maximum false alert probability $P_{FA_{max}}$. In this paper we define a false alert as an alert given nominal conditions. A missed detection as no alert under spoofed conditions. From here on out we refer to nominal conditions as the null hypothesis H_0. A missed detection is the event of not raising an alarm given spoofed conditions. Spoofed conditions are referred to as H_1, or H_h with $h > 1, 2, \dots$ if we want to differentiate between different specific threats. The detection problem can be phrased as an optimization problem.

$$\begin{aligned} \min_{\gamma} \quad & \max_{h} P(\log \Lambda(y) \geq \gamma \mid H_h) \\ s.t. \quad & P(\log \Lambda(y) < \gamma \mid H_0) \leq P_{FA_{max}} \end{aligned} \tag{1}$$

for the decision variable $\log \Lambda(y)$ given by

$$\log \Lambda(y) = \log \frac{\max_{\theta_0 \in \Omega_0} p(y \mid \theta_0)}{\max_{\theta_1 \in \Omega_1} p(y \mid \theta_1)} \tag{2}$$

for distribution parameters θ_i and the parameter space under the ith hypothesis Ω_i. The parameters θ_i represent unknown aspects about the distribution of the measurements y under either hypothesis, for example its mean.

The detection threshold γ is determined by the distribution under nominal conditions.

$$P_{FA_{max}} = \int_{-\infty}^{\gamma} p\left(\log \Lambda | H_0\right) d \log \Lambda \tag{3}$$

A solution of the P_{MD} requires an expression of $\log \Lambda$ under spoofed conditions H_h:

$$P_{MD,h} = \int_{\gamma}^{\infty} p(\log \Lambda | H_h) d \log \Lambda \tag{4}$$

For the most part of the paper we will refer to the spoofed hypothesis as H_1 to keep the notation light. We keep in mind that H_1 really comprises a range of different attack scenarios, until we specifically analyze the defense's performance against different threats in Section II.3.

An online monitor has to be able to solve at least for the threshold γ in real time, requiring a description of the probability distribution of $p(\log \Lambda | H_0)$. In [7] we show that many observables used for spoofing detection follow a common form: they are modeled with Normally distributed measurement noise with covariance Σ under both H_0 and H_1. The hypothesis differ only in the means of the distributions, hence $\theta_i = \mu_i$. This is a special case of the general Gaussian problem [22].

$$y | H_0 \sim N(\mu_0, \Sigma); \quad \mu_0 \in \Omega_0 \tag{5a}$$

$$y | H_1 \sim N(\mu_1, \Sigma); \quad \mu_1 \in \Omega_1 \tag{5b}$$

Three scenarios can be considered. Either, the mean under nominal conditions μ_0 is known, but μ_1 is unknown. This is the most common case and the GLRT turns into a simple vs. composite hypothesis test. It applies to the power measurement, many signal quality metrics and pseudorange residuals. Alternatively, μ_0 could be unknown but μ_1 is known, resulting in a composite vs. simple test. This scenario is phrased e.g. in [23] for a dual antenna setup. The most powerful test can be phrased when both μ_0 and μ_1 are known. In [24] we phrase several Direction of Arrival based techniques as this simple vs. simple test.

The metrics considered in this paper all belong to the first case. Depending on the metric, μ_0 is known by definition, obtained through calibration or parameter estimation prior to the spoofing detection test. Similarly, the covariance Σ is usually obtained by calibration during trusted conditions. In the meantime, we are unsure about the metric's mean under spoofed conditions (we for example don't know the pseudoranges and resulting residuals generated by a spoofer). For N measurements $\Omega_1 = \mathbb{R}^N$, μ_1 is generally set to its Maximum Likelihood Estimate (MLE). For a normal distribution the MLE of the mean is simply given by the measurement itself. We therefore have $\mu_1 = y$.

Substituting the probability density functions (pdf)s of the Normal distributions given in Eq. (5) with $\mu_1 = y$ into Eq. (2) results in

$$\log \Lambda(y) = -\frac{1}{2} \left(y - \mu_0\right)^T \Sigma^{-1} \left(y - \mu_0\right) \tag{6}$$

The GLRT essentially turns into a χ^2-test of how well the measurements match H_0. Under H_0, $-2 \log \Lambda_{SC}$ follows a χ^2 distribution with k degrees of freedom. Under H_1 it follows a noncentral χ^2 with noncentrality parameter λ.

$$-2 \log \Lambda \mid H_0 \sim \chi^2_k \tag{7a}$$

$$-2 \log \Lambda \mid H_1 \sim \chi^2_{k,\lambda} \tag{7b}$$

$k \leq N$, depending on the number of independent measurements in y. λ and therefore P_{MD} depend on the attack mode. With the distributions in Eq. (7) both γ and P_{MD} (for a specific threat) can be solved for in real time.

2. Combining metrics

The derivations so far have been rather general. Let us now finally consider measurements from M different metrics.

a) Separate Tests

A straight forward way to employ multiple metrics for spoofing detection is to process each metric in a separate detection test as the one defined in the previous section. The overall decision is then a combination of individual decisions based on logical gates. $P_{FA_{max}}$ needs then to be budgeted among the different monitors. For the sake of simplicity we will limit the derivations in this section to logical OR and AND gates between individual monitors, but the presented concepts can easily be expanded to other logical gates. OR gates are used to combine complementary metrics in [25], [26], [27], [28] or to test for multiple hypothesis in parallel as in (Advanced) RAIM [29]. AND gates are often used to represent specific knowledge or assumptions about characteristics of the attack as in [30].

Boolean AND and OR operators have the associative property, a sequence of operations can be broken down into a recursive series of operations with two arguments each. The simple statements considered here are therefore sufficient for combinations of any number of monitors. This concept is well known in the integrity community as a continuity fault tree [31], [32], [33].

We following brief statements are justified in detail in [7]. For the remainder of this paper we assume that measurements from different metrics can be considered mutually statistically independent given a hypothesis. This assumption is leveraged in most statements, with Eq. (8) being a notable exception. We will attempt to justify this assumption for flight data in Section III.

Given M monitors connected through OR gates, the P_{FA} budget of the mth monitor can hence be described by

$$\sum_{m=1}^{M} P_{FA,m} = P_{FA_{max}} \tag{8}$$

The choice of individual $P_{FA,m}$ is up to the designer. The $P_{FA_{max}}$ budget is shared among the monitors, individual thresholds have to be set more conservatively than if monitors were used by themselves.

For M monitors connected through AND gates, the P_{FA} budget of the mth monitor is given by

$$\prod_{m=1}^{M} P_{FA,m} = P_{FA_{max}} \tag{9}$$

Once again the choice of $P_{FA,m}$ is up to the designer. The $P_{FA,1}$ budget of an individual monitor is increased when combined with a second monitor with $P_{FA,2} < 1$. Detection thresholds can overall be set more aggressively than if monitors were used by themselves.

Similarly, we can compute the missed detection probability when combining multiple monitors of mutually independent measurements. The P_{MD} of M monitors connected through OR gates is equal to the product of individual missed detection probabilities.

$$P_{MD_{OR}} = \prod_{m=1}^{M} P_{MD,m} \tag{10}$$

To calculate he missed detection probability of monitors combined through an AND gate, let us consider the probability of detection. It is equal to the product of individual detection probabilities. With $P_{detection} = 1 - P_{MD}$, the overall $P_{MD_{AND}}$ is therefore given by

$$P_{MD_{AND}} = 1 - \prod_{m=1}^{M} (1 - P_{MD,m}) \tag{11}$$

b) The joint GLRT

To consider all measurements in a single GLRT, let us reconsider Eq. (2) with measurements from M metrics stacked in the vector y. If the measurements from different metrics can be considered mutually independent, we can factor the likelihood ratios and rephrase the expression for the decision variable as

$$\log \Lambda_{1:M} = \sum_{m=1}^{M} \log \Lambda_m \tag{12}$$

Fortunately, the family of χ^2 distributions is closed under addition.

$$-2 \sum_{m=1}^{M} \log \Lambda_m \mid H_0 \sim \chi^2_{k_M} \quad \text{with} \quad k_M = \sum_{m=1}^{M} k_m \tag{13a}$$

$$-2 \sum_{m=1}^{M} \log \Lambda_m \mid H_1 \sim \chi^2_{k_M,\lambda_M} \quad \text{with} \quad \lambda_M = \sum_{m=1}^{M} \lambda_m \tag{13b}$$

The distribution parameters are defined as the sums of the parameters of the individual distributions defined in Eq. (7). The computation of γ and P_{MD} is once again straightforward.

3. A first performance analysis

In the previous subsection we have briefly summarized how to combine measurements from different metrics for a decision about the presence of a spoofing attack under the Neyman-Pearson paradigm. More details and a deeper theoretical analysis is given in [7].

Before analyzing flight data and spoofing scenarios from the TEXBAT dataset, let us consider an analytic performance comparison of a popular metric combination in a simplified simulation: a power measurement combined with means to monitor signal distortion of a single satellite. Attacks either have to strongly overpower the authentic signals (visible to a power monitor) or cause a distortion of the autocorrelation function due to the superposition of authentic and malicious signals with similar power levels. [17, 18] leverage the complementary characteristic of the two defenses in a Bayesian monitor to discern between nominal conditions, multipath, jamming and spoofing, [19] offers a combined monitor where the detection threshold is set empirically based on data recorded in flight and by WAAS stations. [34] suggests and [35] analyses in detail the combination of power and CAF monitoring.

The goal in this analysis is to compare the performance of individual monitors to combinations through an AND and OR gate to the joint GLRT. The considered attack scenario is a lift-off attack with different power advantage. The results should only be interpreted on a qualitative basis. The exact behavior of the various power metrics suggested in the literature is difficult to model and depends on the antenna front end and the exact implementation of e.g. the receiver's AGC gain. Similarly the measured autocorrelation depends on the correlator tap spacing and tracking speed, the background noise and even antenna temperature. The code underlying this simulation is publicly available at `https://github.com/stanford-gps-lab/spoofing-detection.git` for the reader's benefit and to encourage further analysis.

To simplify this analysis, let us use C/N_0 as a power measurement and consider a normalized Delta metric D on the in-phase signal S_I to measure the signal distortion as originally suggested in [10] and summarized in [27].

$$D = \frac{S_I(-\Delta\tau) - S_I(\Delta\tau)}{S_I(0)} \tag{14}$$

Both metrics belong to the simple vs. composite scenario described in Section II. The decision variables $\log \Lambda$ are defined by Eq. (6) and follow the the distribution of Eq. (7).

The following simulation parameters are loosely based on the flight data calibration campaign of a Novatel GIII receiver presented in Chapter III.

- Nominal signal power $\mu_P = 40$ dB

- Nominal power standard deviation $\sigma_P = 2$ dB
- Nominal delta metric $\mu_D = 0$
- early/late correlator pairs spaced $\Delta\tau = 0.052$ chips from the prompt
- Nominal Delta metric standard deviation $\sigma_D = \frac{0.1}{C/N_0 - 30}$
- The $P_{FA_{max}} = 10^{-8}\frac{1}{s}$ budget is shared equally among monitors in the OR and AND combination case

To model and analyze the behavior of the Delta metric during the attack, we need to be able to map signal strength in dB to the output of the correlator taps. This is once again dependent on the employed receiver. Here we leverage the analysis of the flight data in Chapter III. Following the signal model in [36], it has shown that the relationship between the signal amplitude recorded by the prompt correlator $\sqrt{C}$ and the recorded C/N_0 for the used Novatel GIII receiver is given by the following function:

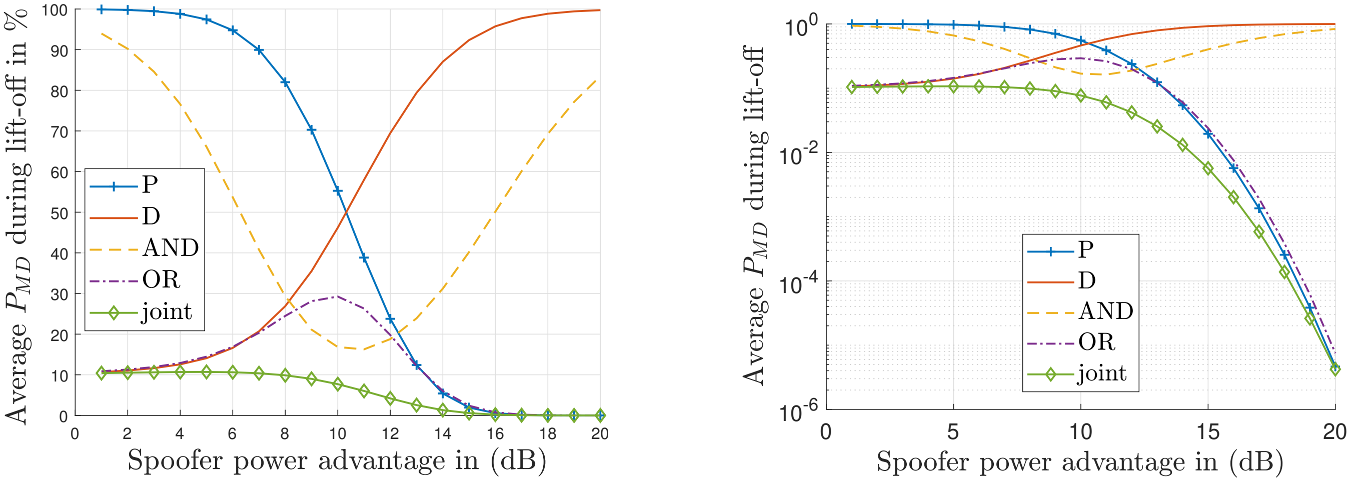
$$\sqrt{C} = \sqrt{10^{(C/N_0 + N_0)/10}} \cdot 10^{15}/2 \tag{15}$$

for background noise of $N_0 = -203$ dB-Hz.

Within the scope of this analysis we consider an attack with constant signal strength that alters its code phase at a constant rate. During the lift-off, the Delta metric can only show an indication of the attack while the code phase difference between authentic and spoofed signal $\Delta\tau_s$ is not more than $1 + \Delta\tau$ chips.

At every epoch during the lift-off, we run a Monte Carlo analysis for 10^6 instances of correlation peaks of the nominal and spoofed signal plus noise. We identify the largest value as what a receiver would track as a peak and compute the Delta metric for every instance. The peak amplitude is then converted back to a C/N_0 value by inverting Eq. (15) for a power measurement.

Figure 1 shows the average P_{MD} across the 107 epochs throughout the attack, both in $\%$ and on a logarithmic scale. The figures contain a lot of information: the Delta (D) metric is a powerful means to detect attacks with a low power advantage, monitoring the received power (P) allows for detection of high power attacks. Each metric alone results in strong results in some attack scenarios, but not all. A combination through an AND gate gives good results in the intermediate case but fails in either extreme case. An OR combination of the complementary techniques leads to overall good results. A significant improvement with respect to the goal of $\min \max_h P_{MD}$ across all possible spoofed hypothesis is achieved by combining the metrics in a single, joint GLRT. It reduces the maximum risk by around 60%.



(a) Linear scale

(b) Logarithmic scale

Figure 1: Average P_{MD} during the lift-off attack across all epochs where $\Delta\tau_s \leq 1 + \Delta\tau$. Shown results are for defenses monitoring either power (P) or signal distortion (D) individually, and for three possible combinations of the two metrics.

Overall the most challenging to detect scenario in this analysis are attacks with very low power advantage. We will see in Section IV that monitoring the pseudorange residuals can be especially helpful in these cases, significantly lowering P_{MD} even further.

A very detailed examination of the results in especially in Figure 1b allows us to reiterate on more subtle points observed throughout this chapter. For attacks with a power advantage of more than 13 dB, considering the power metric alone outperforms the OR monitor. The P_{FA} budget is shared between two monitors combined through an OR gate, resulting in less aggressive individual thresholds. This results in fewer detections through the Power metric. The Delta metric in the meantime does not raise an alarm and "wastes" its P_{FA} budget. Considering it has slightly reduced the detection performance towards high power advantage attacks. Extrapolating the shown results beyong 20 dB, the Power metric alone would similarly begin to outperform the joint monitor as well. While the defenses using combinations of metrics offer significantly more robustness across most attack scenarios, specific defenses show higher detection performance in certain corner cases.

III. FLIGHT DATA ANALYSIS

The spoofing detection framework developed in [7] relies, just like any other detection strategy, on measurement models. In this section we calibrate these models for a power measurement, a delta metric on the absolute values of the correlator tracking taps and the pseudorange residuals. We show in [7] how to phrase the exact $\log \Lambda$'s for each of the variables. All measurements were recorded by a Novatel GIII receiver during a flight test campaign conducted by Zeta Associates. This data has been reported on previously e.g. in [37]. We then validate that using these three metrics together, the detection test does not exceed the false alert probability constraint. All shown data is limited to ground speeds above 50 $\frac{m}{s}$ to exclude epochs on the ground during taxi.

1. Measurement Models

In this section we define measurement models for the three considered metrics. The interested reader is referred to [7] for details on how these models translate to $\log \Lambda$ through Eq. (6).

a) Power Measurement

Inspired by the work in [19] we combine AGC and C/N_0 for a power measurement. A major challenge in spoofing detection through power measurements is that numerous other phenomena such as (un-)intentional radio frequency interference (RFI), geometry changes and signal blockages affect the measurement as well [37]. A detailed analysis of robust thresholds around a power metric is given by [38].

Within the scope of this paper we limit ourselves to combining AGC and C/N_0 into a single, "surrogate" power metric in units of dB.

$$\Delta C_{surrogate} = \Delta(C/N_0)_{max} - \Delta AGC_{dB} \tag{16}$$

where $\Delta(C/N_0)_{max}$ is the difference between the current largest C/N_0 and the average largest C/N_0 during the reference period. Similarly ΔAGC_{dB} is the difference between the current AGC and the average AGC during the reference period converted to dB. Formally, the Δ operator can be defined for a generic metric x at epoch t as

$$\Delta x(t) = x(t) - \frac{1}{T_{ref}} \sum_{\tau=t-\Delta t-T_{ref}+1}^{t-\Delta t} x(\tau) \tag{17}$$

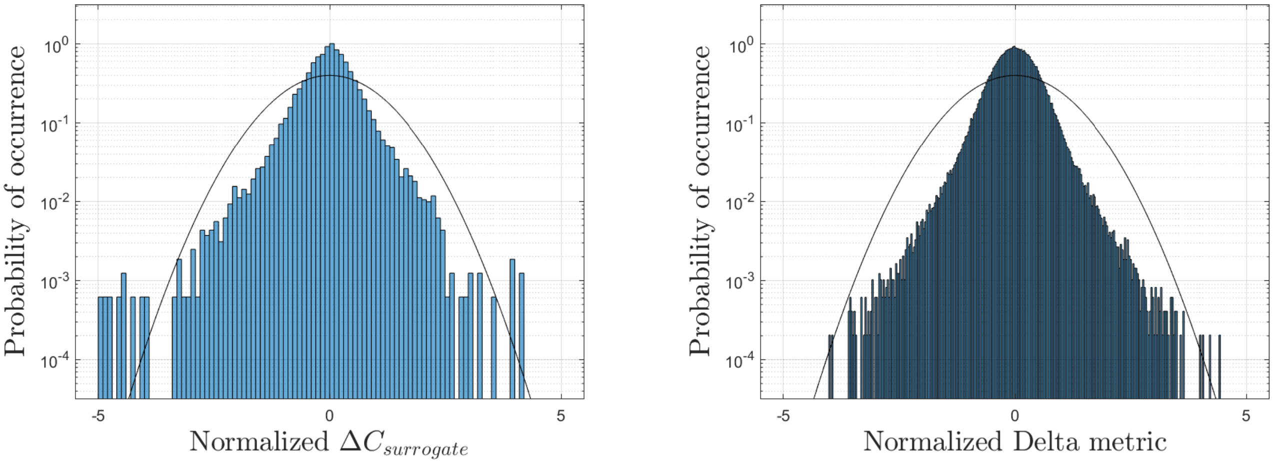
for a reference window length of T_{ref} epochs, which is trailing Δt epochs behind the current epoch. The reference period defines what are considered "nominal" conditions; its ideal values for T_{ref} and Δt vary with the application and threat model.

The used GIII receiver records AGC in units of pulse width. Experiments documented in [39] indicate that for this receiver, a change in 100 pulse width units corresponds to a change in received power of around 1.6 dB. An increases in the AGC value indicates a lower received signal strength and vice versa.

In Figure 2a we show the normalized power metric for a flight from Mather Airport outside Sacramento, CA to Atlantic City International Airport, NJ. We will want to be able to apply this exact model to the TEXBAT data, which is around 400 seconds in length. We therefore set the metric's calibration time window length to a short $T_{ref} = 10$ seconds trailing only $\Delta t = 10$ seconds behind the current epoch. Based on this flight, the power measurement standard deviation is set then to $\sigma_{\Delta C} = 0.5$ dB.

b) Signal Quality Measurement

Several models exist for correlator tap and metric behavior. Early work is summarized in [10], a simple model is e.g. given in [27]. A more detailed model developed by [40] takes into account receiver-specific parameters such as front end bandwidth and code pre-detection integration time. All models generally have an unknown parameter that needs to be determined through analysis of collected data. In the model in [40] this is the loop bandwidth B_L. The models further generally assume Normally distributed noise. Unfortunately, this is not the case in flight and we use a Gaussian over-bound of the measurement noise by setting $B_L = 15Hz$. We show the resulting distribution in Figure 2b after normalization by the measurement model.



(a) Power metric

(b) Delta metric

Figure 2: Histograms of the power and Delta metric normalized by their over-bounding noise models. The data was recorded during a 5.5 hour flight from Sacramento, CA to Atlantic City, NJ.

We can observe in Figure 2 that both measurement models mostly bound the distribution tails, at least for the analyzed flight.

c) Pseudorange Residuals

A lot of work has been done to characterize pseudorange error models for an aircraft environment (the interested reader is referred to e.g. [41]). Within the scope of this proof of concept validation, a simple pseudorange error model is employed. Following the standard notation of the linearized position update in [36], N pseudorange updates $y \in \mathbb{R}^{N \times 1}$ are modeled as independent measurements with the information matrix W.

$$y = Gx + \epsilon \quad \text{with} \quad \epsilon \sim N(0, W^{-1}) \tag{18a}$$

$$W = diag\left(\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_N^2}\right) \tag{18b}$$

$$\sigma_i = \frac{20}{C/N_0 - 33} \tag{18c}$$

for C/N_0 values in dB-Hz, the geometry matrix G and the position update x. Range measurements with $C/N_0 < 33$ dB-Hz are not used for position computation of spoofing detection.

2. Model Validation

An extensive measurement campaign would be necessary to validate the derived models for all flight conditions, and significantly updated models are likely necessary for other applications such as a car in an urban environment. Within the scope of this paper let us simply analyze the behavior of $\log \Lambda$ of the three metrics throughout a separate flight, specifically the return flight from Atlantic City to Mather Airport. We show the results in Figure 3. The figure shows the individual $\log \Lambda$ values of the three metrics, as well as the joint (summed) metric. Specifically we depict all values multiplied by -2. The $\log \Lambda$'s then follow χ^2 distributions, and the positive values can be plotted on a log-scale. We also show the detection threshold for an unrealistically high $P_{FA_{max}} = 10^{-3}\ 1/s$ as a dashed line. The threshold is set deliberately aggressive to cause a significant number of expected and actual false alerts, such that satisfactory behavior w.r.t. the constraint can be confirmed.

13 false alerts are triggered during the shown flight, but 19 would be expected from around 19,000 measurement epochs. The constraint is satisfied.

Apart from the false alert probability, we would like to validate an important assumption: that the measurements are mutually independent. Once again, an extensive measurement campaign would be necessary for a bullet-proof test of the assumption. Within the scope of this paper, let us limit the assumption to measurements we might worry are not independent. One such

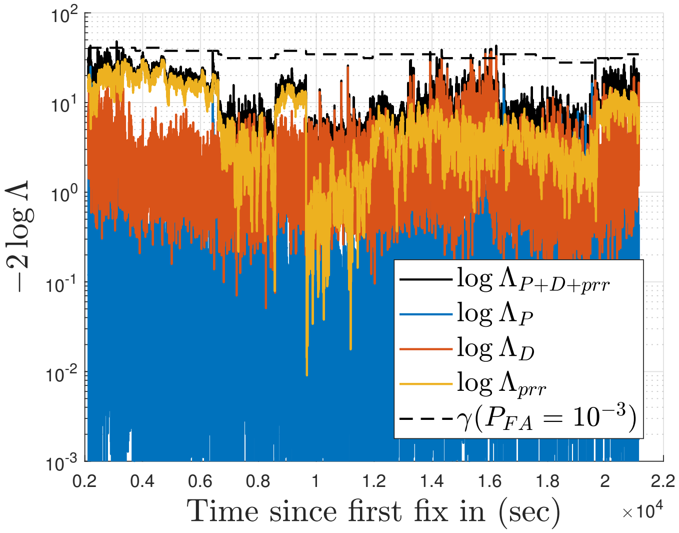
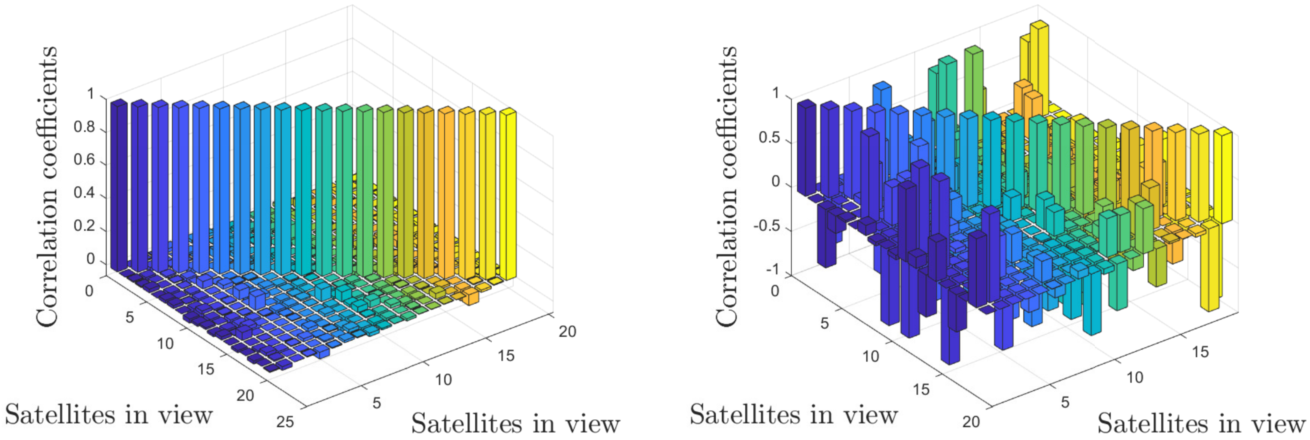


Figure 3: $-2\log\Lambda$ for the power and delta metrics as well as the pseudorange residuals for a flight from Atlantic City to Mather Airport outside Sacramento using the measurement models defined in Section III.1.

concern lies within the decision variable $\log\Lambda_D$ in Figure 3: it is the sum of all $\log\Lambda_{D(s)}$ of the individual satellites. We plot the Pearson correlation coefficients among the various $\log\Lambda_{D(s)}$ of the individual satellites in Figure 4a. The strongest correlation value has a rather small magnitude of 0.0624. A small correlation seems to exist in the data. Since the false alert constraint is satisfied, the over-bounding measurement model appears to compensate for the unmodeled correlation. For an easy visual comparison to clearly correlated data, we plot the correlation among the pseudorange residuals in Figure 4b. The two plots paint drastically different pictures, the residuals obviously show stronger correlation (that is of course accounted for the computation of $\log\Lambda_{prr}$).



(a) Delta metric

(b) Pseudorange residuals

Figure 4: Bar plots of the correlation coefficients during a flight from Atlantic City to Mather Airport outside Sacramento using the measurement models defined in Section III.1.

Let us conclude this brief data analysis with some lessons learned. The measurement models offered in the literature for SQM metrics such as the here employed delta metric do not capture the behavior of the Novatel GIII receiver in flight, and the result is likely the same for other receivers. Similarly, the power metric showed non-Gaussian behavior. We therefore have to use generous over bounds to satisfy our false alert criterion, giving up detection power. As somewhat expected, effects such as banking and (un-) intentional RFI but also other, unexpected effects such as the takeoff rotation causing significant fluctuations in the received power make tight measurement models a challenge. To guarantee robust behavior throughout different flight regimes, an extensive data analysis campaign is likely necessary, to characterize the effect of known and unknown unknowns.

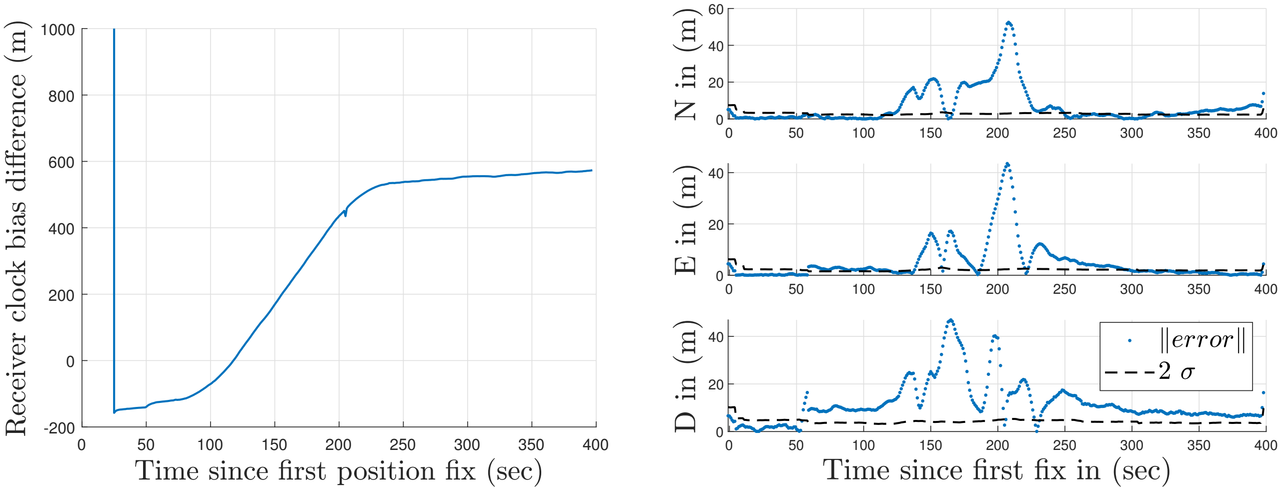
IV. TEST AGAINST TEXBAT

After having validated measurement models (at least to the extent possible in this paper), let us analyze their detection capabilities against spoofing scenarios. For a lack of real world spoofing data we validate the algorithms against the TEXBAT dataset. The dataset is detailed in [42, 43]; in this paper we consider scenarios 2 and 4. Both have been processed through a Novatel GIII receiver, just like the flight data.

1. The Scenarios

We choose TEXBAT scenarios 2 and 4 as they present two different lift-off attacks. In scenario 2, the spoofer has a high power advantage of around 10 dB. The attacker performs a time push, introducing a 600 m offset in the receiver's clock bias. The effect on the GIII receiver is shown in Figure 5a. The spoofer is turned on 50 sec into the scenario, the clock bias offset is introduced between ~ 75 sec and ~ 225 sec. The attack was successful.

Scenario 4 on the other hand is a position push attack introducing a 600 m offset in the ECEF z direction with a very low power advantage of 0.4 dB. Figure 5b shows the offset introduced by the spoofer in a local north (N), east (E), down (D) coordinate system, together with a 2σ uncertainty bound. During the first 50 sec under nominal conditions the uncertainty comfortably bounds the error. Once the spoofing attack starts, the error is increased to several tens of meters, but never reaches the attempted 600 m. After the lift-off is completed around 250 sec, the error drops back close to its original value. This is an indication that the spoofer failed to capture the receiver's tracking correlator tap pairs for most satellites. The transient position error is mostly caused by heavily distorted autocorrelation functions during lift-off. The attack is overall unsuccessful in scenario 4.



(a) Scenario 2

(b) Scenario 4

Figure 5: Clock and position bias introduced during TEXBAT scenarios 2 and 4 when processed through a GIII receiver. The position bias is given in local north (N), east (E), down (D) coordinates. Figure 5b shows as a dashed line the 2σ position covariance for orientation.

2. Performance of the Detector

Let us now evaluate the detector's performance against the two attack scenarios. According to our brief analysis in Section II.3, during scenario 2 (10 dB power advantage) both the power and delta metric should give some tell of the attack. During scenario 4 (0.4 dB power advantage) on the other hand, the delta metric should be the main tell of the attack. We will further see how the pseudorange residuals contribute to the detection especially in the latter case.

We now set $P_{FA_{max}} = 10^{-8}\ 1/s$ to explore the detector's performance under a more realistic false alert probability constraint. Apart from $P_{FA_{max}}$, every parameter and measurement model is equal to its value in Section III.

a) Scenario 2 (10 dB power advantage)

Figure 6 shows the decision variable $\log\Lambda$ for the individual metrics as well as the joint value as solid lines. Each corresponding detection threshold is depicted as a dashed line in the color matching its $\log\Lambda$ value. The thresholds for the individual metrics have been calculated using $P_{FA_{max}}/3$ according to Eq. (8), assuming that the three metrics are combined through logical OR gates. All depicted values have once again been multiplied by -2 such that we can analyze positive values on a log scale.

Figure 6 contains a lot of information. Right when the spoofer is turned on 50 sec into the scenario, we can observe a spike in $\log \Lambda_P$ and $\log \Lambda_D$ due to the attack's high power advantage and superposition of the two signals. $\Delta t = 10$ sec later, the increased power becomes the "new normal" and the power metric drops again. This shows us that Δt should be selected larger for the elevated power metric to persist for longer. The small value here is only due to the short duration of the nominal period in the dataset. The receiver's tracking taps quickly start to track the much stronger signal of the spoofer, reducing the measured distortion significantly. This causes however the pseudorange residuals to be elevated. When the lift-off starts at 75 sec, the signal distortion once again spikes for a short time frame before the receiver tracks the spoofer's signal perfectly reducing the detected distortion and pseudorange residuals. The measured distortion and residuals remain slightly elevated throughout the attack, occasionally exceeding the detection threshold.

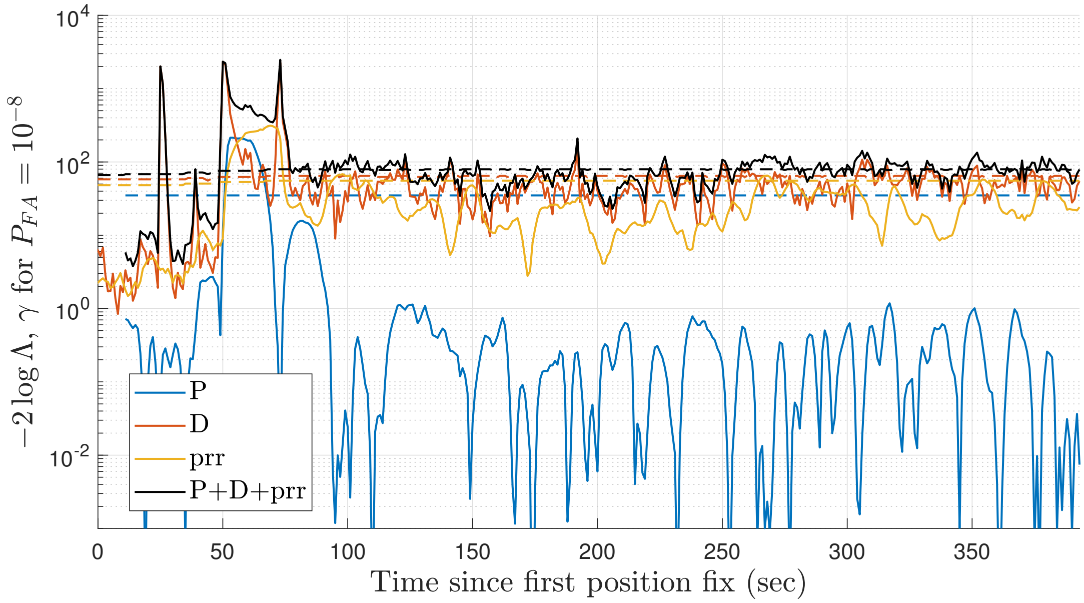


Figure 6: $-2\log\Lambda$ and -2γ during TEXBAT scenario 2. Alarms are raised if $-2\log\Lambda > -2\gamma$. Solid lines represent $\log\Lambda$ values, dashed lines detection thresholds.

Figure 6 vividly shows the transient nature of the employed detection metrics. The attack is detectable during lift-off, but once the spoofer has fully captured the receiver a detection becomes much more challenging.

Already before the attack starts, something noteworthy happens. At around 25 sec we can see the joint detector, driven by the distortion measurement, greatly exceeding the threshold causing a false alert. In Section III we saw compliance with the false alert criterion over hours of flight data. How did a false alert appear now, within less than 30 seconds and $P_{FA_{max}} = 10^{-8}$ $1/s$? A hint at the answer can be found in Figure 5a: the spike corresponds to a reset of the receiver clock. A detailed analysis omitted here for brevity reveals that all delta metrics show almost identical, negative values during the spike. In this particular receiver, the clock reset apparently causes all tracking correlators to be shifted off the peak to the early side. The effect only lasts for a few seconds until the tracking controller has successfully captured the peak again. This points back to our note at the end of Section III.2: significantly more measurement data is necessary to discover and characterize the known and unknown unknowns affecting the measurement model.

b) Scenario 4 (0.4 dB power advantage)

In Figure 5b we observed that the attack in scenario 4 is mostly unsuccessful when processed through the GIII receiver. We can confirm this observations if we once again consider a plot of $-2\log\Lambda$ in Figure 7. From the start of the attack, the pseudorange residuals are elevated. The effect gets worse during lift-off. The distortion metric shows a similar initial spike when the spoofer is turned on, but mainly increases during the lift-off phase. This agrees with our analysis in Section II.3, as the distortion should be a strong tell during a low power advantage attack. The power metric on the other hand is not much help. It never reaches its own threshold in an OR detector and only contributes little to the joint GLRT.

After lift-off the pseudorange residuals remain elevated. A more detailed analysis reveals that the spoofer succeeded in capturing a single satellite which is now disagreeing with the remaining authentic signals. The effect is comparable to long-delay multipath. In this lucky situation, a fault detection and exclusion (FDE) algorithm often implemented in the scope of RAIM [29] would likely be capable of excluding the compromised satellite.

Once again the clock reset at 25-30 sec causes a false alarm, even though the tracking correlators seem less affected. It is unclear whether this is a coincidence or e.g. due to the 10% smaller magnitude of the clock bias before the reset.

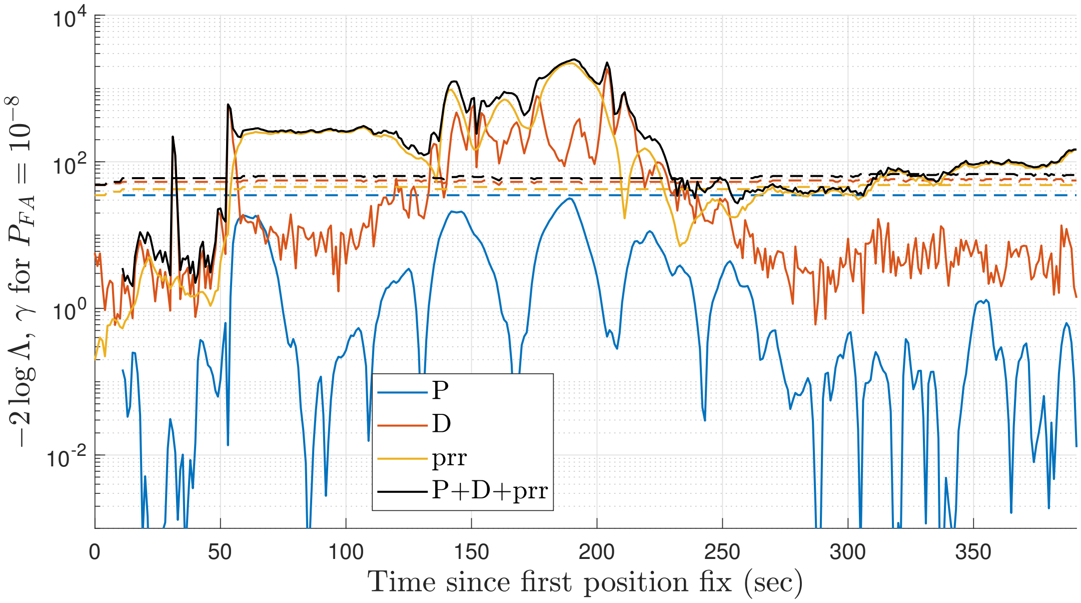


Figure 7: $-2\log\Lambda$ and -2γ during TEXBAT scenario 4. Alarms are raised if $-2\log\Lambda > -2\gamma$. Solid lines represent $\log\Lambda$ values, dashed lines detection thresholds.

Both attacks are detected successfully before any error in the pvt solution is induced while maintaining a stringent constraint on false alerts (it is assumed that the signal distortion caused by the clock bias reset is a receiver-specific artifact that occurs at known epochs, such that the caused alarms can be safely ignored). Different metrics show strong tells at different stages during the attack and their effectiveness varies with the attack scenario. Together, power monitoring, signal distortion monitoring and pseudorange residuals form a robust defense that is available in any receiver.

V. SUMMARY AND CONCLUSION

This paper provides an application example to the general detection framework for spoofing detection using an arbitrary number of metrics presented in [7]. We briefly summarize the framework that calculates a decision variable for each employed metric and offers a joint detection decision. In this paper we focus on the received signal power, autocorrelation function distortion and pseudorange residuals, since these metrics are available in any receiver and show highly complementary characteristics. We calibrate measurement models using several hours of flight data and confirm satisfaction of the false alert probability constraint. We then apply the calibrated models to TEXBAT scenarios 2 and 4, robustly detecting both attacks before any pvt error is introduced.

Several open questions remain for future work. On several occasions we have noted the necessity for extensive measurement campaigns to fully characterize the measurement model. We have already observed several unexpected phenomena affecting the metric's behavior, such as the aircraft's rotation on takeoff or the receiver's clock reset. The employed over-bounding error models captured these effects, but at the cost of some detection performance.

ACKNOWLEDGMENTS

Several people have supported the work leading up to this paper. Karl Shallberg and the team at Zeta Associates by providing the flight data, Prof. Dennis Akos by supplying TEXBAT recordings of the GIII receiver and the associated parser, and Prof. J. David Powell with his invaluable advice and guidance during this research. The authors would also like to thank the Federal Aviation Administration (FAA) for sponsoring this research.

REFERENCES

[1] C4ADS, “Exposing GPS Spoofing in Russia and Syria C4ADS innovation for peace Above Us Only Stars,” Tech. Rep., 2019. [Online]. Available: www.c4ads.org

[2] I. GNSS, “Sinister Spoofing in Shanghai,” 2019. [Online]. Available: https://insidegnss.com/sinister-spoofing-in-shanghai/

[3] John A. Volpe National Transportation Systems Center, “VULNERABILITY ASSESSMENT OF THE TRANSPORTATION INFRASTRUCTURE RELYING ON THE GLOBAL POSITIONING SYSTEM,” Tech. Rep., 2001. [Online]. Available: https://www.hsdl.org/?view{&}did=1815

[4] A. Jafarnia-Jahromi, A. Broumandan, J. Nielsen, and G. Lachapelle, “GPS vulnerability to spoofing threats and a review of antispoofing techniques,” *International Journal of Navigation and Observation*, 2012.

[5] C. Günther, “A Survey of Spoofing and Counter-Measures,” *Navigation, Journal of the Institute of Navigation*, 2014.

[6] M. L. Psiaki and T. E. Humphreys, “GNSS Spoofing and Detection,” *Proceedings of the IEEE*, vol. 104, no. 6, pp. 1258–1270, 2016.

[7] F. Rothmaier, Y.-h. Chen, S. Lo, and T. Walter, “A Framework for GNSS Spoofing Detection through Combinations of Metrics,” *IEEE Transactions on Aerospace and Electronic Systems*, 2021.

[8] L. Scott, “Anti-Spoofing and Authenticated Signal Architectures for Civil Navigation Systems,” *Proceedings of the 16th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS/GNSS 2003)*, 2003.

[9] Ç. Tanıl, S. Khanafseh, M. Joerger, and B. Pervan, “An INS Monitor to Detect GNSS Spoofers Capable of Tracking Vehicle Position,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 1, pp. 131–143, 2018.

[10] R. E. Phelts, “Multicorrelator Techniques for Robust Mitigation of Threats To GPS Signal Quality,” Ph.D. dissertation, Stanford University, 2001.

[11] E. G. Manfredini, F. Dovis, and B. Motella, “Validation of a signal quality monitoring technique over a set of spoofed scenarios,” in *2014 7th ESA Workshop on Satellite Navigation Technologies and European Workshop on GNSS Signals and Signal Processing, NAVITEC 2014 - Proceedings*. Noordwijk, Netherlands: IEEE, 2015, pp. 1–7.

[12] M. Troglia Gamba, M. D. Truong, B. Motella, E. Falletti, and T. H. Ta, “Hypothesis testing methods to detect spoofing attacks: a test against the TEXBAT datasets,” *GPS Solutions*, vol. 21, no. 2, pp. 577–589, 2017.

[13] A. Broumandan, R. Siddakatte, and G. Lachapelle, “Feature article: An approach to detect GNSS spoofing,” *IEEE Aerospace and Electronic Systems Magazine*, pp. 64–75, 2017.

[14] D. M. Akos, “Who’s Afraid of the Spoofer? GPS/GNSS Spoofing Detection via Automatic Gain Control (AGC),” *NAVIGATION, Journal of the Institute of Navigation*, vol. 59, no. 4, pp. 281–290, 2012.

[15] A. Jafarnia-Jahromi, A. Broumandan, J. Nielsen, and G. Lachapelle, “Pre-despreading authenticity verification for GPS L1 C/A signals,” *Navigation, Journal of the Institute of Navigation*, vol. 61, no. 1, pp. 1–11, 2014.

[16] B. W. Parkinson and P. Axelrad, “Autonomous GPS Integrity Monitoring Using the Pseudorange Residual,” *Navigation*, vol. 35, no. 2, pp. 255–274, 1988.

[17] K. D. Wesson, J. N. Gross, T. E. Humphreys, and B. L. Evans, “GNSS Signal Authentication Via Power and Distortion Monitoring,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 2, pp. 739–754, 2018.

[18] J. N. Gross, C. Kilic, and T. E. Humphreys, “Maximum-likelihood power-distortion monitoring for GNSS-Signal authentication,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 1, pp. 469–475, 2019.

[19] D. Miralles, A. Bornot, P. Rouquette, N. Levigne, D. M. Akos, Y. H. Chen, S. Lo, and T. Walter, “An Assessment of GPS Spoofing Detection Via Radio Power and Signal Quality Monitoring for Aviation Safety Operations,” *IEEE Intelligent Transportation Systems Magazine*, vol. 12, no. 3, pp. 136–146, 2020.

[20] H. Tao, H. Li, and M. Lu, “A method of detections’ fusion for GNSS anti-spoofing,” *Sensors (Switzerland)*, vol. 16, no. 12, p. 18, 2016.

[21] A. Molina-Markham and J. J. Rushanan, “Positioning, Navigation, and Timing Trust Inference Engine,” in *Proceedings of the 2020 International Technical Meeting (ITM) of The Institute of Navigation ION*. San Diego, CA: The Institute of Navigation, 2020, p. 15.

[22] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York: John Wiley & Sons, Inc., 2001.

[23] D. Borio and C. Gioia, “A sum-of-squares approach to GNSS spoofing detection,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 4, pp. 1756–1768, 2016.

[24] F. Rothmaier, Y.-H. Chen, S. Lo, and T. Walter, “GNSS Spoofing Detection through Spatial Processing,” *Navigation, Journal of the Institute of Navigation*, vol. 68, no. 2, pp. 243–258, 2021.

[25] A. Jovanovic, C. Botteron, and P. A. Fariné, “Multi-test detection and protection algorithm against spoofing attacks on GNSS receivers,” in *Record - IEEE PLANS, Position Location and Navigation Symposium*, Monterey, CA, 2014, pp. 1258–1271.

[26] A. Pirsiavash, A. Broumandan, and G. Lachapelle, “Two-Dimensional Signal Quality Monitoring For Spoofing Detection,” in *Navitec 2016*. Noordwijk, Netherlands: ESA/ESTEC, 2016.

[27] C. Sun, J. W. Cheong, A. G. Dempster, H. Zhao, and W. Feng, “GNSS spoofing detection by means of signal quality monitoring (SQM) metric combinations,” *IEEE Access*, vol. 6, pp. 66 428–66 441, 2018.

[28] A. M. Khan, N. Iqbal, A. A. Khan, M. F. Khan, and A. Ahmad, “Detection of Intermediate Spoofing Attack on Global Navigation Satellite System Receiver through Slope Based Metrics,” *Journal of Navigation*, vol. 73, no. 5, pp. 1052 – 1068, 2020.

[29] J. Blanch, T. Walter, P. Enge, Y. Lee, B. Pervan, M. Rippl, A. Spletter, and V. Kropp, “Baseline Advanced RAIM User Algorithm and Possible Improvements,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 1, pp. 713–732, 2014.

[30] K. Ali, E. G. Manfredini, and F. Dovis, “Vestigial signal defense through signal quality monitoring techniques based on joint use of two metrics,” in *IEEE PLANS, Position Location and Navigation Symposium*. Monterey, CA: IEEE, 2014, pp. 1240–1247.

[31] S. Lo, L. Boyce, T. Walter, P. Enge, B. Peterson, B. Wenzel, and K. Carroll, “Loran Fault Trees for Required Navigation Performance 0 . 3,” in *ION NTM*, Anaheim, CA, 2003, pp. 352–361.

[32] J. Fernow and D. O’Laughlin, “Estimating continuity of GNSS,” in *Proceedings of the 17th International Technical Meeting of the Satellite Division of the Institute of Navigation, ION GNSS 2004*, 2004, pp. 2113–2123.

[33] M. Joerger, Y. Zhai, I. Martini, J. Blanch, and B. Pervan, “ARAIM continuity and availability assertions, assumptions, and evaluation methods,” in *ION 2020 International Technical Meeting*. The Institute of Navigation, 2020, pp. 404–420.

[34] C. Hegarty, B. W. O’Hanlon, A. Odeh, K. Shallberg, and J. Flake, “Spoofing detection in GNSS receivers through cross-ambiguity function monitoring,” in *Proceedings of the 32nd International Technical Meeting of the Satellite Division of the Institute of Navigation, ION GNSS+ 2019*. Miami, Florida: Institute of Navigation, 2019, pp. 920–942.

[35] T. Zhang, X. Chen, D. He, and Y. Jin, “Performance Analysis and Tests for GNSS Spoofing Detection Based on the Monitoring of Cross Ambiguity Function and Automatic Gain Control,” in *Proceedings of the 2021 International Technical Meeting of The Institute of Navigation*, 2021, pp. 98 – 110.

[36] P. Misra and P. Enge, *Global Positioning System: Signals, Measurements, and Performance; Revised Second Edition*, 2nd ed. Lincoln, Massachusetts: Ganga-Jamuna Press, 2011.

[37] C. Hegarty, A. Odeh, K. Shallberg, K. D. Wesson, T. Walter, and K. Alexander, “Spoofing Detection for Airborne GNSS Equipment,” in *Proceedings of the 31st International Technical Meeting of the Satellite Division of the Institute of Navigation, ION GNSS+ 2018*, Miami, Florida, 2018, pp. 1350–1368.

[38] S. Lo, F. Rothmaier, D. M. Akos, and T. Walter, “Developing GNSS Interference/Spoof Detection Thresholds for Receiver Power Monitoring,” in *Proceedings of the 34th International Technical Meeting of the Satellite Division of the Institute of Navigation, ION GNSS+ 2021*, 2021.

[39] E. G. Manfredini, “Signal processing techniques for GNSS anti-spoofing algorithms,” Ph.D. dissertation, POLITECNICO DI TORINO, 2017.

[40] J. W. Betz and K. R. Kolodziejski, “Extended theory of early-late code tracking for a bandlimited GPS receiver,” *Navigation, Journal of the Institute of Navigation*, vol. 47, no. 3, pp. 211–226, 2000.

[41] Working Group C, “EU-US Cooperation on Satellite Navigation Working Group C, ARAIM Technical Subgroup, Milestone 3 Report,” Tech. Rep. 1, 2016. [Online]. Available: https://www.gps.gov/policy/cooperation/europe/2016/working-group-c/ARAIM-milestone-3-report.pdf

[42] T. E. Humphreys, J. A. Bhatti, D. Shepard, and K. D. Wesson, “The Texas spoofing test battery: Toward a standard for evaluating GPS signal authentication techniques,” in *Proceedings of the 25th International Technical Meeting of the Satellite Division of the Institute of Navigation 2012, ION GNSS 2012*, Nashville, TN, 2012, pp. 3569–3583.

[43] C. A. Lemmenes, P. Corbell, and S. Gunawardena, “Detailed analysis of the TEXBAT datasets using a high fidelity software GPS receiver,” in *29th International Technical Meeting of the Satellite Division of the Institute of Navigation, ION GNSS 2016*, vol. 5, 2016, pp. 3027–3032.