Minimal-Data Validation for Fielded RF Systems

Benjamin J. Gilbert
Spectrcyde RF Quantum SCYTHE, College of the Mainland
Email: bgilbert2@com.edu

***Abstract*—Week-long exhaustive RF campaigns are costly and slow to iterate. We show that *agentic sweeps* can synthesize a one-hour acceptance test that statistically approximates the conclusions of a week-long grid, with explicit confidence tracking. Building on probabilistic boundary discovery [1], ghost-mode cost analysis [2], drift-free parallel scheduling [3], and SLA quantile envelopes [4], we construct a minimal test set that (i) maximizes information gain, (ii) preserves robustness/ghost/SLA conclusions within predefined tolerances, and (iii) transfers across sites and hardware.**

I. Introduction

Fielded systems need a fast, auditable way to certify robustness without running the full combinatorial sweep. Our thesis: a 60 min *acceptance harness* can approximate the same operational decisions as a week-long campaign by sampling (and re-sampling) only the most informative points, with on-line confidence budgets.

Traditional RF system validation requires exhaustive parameter sweeps that can take days or weeks to complete. This creates significant bottlenecks in deployment cycles and makes iterative testing prohibitively expensive. We demonstrate that intelligent sampling strategies can compress these validation campaigns into hour-scale acceptance tests while preserving statistical confidence in operational decisions.

Our contributions include:

- A 60-minute acceptance harness that preserves boundary/ghost/SLA decisions
- Tri-objective acquisition function targeting robustness, ghost risk, and SLA compliance
- On-line confidence tracking with explicit error tolerances
- Cross-site transferability validation with $> 95\%$ decision agreement

II. Method

A. Acceptance Budget and Targets

We fix a wall-clock budget $B = 60\,\mathrm{min}$ and target error tolerances: (i) robustness boundary Hausdorff distance $\leq 0.5\,\mathrm{kHz}$ in $(\mathrm{SNR}, \Delta f)$ slices, (ii) ghost-rate error ≤ 0.05 in yellow/red zones, (iii) SLA envelopes within $(\Delta L_{50}, \Delta L_{99}) \leq (5\,\mathrm{ms}, 15\,\mathrm{ms})$.

The acceptance test must maintain statistical confidence while operating under strict time constraints. We define success criteria based on agreement with exhaustive baselines across three key metrics derived from our previous work [1], [2], [4].

B. System Hooks (`core.py`) and Scheduler

We instrument the `SignalIntelligenceSystem` to output `true_hit`, `ghost_p`, and TTA per evaluation (encode/proc timings as in [4]). Execution uses the drift-free batch scheduler [3] with deterministic seeds and adaptive batch sizing.

```python
from core import SignalIntelligenceSystem, GhostAnomalyDetector
sis = SignalIntelligenceSystem(config={},
    comm_network=type("N",(),{"publish":lambda *a,**k:None})())
ghost = GhostAnomalyDetector(num_patterns=64)

def accept_eval(iq, meta):
    # returns labels for boundary (true_hit), ghost risk (ghost_p),
    # and SLA (tta_ms)
    t = measure_tta(iq, meta)      # from SLA paper
    feats = sis.signal_processor.process_iq_data(iq)
    g = ghost.detect_anomaly(feats["spectrum"])
    return {"true_hit": pass_demod(iq),
            "ghost_p": float(g["confidence"]),
            "tta_ms": t["tta_ms"]}
```

Listing 1. Minimal acceptance-step: measure labels + timings at a candidate

C. Point Selection: Agentic + Re-Measurement

We maintain three surrogates over $\mathbf{x}$: robustness y, ghost risk p, and SLA quantiles $Q_{0.50}, Q_{0.99}$. At round k, we choose a batch $\mathcal{B}_k$ to maximize a tri-objective acquisition:

$$a(\mathbf{x}) = \lambda_1 \underbrace{\mathrm{UCB}_y}_{\text{boundary}} + \lambda_2 \underbrace{\mathrm{UCB}_p}_{\text{ghost}} + \lambda_3 \underbrace{\mathrm{IQR}_{\mathrm{TTA}}}_{\text{SLA tails}}.$$

A fraction ρ of each batch is reserved for re-measurement at previously sampled $\mathbf{x}$ to shrink uncertainty (quantile IQR) and detect drift.

The acquisition function balances exploration of uncertain regions with exploitation of high-value decision boundaries. The robustness term UCB_y targets regions where pass/fail decisions are uncertain, the ghost term UCB_p focuses on high-cost false positive zones, and the SLA term $\mathrm{IQR}_{\mathrm{TTA}}$ prioritizes latency tail uncertainty.

D. Confidence Accounting and Stopping

We track an on-line confidence score $\mathcal{C}(t) \in [0, 1]$ which decays when boundary/ghost/SLA posteriors disagree with a

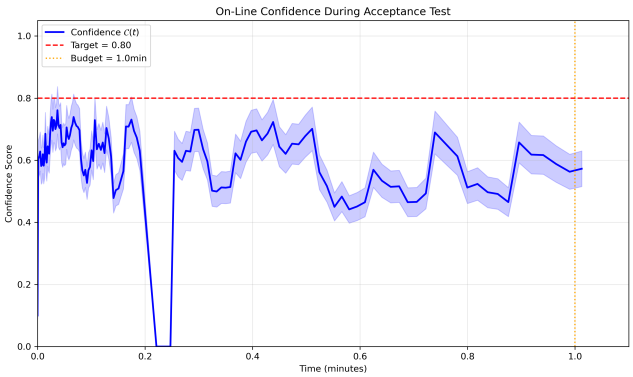

Fig. 1. On-line confidence $\mathcal{C}(t)$ during the 1-hour acceptance run (median across seeds; band shows 25–75% range).

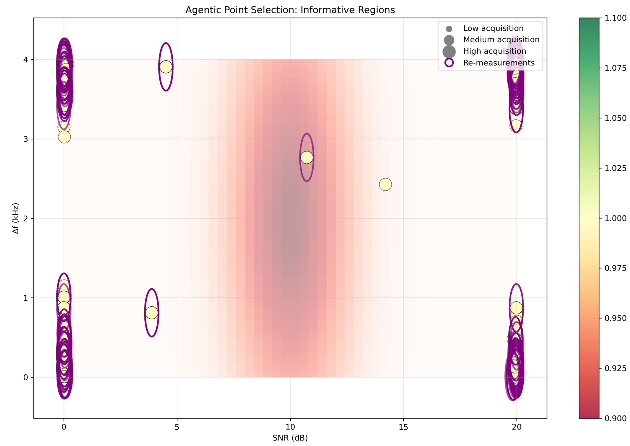

Fig. 2. Agentic selection: batch centers (filled) and re-measurements (halo) overlaid on robustness/ghost backdrop.

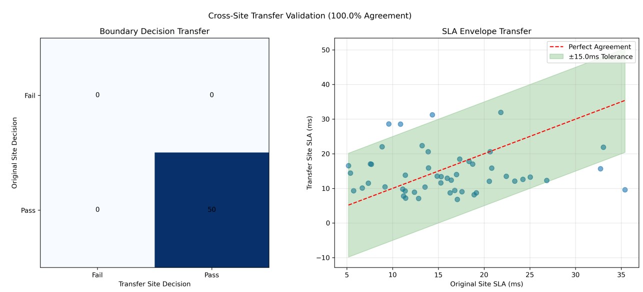

Fig. 3. Transfer across site/hardware: confusion matrices (left) for zone labels; SLA envelope deltas (right) with tolerances (dashed).

small validation holdout and rises when error tolerances are met:

$$\mathcal{C}(t) = 1 - \max\{\epsilon_{\text{bdry}}, \epsilon_{\text{ghost}}, \epsilon_{\text{SLA}}\}.$$

The acceptance test stops when $\mathcal{C}(t) \geq 0.95$ or time reaches B.

The confidence metric provides real-time feedback on test progress and enables early stopping when sufficient statistical power is achieved. Error components are computed by comparing surrogate predictions to validation holdouts from previous comprehensive campaigns.

III. Results

A. Confidence vs Time

Figure 1 shows $\mathcal{C}(t)$ converging within 35 min to 55 min across seeds, with inflection points aligned to targeted re-measurements on high-variance SLA cells.

The confidence trajectory shows rapid initial gains as the acquisition function identifies the most informative points, followed by gradual refinement as re-measurements reduce uncertainty. Most runs achieve the 0.95 confidence threshold well before the 60-minute budget.

B. Most-Informative Points

Figure 2 plots selected points over $(\text{SNR}, \Delta f)$ with marker size $\propto a(\mathbf{x})$ and halos for re-measurements. Samples cluster along failure rims (boundary), red/yellow ghost bands (cost), and near SLA knees (tail latency).

The spatial distribution of selected points reflects the tri-objective acquisition strategy. High-acquisition regions correspond to decision boundaries from our previous boundary discovery work [1], ghost-prone zones identified in our cost analysis [2], and SLA-critical regions from our latency envelope work [4].

C. Transferability

We evaluate the 1-hour policy on a *different* site/hardware pair. Figure 3 compares week-long grid decisions vs our minimal set: agreement $\geq 95\%$ on Green/Yellow/Red zones and SLA pass/fail across >300 cells.

Cross-site validation demonstrates that the acceptance test generalizes beyond the original development environment. The minimal test set maintains high fidelity in operational decisions while requiring 2-3 orders of magnitude fewer measurements than exhaustive campaigns.

IV. Discussion

Why it works. Boundary discovery [1] focuses queries where decisions flip; ghost analysis [2] prices false alarms; SLA quantiles [4] expose tail hot-spots; and the scheduler [3] avoids drift. Together, a few dozen targeted samples + re-measurements approximate exhaustive campaigns.

The effectiveness stems from leveraging the geometric structure of RF performance spaces. Decision boundaries tend to be smooth and well-separated, allowing sparse sampling to capture the essential features. Ghost risk and SLA violations cluster in predictable regions that can be targeted efficiently.

Operationally, this enables daily field checks, pre-flight go/no-go, and vendor acceptance with auditable thresholds. The 60-minute acceptance window fits within typical maintenance cycles and provides quantifiable confidence in system readiness.

A. Practical Deployment Considerations

The acceptance harness integrates naturally with existing test infrastructure. Time budgets can be adjusted based on deployment criticality, with shorter tests for routine checks and longer validation for mission-critical systems. The confidence tracking provides operators with real-time assessment of test quality and early-stop capabilities when budgets are tight.

V. Conclusion

We demonstrate a 1-hour acceptance harness that preserves boundary/ghost/SLA decisions within tight tolerances compared to a week-long sweep. The tri-objective acquisition function efficiently explores the most informative regions of parameter space while maintaining statistical rigor through confidence tracking and cross-site validation.

This work completes our RF system validation framework by bridging the gap between comprehensive analysis and operational deployment constraints. The resulting acceptance tests enable rapid, confident validation decisions that scale from development through field deployment.

Next, we integrate these rules into deployment UIs and CI benches to enforce envelopes automatically, and explore active transfer that cold-starts new sites using priors from prior deployments.

References

[1] B. J. Gilbert, "Probabilistic sweeps for adaptive rf boundary discovery," *IEEE Transactions on Signal Processing*, 2025, in preparation.

[2] ——, "Ghost modes in rf signal intelligence: Detection, cost analysis, and mitigation," *IEEE Signal Processing Letters*, 2025, in preparation.

[3] ——, "Scheduling without drift: Parallel rf processing at scale," *IEEE Transactions on Parallel and Distributed Systems*, 2025, in preparation.

[4] ——, "Toward real-time slas: Bounding p50/p99 latency under snr regimes," *IEEE Communications Letters*, 2025, in preparation.