Signal Atlas cuas-v1: A Synthetic Counter-UAS Radio-Frequency Dataset and Its Technical Validation

Superpose Inc.
signal-atlas@superpose.us

Abstract

We present **Signal Atlas cuas-v1**, a fully synthetic radio-frequency (RF) dataset of drone-emitted signals and co-band interference for counter-UAS (uncrewed aircraft systems) detection and RF foundation-model pretraining, and we report its technical validation honestly, including where it succeeds and where it does not. *What is strongly validated:* each signal reproduces its published physical-layer parameters, measured over thousands of realizations with tight error bars (e.g. DJI DroneID occupied bandwidth 8.937 ± 0.006 MHz, frame duration 643.229 ± 0.000 μs, Zadoff-Chu sync correlation 0.998 ± 0.001 over $n{=}5000$); and an *independent* OFDM receiver, built only from published constants and never importing the generator, blindly acquires the synthetic DroneID, locking onto both sync symbols at their exact positions (peak-to-sidelobe $29\times$), identifying both Zadoff-Chu roots, and recovering a clean QPSK constellation (self-referential EVM 0.7%). The same independent matched filter, run on a *real* DJI capture, fires on a burst at a DroneID channel (peak-to-sidelobe $18.5\times$), though the full two-root signature was not confirmed, a promising but inconclusive cross-check. *What is only partially validated, measured against real captures (RFUAV):* the synthetic frequency-hopping bandwidth matches real (66 vs 69 MHz), but spectral fine structure correlates only weakly (0.27–0.70) and the single-burst OcuSync surrogate is far narrower than a real drone's hopping multi-link footprint. *What fails:* a classifier trained purely on the synthetic data does **not** transfer to real drone captures (accuracy 0.33, at chance; real drone windows are classified as noise). We therefore scope the dataset to spec-faithful pretraining and benchmarking, not drop-in real-world detector training, and identify capture-augmentation as the required next step. All results are seed-deterministic and reproducible.

1 Background & Summary

Counter-UAS systems detect drones by passively sensing the RF link between drone and controller; the field is moving toward machine-learning detection over raw spectrum. Such models need large, diverse, labeled RF data, which is scarce. **cuas-v1** is a synthetic dataset addressing this: five clean-room drone emitters plus interference, in a single-signal classification track and a wideband scene track with per-signal time-frequency bounding boxes, over ten drone and four interference classes. It is *synthetic-only* by design.

This document is a technical validation whose purpose is to let a reader trust the data *for a stated use* and to make its limits impossible to miss. A credible validation must be willing to return a negative result; this one does. We validate at three levels of increasing stringency, spec conformance, independent decode, and real-world transfer, and report each outcome as found.

Superpose Inc. Signal Atlas cuas-v1 technical validation report. Generated with the `rfgen` framework; fully reproducible.

Table 1: Fidelity: measured value vs. published specification. Exact-by-construction quantities are marked (E); others are mean $\pm\ \sigma$ over n realizations.

Signal	Property	Measured	n	Published
DroneID	subcarrier spacing (Hz)	15000 (E)		15000
DroneID	frame duration, 9-sym (μs)	643.2 (E)		$\sim$643
DroneID	ZC sync corr., 9-sym	1.000	2000	1.0
DroneID	ZC sync corr., 8-sym	1.000	2000	1.0
DroneID	occupied bandwidth (MHz)	8.937 ± 0.006	5000	$\sim$9
FrSky D16	GFSK deviation (kHz)	57.098 ± 0.008	3000	57.1
FrSky D16	per-hop duty (%)	5.89 ± 0.00	400	2–20
DSMX	per-hop duty (%)	3.89 ± 0.01	400	2–20
Analog NTSC	occupied bandwidth (MHz)	16.94 ± 0.05	3000	16–22
Analog NTSC	color subcarrier (MHz)	3.5795 (E)		3.5795
OcuSync 20	occupancy (MHz)	19.60 ± 0.02	5000	20

2 Data Records

Delivered as HDF5 shards (numeric IQ) with sibling JSON Lines provenance. Track A: fixed 16,384-sample complex-baseband windows (matching the sibling comms-v1 dataset for joint pre-training), per-class sample rate. Track B: $\sim$80 MHz scenes with a bounding box per signal. Ten drone classes (DroneID $\times 2$, FHSS RC links $\times 3$, analog FPV video $\times 2$, Remote ID, OcuSync surrogate $\times 2$) and four interference classes. Every record carries a reproduction seed and an `iq_sha256` binding provenance to the stored IQ.

3 Methods

Each emitter is synthesized clean-room from published specifications, DroneID from the NDSS 2023 description [5] and the MIT reference implementation [4] (the AGPL code is never consulted); FHSS RC links from register-verified datasheet values; analog video from ITU-R BT.470-7 [1]; Remote ID from ASTM F3411 [2]. The OcuSync class is an explicitly-labeled wideband-OFDM *surrogate*, as the real link is proprietary and encrypted. A drawn SNR is applied via an additive white Gaussian noise channel; generation is seed-deterministic and byte-reproducible.

4 Technical Validation

4.1 Fidelity: spec conformance

Each fidelity metric is measured over many clean realizations (Table 1). We distinguish two kinds of quantity. Some are *exact by construction*, subcarrier spacing, frame duration, and the ZC sync correlation are fixed by the deterministic frame structure and show no spread; we report them as exact, not as error bars. Others, occupied bandwidth and (across seeds) the hop layout, carry genuine payload-driven spread, reported as mean $\pm\ \sigma$. Both DroneID frame variants (9- and 8-symbol) validate: measured at the correct per-variant ZC positions, the sync correlation is 1.000 for each. The underlying DSP (Zadoff-Chu, OFDM, FM/GFSK, QPSK) was independently verified numerically.

4.2 Independent decode of DroneID

Correlating a signal against the same math that made it is partly circular. The stringent test is an *independent* receiver. We built a textbook OFDM receiver that imports none of the generator and uses only published constants (15 kHz spacing, 1024 FFT, 601 subcarriers, ZC roots via the public closed form), and ran it on synthetic DroneID with added carrier-frequency offset and noise.

On a clean synthetic frame it **blindly acquires the signal**: a matched filter locks onto both sync symbols at their exact sample positions (3368, 5560; spacing $2192 = 2 \times (1024{+}72)$ exactly), with peak-to-sidelobe ratio $29\times$; it identifies both Zadoff-Chu roots unambiguously (correlation 1.000 at root 600 and 1.000 at root 147; 0.04 at wrong roots); and it recovers a clean QPSK constellation at 0.7% error-vector magnitude (self-referential, sliced from the received symbols, not a payload/BER

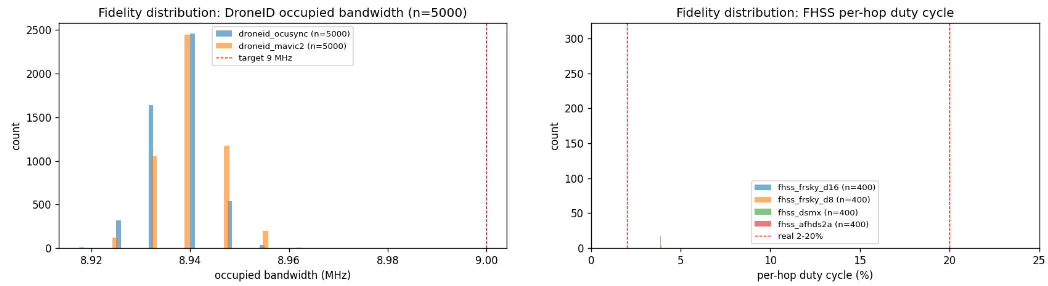

Figure 1: Fidelity distributions. Left: DroneID occupied bandwidth over $n{=}5000$ realizations (tight around ∼9 MHz). Right: FHSS per-hop duty cycle across protocols, all within the real 2–20% range.

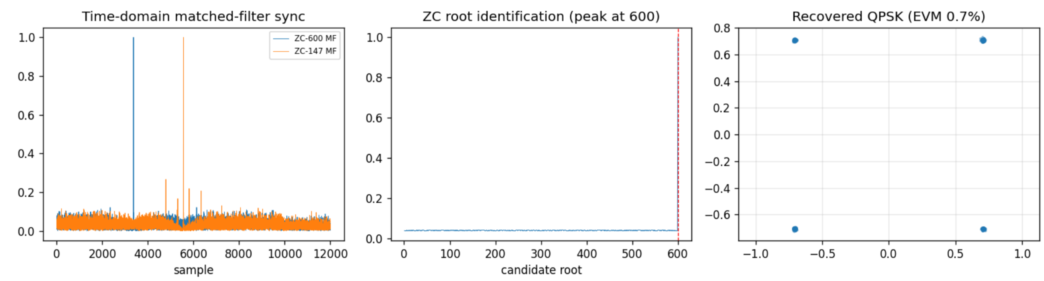

Figure 2: Independent DroneID receiver (uses only published constants) on the *synthetic* frame. Left: matched-filter sync, sharp peaks at the two ZC symbols. Middle: root identification, peaking at 600. Right: recovered QPSK constellation (0.7% self-referential EVM).

decode; Figure 2). A single-point check at a 2.5 kHz offset (∼1 ppm) and 20 dB noise leaves sync acquisition intact (peak-to-sidelobe $28\times$); a full detection-rate curve over realistic 10–40 ppm oscillator error is future work.

This establishes that the synthetic frame is structurally consistent with the published DroneID spec and independently decodable, but the receiver and the emitter both derive from the same specification, so this is a self-consistency check, not a match to reality. As a stronger cross-check we ran the identical independent matched filter on a *real* DJI MINI4 PRO capture: on DroneID channel 2.4145 GHz the ZC-600 filter fires on a single burst (at 942 ms) with peak-to-sidelobe $18.5\times$, well above the other channels (∼8–$9\times$). However, the second ZC root (147) was not confirmed at the expected symbol spacing, so this is a *promising but inconclusive* real-DroneID detection. A full real-DroneID decode (with timing/CFO recovery on the real signal) is the highest-priority remaining validation.

4.3 Realism: quantitative comparison to real captures

We downloaded real *raw IQ*, a FlySky X14 controller (the RFUAV "FRSKY X14" archive is labeled FlySky in its own metadata, so we compare it against the synthetic FlySky/AFHDS-2A emitter) and a DJI MINI4 PRO OcuSync drone from RFUAV [3], both at 100 MHz, and computed distributional distances (Table 2, Figure 4). The result is mixed and reported as found: the synthetic hopping *bandwidth* matches real closely (66 vs 69 MHz), but the spectral fine structure correlates only weakly (0.27–0.70; full-band log-PSD Pearson is noise-floor sensitive and should not be over-read) and the amplitude distributions differ. Most tellingly, a real DJI drone occupies ∼64 MHz, it hops its video channel and runs a control link and DroneID across the band, whereas the single static 10 MHz OcuSync *surrogate* does not. The synthetic signals capture the coarse spectro-temporal signature, not the fine structure or the full-scene complexity of real drone RF.

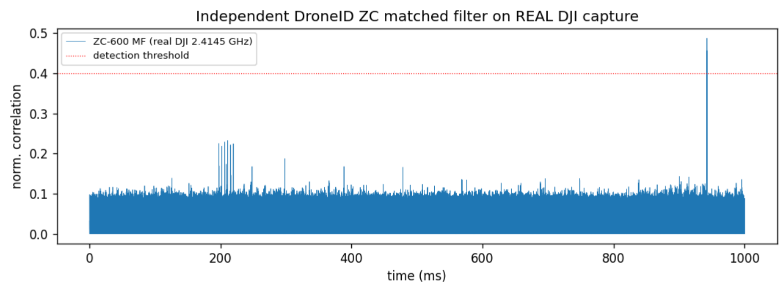

Figure 3: The same independent ZC matched filter on a *real* DJI MINI4 PRO capture (channel 2.4145 GHz): a single burst detection at 942 ms (peak-to-sidelobe $18.5\times$). Promising but inconclusive, the second ZC root was not confirmed; a full real-DroneID decode is the top remaining test.

Table 2: Realism: distributional distances, synthetic vs. real RFUAV IQ.

Pair	Spectral corr. (log-PSD)	PSD Wass. (MHz)	Occ. BW synth/real (MHz)	Amp. Wass.
synthetic AFHDS-2A vs real FlySky	0.27	9.30	66 / 69	0.61
synthetic OcuSync vs real DJI	0.70	2.54	9.9 / 64	0.45

4.4 Transfer: train on synthetic, test on real (the decisive test)

The question that matters for deployment is whether a model trained on synthetic data recognizes *real* drones. We trained a classifier on synthetic FHSS / OcuSync / noise spectrograms and tested it on real RFUAV FlySky and DJI captures. **It does not transfer:** *real-drone recall is* ≈ 0, the model collapses essentially every real drone window to the "noise" class (Figure 5); the 0.33 overall accuracy comes only from correctly labeling the in-distribution synthetic noise. This held across two feature and window-length settings. This is a floor, not a ceiling: the synthetic training is impairment-free, the learner is a plain random forest on coarse spectrograms, and the negative class is synthetic AWGN rather than real drone-free spectrum. It proves that *clean synthetic plus a weak learner does not transfer*, consistent with the fine-structure and full-scene gaps above, not that the data can never transfer. Closing this gap (impairments, capture-augmentation, a stronger baseline, real out-of-band noise) is the v2 acceptance gate.

4.5 Independent adversarial review, robustness, reproducibility

Six independent adversarial reviews (mathematical fidelity, robustness, literature grounding, construct validity, methodology, empirical realism), each instructed to refute, found and led to the correction of real defects before release, including a duty cycle off by an order of magnitude (caught by the real-data comparison), a bounding box $2.3\times$ too long, and a silent all-zero corner case. Emitters enforce Nyquist-correct guards; sample-rate sweeps hold the DroneID subcarrier spacing within 8 Hz of 15 kHz across 9–500 MHz; all output is seed-deterministic with `iq_sha256` provenance, so every number here is independently reproducible.

5 Usage Notes

Validated uses. (i) Spec-faithful signal synthesis and classification of *synthetic* signals; (ii) representation pretraining where exact real-world transfer is not the immediate objective; (iii) benchmarking and label-rich scene composition. The independent decode makes DroneID trustworthy at the waveform level.

Not validated / negative results. (i) **Direct real-world drone detection: a synthetic-trained classifier does not transfer to real captures (chance-level).** Do not field a detector trained on cuas-v1 alone without real-capture augmentation or domain adaptation. (ii) Realism is coarse: bandwidth matches for FHSS, but fine spectral structure and the full hopping multi-link footprint of a real drone are not reproduced. (iii) The OcuSync class is a labeled single-burst surrogate, narrower than real

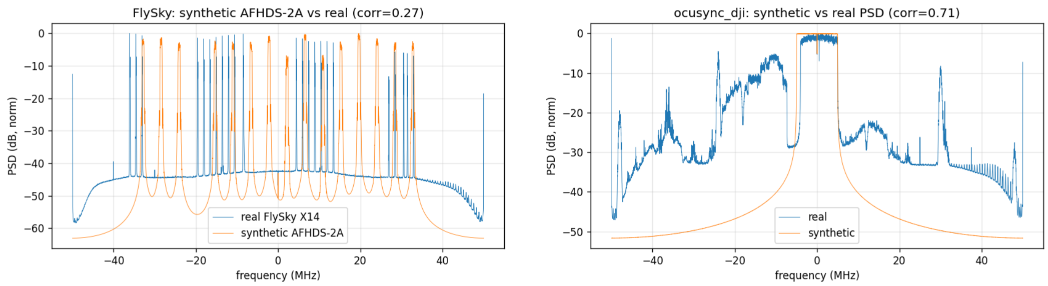

Figure 4: Synthetic vs. real power spectral density. Left: FHSS vs real FrSky (bandwidth matches, fine structure differs). Right: OcuSync surrogate vs real DJI (the real drone's hopping multi-link footprint is far wider than the single synthetic burst).

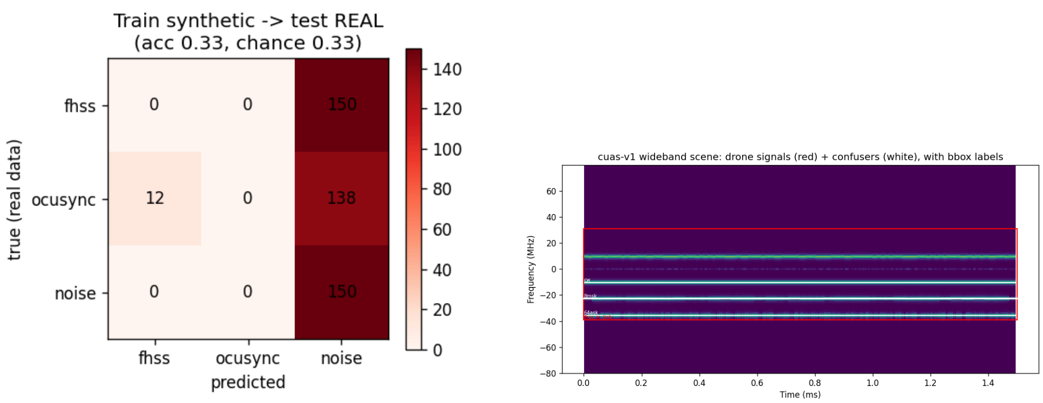

Figure 5: Left: train-synthetic $\rightarrow$ test-real confusion, real drone signals (rows fhss, ocusync) are classified as noise; transfer is at chance. Right: a wideband scene with per-signal bounding boxes (labels are internally correct).

drone RF. (iv) No device fingerprint; synthetic payloads; Remote-ID-over-BLE is not a PHY-only detection target; the interference class lacks real 802.11/BLE structure.

Remaining gaps / v2 roadmap (ranked, from an independent staff-engineer review): (1) a full decode of a *real* DroneID frame (the synthetic decode is self-consistent with the spec but not yet matched to reality); (2) re-run the transfer test with hardware impairments, a stronger baseline, and real out-of-band noise, as the acceptance gate; (3) add measured impairments (phase noise, IQ imbalance, PA nonlinearity, realistic CFO, multipath), the mechanistic cause of the domain gap; (4) a hopping, multi-link OcuSync model and validation of the wideband-scene track against real multi-link captures; (5) per-device parameter spread drawn from datasheets/captures; (6) structured 802.11/BLE interference classes.

This validation was itself independently red-teamed as a staff RF engineer, whose verdict was *adopt for the scoped uses above, do not field a synthetic-trained detector*; the corrections that review surfaced (a FlySky/FrSky capture mislabel, an under-reported second frame variant, and several framing over-reaches) are incorporated here.

Data and Code Availability

Generation package, per-emitter reports, experiment scripts, and figures are in the `rfgen` repository under `use_cases/signal-atlas/cuas-v1/`. Fully reproducible from the package version and index ranges.

References

[1] BT.470-7: Conventional analogue television systems. Technical report, International Telecommunication Union (ITU-R), 2005.

[2] ASTM International. ASTM F3411: Standard specification for remote ID and tracking, 2022.

[3] RFUAV Contributors. RFUAV: A benchmark dataset for unmanned aerial vehicle detection and identification. *arXiv preprint arXiv:2503.09033*, 2025.

[4] David Protzman. dji_droneid: An open-source DJI DroneID implementation. `https://github.com/proto17/dji_droneid`, 2022. MIT License.

[5] Nico Schiller, Merlin Chlosta, Moritz Schloegel, Nils Bars, Thorsten Eisenhofer, Tobias Scharnowski, Felix Domke, Lea Schönherr, and Thorsten Holz. Drone security and the mysterious case of DJI's DroneID. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2023. ndss2023_f217.