Calibrated Simulation for Long-Range UAV RF Dataset Generation and Counter-UAS Evaluation

Karan Kapoor, Jared Surajballi, and Hao Wang
Stevens Institute of Technology, USA
{kkapoor4, jsurajba, hwang9}@stevens.edu

Kang Jun Bai
Air Force Research Laboratory, USA
kang.jun.bai@us.af.mil

***Abstract*—Passive RF sensing is a cornerstone of counter-UAS defense, and its detection models are trained on public UAV RF datasets. Yet every mainstream public dataset was captured within 150 m of the emitter—most within 30 m—while operational requirements specify detection of Group 1 UAS at 2 km and beyond. The prevailing workaround, rescaling additive noise on bench captures, varies SNR as a unitless knob and reproduces none of the physics of range. We present *RF-UAVSim*, an open-source, scenario-driven framework coupling 3D UAV mobility to an absolute-power channel model: transmit power in dBm, free-space path loss, antenna gains, and a thermal noise floor. Standoff distance therefore maps to received SNR through a link budget rather than a label, while scenario kinematics and geometry set Doppler and fading. Its channel is addressed by absolute sample index, verified bit-identical across 36 packet-partition and 7 tick-rate schedules and agreeing with closed-form propagation on 11 analytical tests. A calibrated energy detector, frozen on 10,000 noise-only windows, holds an empirical detection probability of 1.0 to 1.5 km and falls to 0.326 at 3 km at 644 false-positive windows/hour—a range-conditioned operating curve no public dataset can support. Comparing 14,484 released DroneRFb-Spectra arrays against 2,560 synthetic records then bounds fidelity: the cross-domain median marginal RMSE is 0.245 against a 0.063 real–real baseline, so generic families are *not* substitutes for measured captures. The metric still orders families by spectral-occupancy similarity: OFDM is closest for all 24 classes and FM farthest, consistent with the wideband OFDM structure of these downlinks.**

***Index Terms*—UAV detection, RF signal simulation, counter-UAS, dataset generation, machine learning**

I. Introduction

Small uncrewed aircraft systems (UAS) are a defining threat to military formations and critical infrastructure alike. Passive RF sensing detects and characterizes the command-and-control (C2), telemetry, and video links a drone emits. It remains a cornerstone counter-UAS modality: passive, low-cost, and effective against the small radar cross-sections of Group 1–2 airframes. The operational requirement, however, is measured in kilometers. Recent U.S. Department of Defense solicitations require detection of Group 1 UAS at a minimum range of 2 km [1]. Protocol-aware research prototypes have decoded DJI identification signals out to 3.7 km, though that result is specific to those emitters and to a decoding receiver rather than generic passive classification [2].

The data that trains and evaluates RF-based detection models tells a different story. As we survey in Section II, every mainstream public UAV RF dataset was captured within 150 m of the emitter—and most within 30 m. DroneRFa, the longest-range dataset with labeled raw IQ, tops out at 150 m [3]; datasets released as recently as 2026 still capture at 2–30 m [4]. The consequence is physical, not incidental. Moving from 150 m to 3 km adds $\sim$26 dB of free-space path loss. That pushes a link received at 30–50 dB SNR on the bench into the single-digit and negative-SNR regime, with air-to-ground multipath and co-channel interference from a far wider footprint. RF classifiers degrade severely under such distribution shift [5], and the state of the art already discards the far bucket of existing data as too weak [6]. In short: models guarding the 2 km perimeter are trained on 20 m data.

Closing this gap with measurement alone is hard. Long-range captures need large instrumented ranges, aviation and spectrum authorizations, and per-emitter flight campaigns; the resulting raw IQ may be restricted in military settings. The prevailing workaround—rescaling additive noise on bench captures [7]–[9]—changes SNR as a unitless knob while reproducing none of the physics of range.

This paper presents *RF-UAVSim*, an open-source, scenario-driven simulation framework that makes the long-range regime available for dataset generation and model evaluation. *RF-UAVSim* couples analytic 3D UAV mobility with a channel model expressed in *absolute power*. Absolute transmit power, free-space path loss, antenna gains, and a thermal receiver noise floor jointly determine received SNR. Standoff distance therefore maps to received power and SNR through a link budget rather than a label, while the scenario's kinematics and geometry set Doppler and fading. Waveforms come from synthetic modulation generators or from replaying real near-range captures through the simulated channel. Outputs are SigMF recordings carrying per-link ground truth (range, received power, Doppler) and a scenario hash, which makes every dataset regenerable from its configuration file.

A simulator claiming physical units invites an objection: how do we know the units are right? The channel is addressed by *absolute sample index* rather than buffer offset, so a rendered window is a pure function of scenario and index, independent of how the run was chopped into packets or ticks—and that property is testable. We test it alongside closed-form agreement for path loss, delay, carrier offset, Doppler, fading, and the noise floor before using the simulator for measurement. We report a range-conditioned detection experiment spanning 26.0 dB of received power, and a feature-morphology comparison against the largest public set of released UAV RF feature arrays, which bounds fidelity rather than asserting it. Our contributions are:

- **Gap characterization.** A survey of public UAV RF datasets showing that stated capture ranges end at 150 m, that recent datasets synthesize SNR without physics, and that no public raw-IQ data or simulator exists for the kilometer-scale regime that counter-UAS requirements specify (Section II, Table I).
- **An absolute-power, index-addressed simulator.** *RF-UAVSim*, whose link budget, geometry-derived Doppler and fading, and provenance-first SigMF pipeline turn scenario files into regenerable long-range datasets compatible with existing training code, and whose channel passes 11 analytical, 36 packet-partition, and 7 tick-rate gates (Sections III–IV-B).
- **Two range-aware measurements.** A calibrated energy detector characterized over 11 ranges with a measured false-alarm rate (Section IV-C), and a full-dataset real–synthetic morphology comparison across 24 real classes and 8 waveform families that reports its own baselines and therefore bounds, rather than overstates, synthetic fidelity (Section IV-D).

II. Related Work and the Range Gap

A. Public UAV RF Datasets Stop at 150 m

Table I surveys public UAV RF datasets with their *stated* capture distances, and two patterns emerge. Distance is only a proxy for difficulty—EIRP, bandwidth, noise figure, and interference all matter—but it is the one quantity these datasets consistently report. First, the longest labeled capture range in any mainstream dataset is 150 m. DroneRFa [3] buckets its outdoor flights into 20–40, 40–80, and 80–150 m. Every other dataset that states a distance is shorter, down to 1.2 m for chamber captures [10], and the rest are bench, anechoic-chamber, or unstated near-range field captures. Second, the largest recent datasets manufacture SNR diversity *synthetically*: RFUAV (37 drones, 1.3 TB) records transmitters “directly connected to the receiver… or positioned in close proximity” and rescales AWGN afterwards [7]. CageDroneRF captures in a Faraday cage and augments with AWGN, frequency shift, and interferer mixing [9]. Glüge et al. mix chamber captures with noise at -20 to $+30$ dB [8]. AWGN scaling of a bench capture reproduces none of the Doppler, fading, delay, or front-end effects that accompany low SNR in the field. The sole public artifact beyond 150 m is DroneRFz [11], a small (1.56 GB) set of 512×512 spectrogram *images*—no raw IQ, no per-sample distance labels—collected at up to 1 km.

B. Models Inherit the Gap, and the Bar Keeps Rising

State-of-the-art RF drone recognition is built on these datasets. S3R [16] performs open-set recognition over DroneRFb-Spectra signal semantics, Open-RFNet [6] adds multi-domain supervised contrastive learning. Tellingly, Open-RFNet’s authors restrict evaluation to DroneRFa’s near-range and indoor subsets because the 80–150 m signals are already too weak [6]. SOTA methods struggle at the far edge of a 150 m dataset, and classifier accuracy is known to collapse under channel, environment, and time shift—with cross-environment drops from 99% to as low as 1% [5].

Kilometer-scale passive RF detection is nonetheless practical: protocol-aware decoding of DJI identification signals reached 1.3–3.7 km with commodity receivers [2], and Glüge et al. validated a chamber-trained CNN in field trials at 110–670 m, training on synthetic-noise proxies precisely because no long-range data exists [8]. None of these efforts released long-range captures, while U.S. DoD counter-UAS solicitations require Group 1 detection at a minimum of 2 km [1], an order of magnitude beyond the public data ceiling.

TABLE I
Public UAV RF datasets and their stated capture distances.

Dataset	Year	Stated capture range	Types
DroneRF [12]	2019	indoor lab (∼m)	3
RC controllers [13]	2020	bench	17
DroneDetect [14]	2021	near-range field, n/s	7
GENESYS [10]	2021	1.2–4 m	5
CardRF [15]	2022	8–12 m	6
DroneRFa [3]	2024	**20–150 m**	24
DroneRFb-Spectra [16]	2024	urban, n/s	24
UAVSig [17]	2024	near-range, n/s	8
Noisy Drone RF [8]	2024	chamber + synth. SNR	10
RFUAV [7]	2025	bench + synth. AWGN	37
CageDroneRF [9]	2026	cage/campus + augm.	23
TSMS-Drone [4]	2026	2–30 m	4
DroneRFz [11]	2026	$\leq$1 km (spectr. only)	7

n/s: not stated; synth./augm.: synthetic or augmented SNR; spectr.: spectrogram images only. Types counts distinct emitters released as classes.

C. Synthetic RF Data and Simulators

Synthetic data has strong precedent in RF ML: Scholl showed that carefully impaired synthetic training data reaches up to 95% accuracy on real-world signals [18]. For UAVs, generation tooling is fragmentary: a reverse-engineered DroneID frame crafter, AWGN and interference augmentation bundled with datasets [7], [9], and GAN-based augmentation bound to the near-range source distribution. Among tools that emit labeled UAV RF datasets, we are not aware of one tying received power to an absolute link budget so that distance sets SNR in physical units. General wireless and ray-tracing simulators model geometry-dependent propagation but are not packaged as UAV dataset generators; *RF-UAVSim* contributes the pairing, not the propagation model.

III. System Design

RF-UAVSim is a scenario-driven simulation framework that couples 3D UAV mobility, an absolute-power RF link model, and a reproducibility-first dataset pipeline. Its design goal is not to replace over-the-air measurement, but to make the *long-range, low-SNR* operating regime—which is largely absent from public datasets—available for training and stress-testing

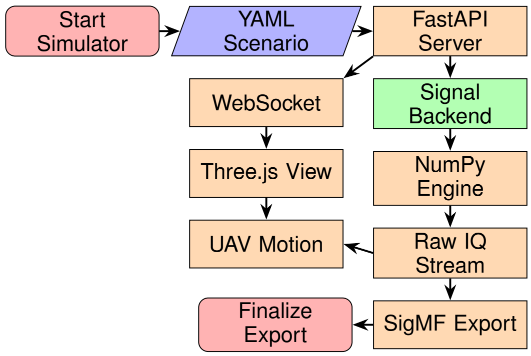

Fig. 1. Software architecture and principal data flow of *RF-UAVSim*. A YAML scenario configures the FastAPI simulation server, which coordinates signal generation, receiver-IQ processing, visualization, and dataset export.

RF-based UAS detection and recognition models. Figure 1 shows the software architecture and data flow.

A. Scenario Model and Mobility

Every experiment is a declarative YAML scenario specifying objects (UAVs, receivers, interferers), motion, emitter and receiver characteristics, channel effects, and capture policy, hashed (SHA-256) into every output recording so a dataset traces back to the exact configuration that produced it.

The mobility engine evaluates closed-form 3D trajectories: constant-acceleration kinematics, circular orbits, and spline-interpolated waypoint paths. Positions, velocities, and headings are analytic, making the induced range, radial velocity, and geometry exactly reproducible. Geometry is evaluated on a grid *independent* of the simulation tick rate, so a trajectory yields the same channel regardless of how finely the run is stepped, as Section IV-B verifies.

B. Signal Sources

RF-UAVSim deliberately separates *what is transmitted* from *how it propagates*, drawing waveforms from four interchangeable source types. Synthetic waveforms come from TorchSig [19], covering ASK, FSK, MSK, PSK, QAM, OFDM, AM, and FM families with configurable bandwidth, burst structure, and transmitter-side impairments. These serve as class proxies for video downlinks and telemetry bursts. Replay passes real SigMF [20] captures of drone links through the simulated channel, carrying authentic microstructure into new geometries. Such a capture already carries its own receiver noise and propagation loss, and we apply no de-embedding, so replay cascades a second channel onto the first and suits microstructure, not absolute-power claims. Approximate IQ can also be regenerated from public spectrogram-feature datasets [16]. GNU Radio flowgraph emitters plus deterministic tones and noise sources support calibration. A burst scheduler gates any source into duty-cycled packets, approximating the on–off structure of C2 and video links.

All sources are seeded and *absolutely indexed*: a block of IQ is a pure function of the scenario seed and the absolute sample index, not of the buffer that requested it. This is what makes the invariance results of Section IV-B achievable at all.

C. Absolute-Power Channel

For transmitter m and receiver r, a received window is

$$y_r[n] = \sum_m a_{m,r} \sum_\ell h_{m,r,\ell}\, s_m[n - \tau_{m,r,\ell}] \times e^{j2\pi(\Delta f_{m,r} + f^D_{m,r})n/f_s} + w_r[n], \tag{1}$$

where $a_{m,r}$ combines antenna gains with free-space loss, $h_{m,r,\ell}$ are fading or multipath coefficients (at most eight configured gain/phase/delay taps), $\tau_{m,r,\ell}$ is propagation plus tap delay applied at fractional-sample resolution, and $w_r[n]$ is receiver noise after the receive filter. Geometry-dependent terms are held constant within a rendering block and re-evaluated between blocks, so Eq. (1) is a quasi-static approximation of a moving link.

The stage is deliberately simple but *dimensionally honest*: sample power corresponds to absolute watts, so range-dependent SNR is a physical consequence of geometry, not a post-hoc label. At distance d, with transmit power P_{tx} (dBm) and antenna gains G_{tx}, G_{rx} (dB),

$$\begin{aligned} P_{rx}(d) &= P_{tx} + G_{tx} + G_{rx} - L_{\text{FSPL}}(d, f_c), \\ L_{\text{FSPL}}(d, f_c) &= 20\log_{10} d + 20\log_{10} f_c - 147.55, \end{aligned} \tag{2}$$

with d in meters and f_c in hertz, for which the -147.55 constant holds. Radial velocity v_r produces $f_D = -v_r f_c/c$, so Doppler follows scenario kinematics, not range. Rather than assuming the noise bandwidth equals the sample rate—a common shortcut that silently misstates the noise floor—the receiver filter defines an equivalent noise bandwidth $B_{\text{eq}} = f_s\big(\sum_n |h[n]|^2\big)/\big|\sum_n h[n]\big|^2$, so the floor is

$$N_{\text{dBm}} = -174 + 10\log_{10}(B_{\text{eq}}) + NF. \tag{3}$$

Because this floor is absolute rather than scaled to the signal, doubling the standoff distance costs exactly 6 dB of SNR under the modeled free-space assumptions with fixed gains and orientation; real sites need not follow a path-loss exponent of two. Consider one representative link, 2.45 GHz with $P_{tx} = 10$ dBm, $G_{tx} + G_{rx} = 6$ dB, $B = 5$ MHz, $NF = 5$ dB; carrier and bandwidth are per-emitter scenario fields, so other bands follow the same budget. A UAV at 20 m is received at roughly 52 dB SNR, at 150 m at 34 dB, at 1 km at 18 dB, and at 10 km at -2 dB. Public near-range datasets occupy a high-SNR sliver of this space; the regime that matters for wide-area counter-UAS is exactly the one a calibrated simulator can populate with unlimited labeled data.

D. Dataset Generation and Provenance

A headless capture mode runs a scenario to completion and writes every transmitted and received stream as SigMF (`cf32`) recordings, each embedding the scenario hash, per-link ground truth over time (range, direct-path received power, Doppler), source type, label, and a split assignment validated for disjointness at export. A dataset builder converts labeled captures into model-ready time–frequency feature sets compatible with the DroneRFb-Spectra format [16], so simulator output can be

Fig. 2. The *RF-UAVSim* browser interface during a run: an analytic 3D scene (left) and live I/Q, rolling spectrogram, and current-window STFT (right). Displays are qualitative; all measurements use unrounded backend IQ.

mixed with or substituted for public datasets without changing training code; generation is resumable by family and shard. Because every stage is seeded, a dataset is regenerable from a scenario file and manifest, reducing dataset distribution to configuration distribution—which matters where raw captures are restricted but scenario descriptions are shareable.

E. Implementation

RF-UAVSim comprises 14 K lines of Python (NumPy, PyTorch, TorchSig, FastAPI), a browser-based Three.js viewer, and a headless batch path. Figure 2 shows the viewer. Feature and open-set scoring *interfaces* let multiple models consume a common observation, but no trained checkpoint is integrated, so we report no model-versus-range accuracy, augmentation gain, or open-set result. A scenario hash fixes configuration but not code or dependencies, so the release pins an environment lockfile alongside the tagged commit.

IV. Evaluation

Our evaluation proceeds in the order of what each result can support: the channel implements the physics it claims and is invariant to run scheduling (Section IV-B); the framework yields a range-conditioned operating curve in absolute units (Section IV-C); and simulator-generated morphology is compared against real data with baselines that make the number interpretable (Section IV-D).

A. Setup

The analytical path-loss sweep of Section IV-B uses a 2.45 GHz carrier at 150, 300, and 500 m and 1, 2, and 3 km. That carrier is an experimental choice: carrier, bandwidth, and emitter set are scenario YAML fields, and many platforms use other bands. The detector of Section IV-C uses eleven ranges: 150, 300, 500, 750 m and 1, 1.25, 1.5, 1.75, 2, 2.5, 3 km. The detector uses a -30 dBm, 312.5 Hz tone, 0 dB combined antenna gain, a 20 kS/s receiver, 10 kHz bandwidth, $NF = 5$ dB, and 1024-sample windows; range is straight-line separation, and fading, interference, and obstruction are disabled for the sweep and enabled individually in their own tests. The comparison dataset of Section IV-D holds 2,560 complex64 records of 525,312 samples across 8 waveform families, 320 records each, generated at 100 MS/s with a 20 dB receiver SNR and TorchSig impairment level 1 (10.02 GiB).

TABLE II
Channel and invariance validation. All tests pass.

Test	Error or result	Tolerance
Free-space path loss	7.65×10^{-7} dB	10^{-5} dB
Fractional delay	2.87×10^{-6} sample	0.02 sample
Carrier offset	1.53×10^{-5} Hz	0.01 Hz
Geometric Doppler	1.09×10^{-6} Hz	0.1 Hz
Thermal noise (B_{eq})	0.0053 dB	0.1 dB
Rayleigh / Rician power	0.0025 / 0.000113	0.03
Channel-grid path	1.71×10^{-6} m	0.03 m
Packet partitions	36/36 pass	bit-exact
Tick-rate schedules	7/7 pass	bit-exact

B. Channel and Invariance Validation

We first check the claim that makes everything else meaningful: that Eqs. (1)–(3) are implemented, not merely described. The gate runs 11 analytical tests against closed-form expectations; Table II lists the eight distinct quantities covered. Measured receiver-IQ loss tracks $20\log_{10}(R/150)$ to 7.65×10^{-7} dB over the full 150 m–3 km sweep. The thermal floor matches Eq. (3) to 0.0053 dB at an equivalent noise bandwidth of 18,793 Hz, a floor that a sample-rate bandwidth assumption would have misstated. Fractional delay, carrier offset, Doppler, fading power, and channel-grid convergence all pass with margins of two to five orders of magnitude.

Sources, channel, and noise are addressed by absolute index. The rendered receiver IQ is therefore *bit-identical* across equal, unequal, and random packet schedules for undegraded, stochastic, delayed, rate-converted, Doppler, fading, multipath, obstructed, and combined channels (36/36), and across static, kinematic, curved, and non-divisible-rate tick schedules (7/7). Without this, dataset contents would depend on buffer sizes chosen for unrelated reasons. These tests check the implementation against the same closed-form expressions the simulator is built from; no traceable source, chamber, or field measurement enters anywhere, so the link model is analytically validated rather than calibrated in the instrumentation sense.

C. Calibrated Range-Conditioned Detection

We next use the channel for the measurement that motivates the paper: how detection degrades with distance in absolute units. A mean-energy detector is calibrated once from 10,000 noise-only windows at a target 10^{-2} per-window false-alarm probability, then *frozen*. We evaluate 5,000 windows at each of 11 ranges and measure the false-alarm rate on a separate signal-absent stream covering 1 simulated hour.

Figure 3 shows an empirical detection probability of 1.0 through 1.5 km, 0.997 at 1.75 km, 0.950 at 2 km, 0.627 at 2.5 km, and 0.326 at 3 km, at an observed 644 false-positive windows per hour (nominal 95% interval [596, 695]). These are independent 51.2 ms window decisions, not adjudicated alarms; clustering them into alarms would lower the operational rate. Direct-path received power falls from -113.8 dBm at 150 m to -139.8 dBm at 3 km, a 26.0 dB span set by geometry alone, not by a rescaled noise level.

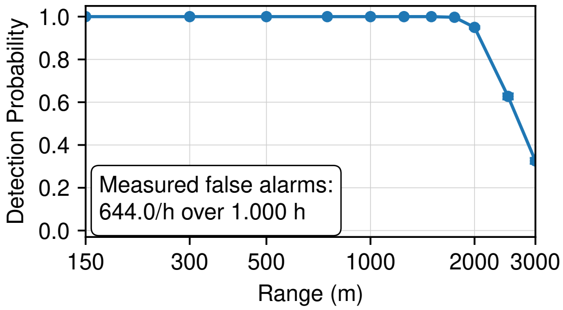

Fig. 3. Calibrated non-overlapping mean-energy detection versus range at a frozen threshold. Bands are nominal 95% Wilson intervals. Empirical detection probability is 1.0 through 1.5 km and collapses between 2 and 3 km as direct-path received power falls from −136.3 to −139.8 dBm.

Two properties distinguish this from an SNR sweep. The independent variable is range in meters: the knee follows from Eqs. (2)–(3) and the stated transmit power, so changing the assumed EIRP moves it predictably. And the false-positive rate is *measured* on signal-absent data at the same frozen threshold rather than assumed. The threshold targets 10^{-2} per window (nominal 703.1 windows/h); the observed 644 corresponds to $p = 0.00916$. This is an energy detector on a tone, calibrated and tested against noise from the same model, reported at one operating point rather than as a ROC curve; it establishes range-conditioned evaluation in physical units, not that a UAV is detectable at 3 km in practice.

D. Real–Synthetic Feature Comparison

We now ask how far synthetic waveform morphology sits from real measured UAV RF data, comparing all 14,484 released DroneRFb-Spectra arrays across 24 classes against 2,560 synthetic records from 8 families. The released arrays are normalized RF-derived features, not raw IQ, so the comparison is restricted to feature *morphology*.

Synthetic receiver IQ passes through the same time–frequency transform as the real path. For FFT length N, hop H, window $w[n]$, and percentiles q_5, q_{99},

$$Y_i(t,k) = \textstyle\sum_{n=0}^{N-1} y_i[tH+n]\, w[n]\, e^{-j2\pi kn/N}, \tag{4}$$

$$S_i(t,k) = 20\log_{10}\max\big(|Y_i(t,k)|, \epsilon\big), \tag{5}$$

$$\widetilde{S}_i = \mathrm{clip}\big((S_i - q_5)/(q_{99}-q_5), 0, 1\big), \tag{6}$$

$$m_i^t(t) = \tfrac{1}{F}\textstyle\sum_k \widetilde{S}_i(t,k), \quad m_i^f(k) = \tfrac{1}{T}\textstyle\sum_t \widetilde{S}_i(t,k). \tag{7}$$

Distances are median pairwise RMSE over concatenated marginals. The transform uses $N = 2048$, $H = 1024$ and the repository S3R window and floor ϵ, resampling to 512×512; pair sampling is seeded (20260828) and capped at 2048 pairs.

A distance is meaningless without a baseline. We therefore report two: the real–real distance between disjoint splits of the measured dataset, and the synthetic–synthetic distance within the generated dataset (Table III). The cross-domain median is 0.245, versus 0.063 real–real, *i.e.*, a factor of 3.9. Generic waveform families are *not* interchangeable with measured captures. The released arrays identify no originating recording or airframe, so one capture may straddle the real–real split and depress that baseline; the synthetic baseline pools

TABLE III
MARGINAL-DISTANCE COMPARISON, ALL 14,484 REAL ARRAYS VS. 2,560 SYNTHETIC RECORDS. MEDIAN PAIRWISE RMSE, WITH THE 5TH–95TH PERCENTILE SPREAD WHERE A PAIRWISE DISTRIBUTION WAS RETAINED; FAMILY-RESTRICTED ROWS ARE MEDIANS OF THE PER-CLASS MATRIX AND CARRY NO PAIRWISE SPREAD.

Pairing	Median RMSE	p05–p95
Real vs. real (baseline)	0.063	0.019–0.158
Synthetic vs. synthetic (baseline)	0.222	0.126–0.322
Real vs. synthetic (all families)	0.245	0.127–0.387
Real vs. synthetic (OFDM only)	0.174	—
Real vs. synthetic (FM only)	0.305	—

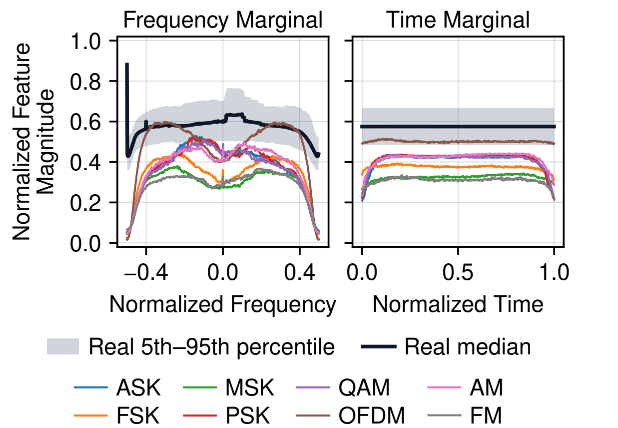

Fig. 4. Frequency and time marginals. The real median and 5th–95th percentile band are compared with the median of each synthetic family. OFDM alone reproduces the broad, flat occupancy of the real band; FM and MSK sit far below it across the whole band.

across families and is correspondingly inflated. The synthetic dataset's own internal spread is comparable (0.222) and itself 3.5× the real dataset's, though pairwise RMSE distributions are not additive, so this indicates rather than decomposes the sources of the gap. Synthetic generation is not a drop-in replacement for real IQ sampling.

The ranking, however, is informative. Figure 5 shows the full 24×8 distance matrix; pooling it by family gives a consistent ordering: OFDM 0.174, AM 0.240, PSK 0.243, ASK 0.248, QAM 0.248, FSK 0.254, MSK 0.293, FM 0.305. OFDM is closest for all 24 classes individually and FM farthest for all 24. No label supervises this ordering; it emerges from the morphology metric and is consistent with the wideband OFDM structure of the DJI video downlinks and 2.4 GHz control links dominating this dataset. Marginals discard joint time–frequency structure and per-record normalization removes absolute power, so this is evidence of similar spectral occupancy, not of modulation identification. Figure 4 shows why: OFDM alone tracks the broad, flat occupancy of the real band (mean magnitude 0.48 vs. the real median's 0.57) while FM and MSK sit far below at 0.29. Restricting the comparison to OFDM brings the cross-domain median to 0.174, within 2.8× of the real–real baseline.

E. Throughput and Reproducibility

A direct NumPy benchmark generates one simulated second as twenty 50 ms chunks at 5 MS/s. Median receiver through-

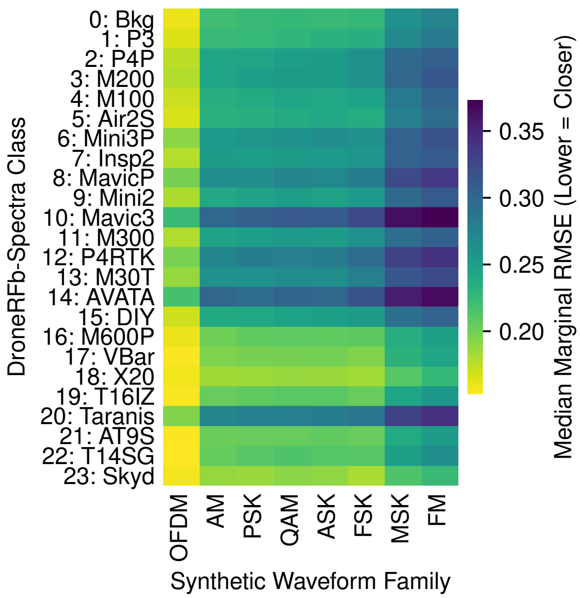

Fig. 5. Median marginal RMSE for every DroneRFb-Spectra class against every synthetic waveform family, families ordered by pooled median. The banding is the result: the family ordering is the same for all 24 classes, with OFDM nearest and FM farthest in every column. Each cell is the median over deterministically sampled record pairs.

put is $4.49\,$MSamples/s (8.97 aggregate across links) with $81.6\,$MiB peak memory on commodity CPU hardware without GPU acceleration, and SigMF export sustains $117\,$MB/s. The median real-time factor of 0.897 (95% CI [0.876, 0.910]) means generation runs 10.3% slower than real time on one core, so regenerating a dataset costs roughly its own duration. Repeated SigMF exports of the same scenario and schedule produce identical TX/RX data hashes, and every figure and table above is bound to a content-hashed artifact manifest.

V. Conclusion

Public UAV RF datasets are captured tens of meters from the emitter, while operational counter-UAS systems must detect and classify at kilometers. *RF-UAVSim* makes that long-range, low-SNR regime available for dataset generation: an absolute-power link budget sets received SNR from geometry, and an index-addressed channel makes receiver IQ a pure function of the scenario. On that basis we measured a range-conditioned detection curve spanning $26.0\,$dB and bounded synthetic fidelity against the full 14,484-array DroneRFb-Spectra dataset, where generic families prove not to be substitutes for measured captures. The recognition matrix this framework was built to support—domain shift, long-range augmentation, and open-set behavior at standoff—is immediate future work.

VI. Acknowledgements

The participation of K. Kapoor and J. Surajballi was supported by an NSF CISE REU Supplement administered by the Computing Research Association under NSF-CSR-2534286. H. Wang was supported in part by grants NSF-CSR-2534286 and NSF-MSI-2523997. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory (AFRL) or the U.S. Government. This material has been cleared for public release, unlimited distribution (PA Case Number: AFRL-2026-4063, 15-Sep, 2026).

References

[1] Defense Innovation Unit, "Counter-UAS Sensing for Homeland and Mobile Defense, Commercial Solutions Opening," 2026.

[2] D. Aouladhadj, E. Kpre, V. Deniau, A. Kharchouf, C. Gransart, and C. Gaquière, "Drone Detection and Tracking Using RF Identification Signals," *Sensors*, vol. 23, no. 17, p. 7650, 2023.

[3] N. Yu, S. Mao, C. Zhou, G. Sun, Z. Shi, and J. Chen, "DroneRFa: A Large-Scale Dataset of Drone Radio Frequency Signals for Detecting Low-Altitude UAVs," *Journal of Electronics & Information Technology*, vol. 46, no. 4, 2024.

[4] S.-K. Han and Y.-H. Jung, "A Time-Synchronized Multi-Sensor Drone Dataset Acquired From Multiple Radars and RF Receiver," *Scientific Data*, 2026.

[5] H. Elyousseph and M. Altamimi, "Robustness of Deep-Learning-Based RF UAV Detectors," *Sensors*, vol. 24, no. 22, p. 7339, 2024.

[6] N. Gao, T. Zeng, B. Chen, D. Cai, S. Jin, and M. Matthaiou, "Multi-Domain Supervised Contrastive Learning for UAV Radio-Frequency Open-Set Recognition," *IEEE Journal on Selected Areas in Communications*, vol. 44, pp. 4083–4098, 2026.

[7] R. Shi, X. Yu, S. Wang, Y. Zhang, L. Xu, P. Pan, and C. Ma, "RFUAV: A Benchmark Dataset for Unmanned Aerial Vehicle Detection and Identification," arXiv:2503.09033, 2025.

[8] S. Glüge, M. Nyfeler, A. Aghaebrahimian, N. Ramagnano, and C. Schüpbach, "Robust Low-Cost Drone Detection and Classification Using Convolutional Neural Networks in Low SNR Environments," *IEEE Journal of Radio Frequency Identification*, vol. 8, pp. 821–830, 2024.

[9] M. Rostami, A. Faysal, H. Xia, H. Kasasbeh, Z. Gao, and H. Wang, "CageDroneRF: A Large-Scale RF Benchmark and Toolkit for Drone Perception," arXiv:2601.03302, 2026.

[10] G. Reus-Muns and K. R. Chowdhury, "Classifying UAVs With Proprietary Waveforms via Preamble Feature Extraction and Federated Learning," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 7, pp. 6279–6290, 2021.

[11] Y. Shi, "DroneRFz and DroneRFB: An Extension RF Spectrogram Dataset for Drone Recognition Across Diverse Electromagnetic Environments," IEEE DataPort, 2026.

[12] M. S. Allahham, M. F. Al-Sa'd, A. Al-Ali, A. Mohamed, T. Khattab, and A. Erbad, "DroneRF Dataset: A Dataset of Drones for RF-Based Detection, Classification and Identification," *Data in Brief*, vol. 26, p. 104313, 2019.

[13] M. Ezuma, F. Erden, C. K. Anjinappa, O. Ozdemir, and I. Guvenc, "Drone Remote Controller RF Signal Dataset," IEEE DataPort, 2020.

[14] C. J. Swinney and J. C. Woods, "RF Detection and Classification of Unmanned Aerial Vehicles in Environments with Wireless Interference," in *Proc. Int. Conf. Unmanned Aircraft Systems (ICUAS)*, 2021.

[15] O. O. Medaiyese, M. Ezuma, A. P. Lauf, and I. Guvenc, "Hierarchical Learning Framework for UAV Detection and Identification," *IEEE Journal of Radio Frequency Identification*, vol. 6, 2022.

[16] N. Yu, J. Wu, C. Zhou, Z. Shi, and J. Chen, "Open Set Learning for RF-Based Drone Recognition via Signal Semantics," *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 9894–9909, 2024, dataset: IEEE DataPort, DOI 10.21227/wv7h-sv64.

[17] T. Zhao, B. W. Domae, C. Steigerwald, L. B. Paradis, T. Chabuk, and D. Cabric, "Drone RF Signal Detection and Fingerprinting: UAVSig Dataset and Deep Learning Approach," in *Proc. IEEE MILCOM*, 2024.

[18] S. Scholl, "RF Signal Classification With Synthetic Training Data and Its Real-World Performance," arXiv:2206.12967, 2022.

[19] L. Boegner, M. Gulati, G. Vanhoy, P. Vallance, B. Comar, S. Kokalj-Filipovic, C. Lennon, and R. D. Miller, "Large Scale Radio Frequency Signal Classification," arXiv:2207.09918, 2022.

[20] B. Hilburn, N. West, T. O'Shea, and T. Roy, "SigMF: The Signal Metadata Format," in *Proc. GNU Radio Conference*, 2018.