Comparing radio-signal searches on Breakthrough Listen data

Ryan Clinton. Version 1.7, 2026-11-24.

Summary

I compared a narrowband radio-signal search of my own (SignalSearchYoke) with two published pipelines, turboSETI and BLISS, on recorded Breakthrough Listen observations from the Green Bank Telescope, by planting synthetic signals in the data (13,310 in the 2017 session, 13,310 in the 2016 one, 7,040 in the 2019 one) and counting how many each pipeline found, and how many false candidates each raised, on the same blocks.

On a 2017 session (KIC8462852), at the cut where each pipeline's measured false-candidate rate is at most 0.1 a block (a point estimate; each 95% interval reaches above it, and the cuts were chosen on the blocks they were judged on), this search found 4,520 of the 13,310 signals in all three ON looks, turboSETI 442. BLISS run as its own pipeline (its GPU hits and its event stage) never reached that false-candidate rate at any cut tested. BLISS's own hits put through this search's ON/OFF test did reach it, and found 4,048.

On a second session, fixed in advance (HIP17147, 2016), this search again found far more than turboSETI (4,337 against 417), but both pipelines **failed** the false-candidate limit set beforehand: 39.1 and 1.27 a block against 0.1, almost all of this search's in known satellite bands. That failure is part of the result.

On a third session, chosen by a rule from the archive's metadata and fixed in advance (NGC6946, a galaxy, at C band in 2019), every criterion fixed beforehand was met at its point estimate: this search found 2,404 of 7,040 signals against turboSETI's 247, and neither raised a false candidate. That band was quiet (3 of 64 blocks crowded), and 64 blocks leave the rates' intervals reaching 0.12, above the limit.

These are three sessions from one telescope. They do not show better sensitivity in general.

Four real cadences of other targets, chosen by a rule fixed beforehand and searched whole, gave 0.029 to 0.31 false candidates a block at S/N 30 (0.125 over all four, bunched in 18 of 1,267 blocks), and the checks left none unexplained. Signals planted in the same cadences show the checks keeping most real-looking signals in complete blocks, but not in blocks at the hit limit, which lie mostly in listed interference bands.

The question

Searches for narrowband technosignatures look for a tone drifting steadily in frequency that appears when the telescope points at the target (ON) and disappears when it points away (OFF). Two things decide how good a search is: how many real signals it finds, and how many interference signals it wrongly keeps. A pipeline can always find more by keeping more, so the fair comparison is how many planted signals each finds at the same rate of false candidates.

Method

- **Data.** Breakthrough Listen's high-resolution spectra (about 2.8 Hz channels) of a whole ON-OFF cadence: three ON looks at the target, each followed by an OFF look elsewhere. In 2017 and 2016, 121 blocks of frequency of each session, 61 crowded with interference and 60 clean, chosen at random within each class; in 2019 all of the session's 64. The false-candidate rates weight each class by its share of the session's blocks; the recovery counts are of the blocks measured, and are also given weighted to the session below.
- **Planted signals.** 110 a block, in the ON looks only, across 11 strengths (S_in 10 to 400, the signal-to-noise a search would give collecting all its power in ideal noise) and five drift bins up to 4 Hz/s. Each look's signal is scaled to that look's own measured noise, so a signal has the same S_in in every ON look: this measures recovery at a controlled signal-to-noise, not of a transmitter of fixed flux whose signal-to-noise would vary from look to look.
- **The planting, checked** (separately, both pipelines, recovery at SNR 10 and 30): a tone added to synthetic voltages before they are squared, as a receiver sees it, against power added to the spectra (at most 2 of 880 signals apart); finer sampling of the drift within a spectrum (the same recoveries); signals crossing the DC bin planted before its blanking, as the sky's would be, against after (at most 2 of 96; only 8 planted signals in all three sessions cross a DC bin); and one flux in all three ON looks against one S_in, on real blocks (at most 2 of 880). The differences are a few signals at most and do not change one pipeline against the other. Flux in Jy Hz on the results pages is nominal, from an assumed SEFD of 10 Jy measured at L and S band only.
- **Found** means detected in all three ON looks within ±6 Hz and ±0.05 Hz/s of the planted signal (the tolerances of Margot et al. 2023).
- **False candidates** are counted by running each pipeline with the wrong looks called ON: the 19 other ways of choosing three of the six looks. Anything a pipeline keeps under a wrong choice is interference or noise. Candidates sharing a hit count once. This is a wrong-cadence control rate, not a calibrated false-alarm probability: the telescope's pointings were not assigned at random, the looks share their interference, and a wrong choice's looks are put in ON, OFF order, not in time order. How well it predicts false candidates in real cadences has not been measured.
- **Matched comparison.** Each pipeline's cut (its signal-to-noise threshold) is set where it gives at most 0.1 false candidates a block, and recovery is compared there.
- **Pipelines.** turboSETI 2.3.2; BLISS at commit 2b98afe with its GPU backend and its own event stage; this search, which sums along every drift up to ±4 Hz/s with the smear removed, keeps hits from S/N 5 as shadow hits, and calls a candidate only when every ON look shows the signal and no OFF look does. Its candidate needs a hit at the cut in one ON look and the signal measured along its track in the others; turboSETI's needs a hit in each. So these are whole pipelines compared, detector and cadence test together; BLISS's hits through this search's test separate the two in part.
- **The second and third sessions** used criteria frozen before their data was read, with a fingerprint of the code checked at each stage; the third was chosen by a rule fixed beforehand, from the archive's metadata alone.

Results: the 2017 session (KIC8462852, AGBT17A_999_68)

pipeline	cut	false candidates a block (95% interval)	signals found of 13,310
this search	30	0.068 (0.023 to 0.13)	4,520

pipeline	cut	false candidates a block (95% interval)	signals found of 13,310
turboSETI	30	0.049 (0 to 0.13)	442
BLISS's hits, this search's ON/OFF test	30	0.052 (0.0065 to 0.12)	4,048
BLISS, its own pipeline	none reaches 0.1	lowest 4.54 (at 100)	—

At its tested cut, each matched pipeline's measured false rate is at or below 0.1 a block. These are point estimates, and they differ; each 95% interval reaches above 0.1, and the cuts were chosen on the same blocks they are judged on. The pipelines are matched on the measured rate against the target, not shown to lie below it. On the same 13,310 signals, this search and BLISS's hits through its test both found 3,829; this search alone 691, BLISS's hits alone 219. The difference, +0.035 of the signals (0.0261 to 0.0456, block bootstrap, 4000 draws, seed 2), holds for this session, these cuts and this signal grid.

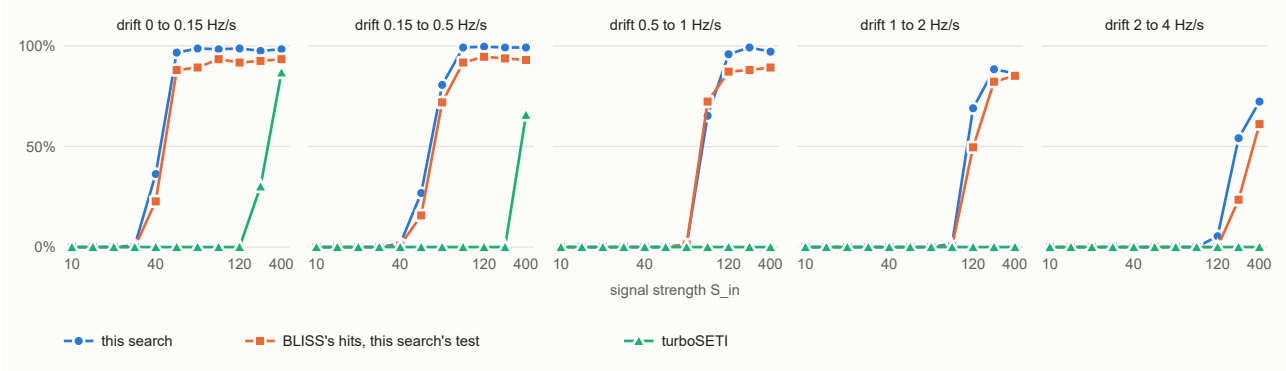


Figure 1. Share of the planted signals found in all three ON looks at each pipeline's matched cut, against signal strength (the eleven steps of the grid, evenly spaced), one panel per drift range (242 signals at each point). At each pipeline's cut its measured false rate is at or below 0.1 a block; the rates differ, and their 95% intervals reach above 0.1 (the table above).

Why BLISS's own pipeline has no matched cut. On the two blocks that carry most of its false candidates, one list of BLISS's hits at S/N 30 was given to each ON/OFF logic in the same run:

False candidates at S/N 30 (distinct, over the 19 wrong choices):

block	BLISS's event stage	this search's test	this search's test, with shadow hits	this search's test, hits grouped at S/N 5
35	1,610	85	85	1
49	0	1	1	1
79	0	0	0	0
129	2,627	205	130	0
131	2	0	0	0
139	10	17	17	9
166	0	0	0	0
185	0	0	0	0

block	BLISS's event stage	this search's test	this search's test, with shadow hits	this search's test, hits grouped at S/N 5
221	1	0	0	0
231	0	0	0	0

With the hits the same, this search's test removes most of them: BLISS's event stage is the larger cause. The rest come from BLISS grouping its hits at the higher cut; grouped at S/N 5 and cut afterwards, almost none are left. These are two blocks, run after the main result was seen.

Results: the 2016 session, fixed in advance (HIP17147, AGBT16A_999_192)

criterion, frozen before the data was read	result
this search's false rate	**fails**
turboSETI's false rate	**fails**
this search's recovery above turboSETI's	**holds**
the pass's false rate	**fails**
the pass's curved recovery above production's	**holds**
the pass loses none of production's recoveries	**holds**

The code changed after the freeze, once: independent.py, logged as a deviation at each of the 4 stages that ran on it (choose, measure, curved, verdict) and printed in the verdict. The reason logged: "choose read search_version from the blocks table, which holds none (it is the run's); the query now goes through store.done_blocks. Plumbing only: the block rule, seed 10 and the measuring code are unchanged." (data/step10/deviations.json).

At the frozen cut (30), this search raised 39.1 false candidates a block (13.6 to 69.2) and turboSETI 1.27 (0.268 to 2.8), against a limit of 0.1. Recovery held: 4,337 of 13,310 against 417; paired, +0.294 of the signals (0.284 to 0.304). A second pass for curved tracks, tested with it, found 2,769 of 7,260 curved signals against production's 389, losing none.

Looked at afterwards, not one of the criteria: nearly all of this search's false candidates lie in blocks overlapping known satellite bands (GNSS, Inmarsat). No band has been masked, and the test has not been changed, because that would be choosing the rule after seeing the result.

Results: the 2019 session, chosen and fixed in advance (NGC6946, AGBT19B_999_104)

The session was chosen by a rule fixed before any listing was read: Breakthrough Listen's target list (Isaacson et al. 2017), telescopes and bands in a fixed order, the first session with a whole ON-OFF cadence and a file layout on which planted signals are recovered as on Green Bank's (checked on synthetic noise). No Parkes session qualified; this is Green Bank at C band, one recording node's 64 blocks, all measured. (A first run of the rule, on a shorter target list, found no session; the list was then extended once, before anything else was read.)

criterion, frozen before the data was read	result
this search's false rate	**holds, its interval reaching above the limit**
turboSETI's false rate	**holds, its interval reaching above the limit**
this search's recovery above turboSETI's	**holds**
the pass's false rate	**holds, its interval reaching above the limit**
the pass's curved recovery above production's	**holds**
the pass loses none of production's recoveries	**holds**

The code changed after the freeze, once: tools/turbo_curves.py, logged as a deviation at each of the stages that ran on it and printed in the verdict. The reason logged: "tools/turbo_curves.py sets a 4 GB floor on blimpy's load limit, which follows free memory: with another program holding the RAM, blimpy kept files on disk and turboSETI refused them (data shape (1,)); the data read and every hit are unchanged" (data/step12/deviations.json).

At the frozen cut (30), both pipelines raised 0 false candidates a block (95% interval 0 to 0.115), against a limit of 0.1: a rate of zero, but 64 blocks cannot show it below the limit. This search found 2,404 of 7,040 signals in all three ON looks, turboSETI 247. The curved pass found 1,621 of 3,840 curved signals against production's 265, losing 0. The band was quiet: 3 of 64 blocks crowded, against 184 of 304 in 2016.

Recovery at every operating point

A single cut does not give one ratio between the pipelines. For each session, the same planted signals at the frozen cut (30 each), the published one (10 each) and each pipeline's matched cut (step 4's rule: the lowest tested cut from which the false-candidate rate's point estimate stays at most 0.1 at every higher cut). Supplementary and exploratory: the frozen verdicts above are unchanged.

session	operating point	cuts (this search, turboSETI)	found in the sample	false a block	paired difference in the sample	weighted to the session
KIC8462852	frozen (30 each)	30, 30	4,520 against 442 of 13,310	0.0682, 0.0487	+0.306 (0.301 to 0.311)	34.4% against 3.5%, +0.309 (0.305 to 0.314)
KIC8462852	published (10 each)	10, 10	8,706 against 1,709 of 13,310	3.3, 0.286	+0.526 (0.514 to 0.536)	66.5% against 13.0%, +0.535 (0.528 to 0.541)
KIC8462852	matched (step 4's rule)	30, 30	4,520 against 442 of 13,310	0.0682, 0.0487	+0.306 (0.301 to 0.311)	34.4% against 3.5%, +0.309 (0.305 to 0.314)
HIP17147	frozen (30 each)	30, 30	4,337 against 417 of 13,310	39.1, 1.27	+0.294 (0.284 to 0.304)	32.0% against 3.0%, +0.290 (0.278 to 0.301)

session	operating point	cuts (this search, turboSETI)	found in the sample	false a block	paired difference in the sample	weighted to the session
HIP17147	published (10 each)	10, 10	8,258 against 1,550 of 13,310	107, 1.95	+0.504 (0.485 to 0.521)	60.9% against 11.3%, +0.496 (0.474 to 0.515)
HIP17147	matched (step 4's rule)	—	no matched cut for this search			
NGC6946	frozen (30 each)	30, 30	2,404 against 247 of 7,040	0, 0	+0.306 (0.302 to 0.311)	34.2% against 3.5%, +0.306 (0.302 to 0.311)
NGC6946	published (10 each)	10, 10	4,702 against 952 of 7,040	0, 0	+0.533 (0.525 to 0.54)	66.8% against 13.5%, +0.533 (0.525 to 0.541)
NGC6946	matched (step 4's rule)	8, 5	5,049 against 1,714 of 7,040	0.0156, 0	+0.474 (0.466 to 0.481)	71.7% against 24.3%, +0.474 (0.466 to 0.482)

The sample's crowded blocks against the session's: KIC8462852 61 of 121 against 61 of 308; HIP17147 61 of 121 against 184 of 304; NGC6946 3 of 64 against 3 of 64. Weighting each class by its share of the session (bootstrap over blocks within each class) moves the difference by at most 0.009, so no conclusion here rests on how the blocks were sampled. The paired difference in the sample has a bootstrap over blocks; resampling instead whole windows of four consecutive blocks of the session (block numbers, gaps of unmeasured blocks kept) barely widens it, but that tests only correlation between neighbouring blocks, and windows of sixteen leave too few windows (20 or fewer) for an interval (all in the record).

At the lower cuts the ratio falls (about ten times as many found at 30 each, about three times at the matched cuts in 2019) while the difference in share grows. In 2016 no tested cut of this search reached the rate.

Limitations

- Three observing sessions, one telescope, one recording pipeline. Nothing here speaks for Parkes, MeerKAT or another receiver.
- Steady drifts within ±4 Hz/s, plus one experiment with curved tracks; no pulsed or broadband signals.
- The 2017 session's cuts were chosen on the same blocks they were judged on. The 2016 and 2019 sessions were fixed in advance: in 2016 both false rates failed, in a band crowded with satellite signals; in 2019 every criterion held, in a quiet band where the rate's interval still reaches above the limit.
- BLISS's hits through this search's test, and the same-hit control, are combinations of two pipelines, designed after the main result was seen.
- Of the planted signals BLISS finds at S/N 10 and this search raises no candidate for, 28 were traced stage by stage. 17 are found strongly in all three ON looks and then called interference

because one OFF look reaches S/N 6 to 8 along the track with no OFF transmission identified: low-level OFF evidence consistent with noise (0.66% of 10,000 random unplanted tracks reach 6; that is the test's rate on noise, not a loss rate for these signals, which were chosen because BLISS found them). 2 meet something real in an OFF look, 7 are missed by the detector (mostly in crowded blocks) and 2 are too faint. The threshold was set beforehand and is not moved for this.

- The 8 signals the detector gave no matching hit for were measured through its stages, the detector unchanged; the first failure seen in each: 6 the hit cap, 2 found, but measured outside the tolerances. The hit cap is the 2,000 hits kept a look: in crowded looks the signal, at S/N 12.3 to 40.6, lies below that many stronger hits, and a cap of 5,000 to 20,000 finds it. Those measured outside the tolerances drift fast (2.0 and 1.6 Hz/s); their measured drift was 0.05 Hz/s off, just outside the tolerance. Neither is changed here: the detector stays as step 10 ran it.
- The false-candidate control is four cadences: enough to show the rate varies between them, not to say how much it varies between sessions in general. Its interval over blocks describes those four observations only.
- The frozen recovery criterion compares two Wilson intervals, which treat the planted signals as independent; signals share their block and their session. The paired difference with a bootstrap over blocks (above) is the better measure, and it is supplementary. All the blocks of a session come from one observation, so none of these intervals speaks for other sessions.
- BLISS's hits through this search's test, supplementary in the third session, could not be run there: the step that hands BLISS each block expects a slice width these blocks do not have, and the frozen rule records that rather than adapting it.

A false-candidate control on four real cadences

The false-candidate counts above judge each block with the 19 wrong choices of three ON looks, which stand in for cadences with no signal. To see what real cadences give, four cadences of other targets, never read by this project, were chosen by a rule fixed before any of their data was read (the first four Green Bank sessions, in the order of Breakthrough Listen's target list, in the layout the search was validated on). Each was searched whole by the search as frozen, in its own time order. No candidate in them is taken to be a signal, so what they give is a false-alarm count on real cadences.

cadence	band	blocks	candidates at S/N 30	a block	at S/N 10	left unexplained by the checks
HIP2	L	308	30	0.0974	59	0
HIP169	L	308	11	0.0357	137	0
HIP171	L	308	9	0.0292	466	0
HIP194	S	343	108	0.315	525	0
all four		1,267	158	0.125	1,187	0

The rate depends on the cadence: at S/N 30, 0.029 to 0.31 a block. Over all four it is 0.125, and drawing runs of 16 consecutive blocks again within each cadence gives 0.022 to 0.31; that describes these four observations only, and four are too few to say how much the rate varies between sessions. The candidates are bunched: all 158 at S/N 30 lie in 18 of the 1,267 blocks, 86 of them in one. 157 are in the 174 blocks where a look reached the search's limit on hits, which are the crowded ones; that the limit itself makes them is not shown.

On the three L-band cadences the rate at S/N 30 is 0.054 a block; one wrong choice of ON looks, weighted as the published rates are, gives 2.1 on the 2016 session and 0.0036 on the 2017 one. So 2016's level of false candidates is not what these L-band cadences give in their real order, but the stand-in is not a steady guide either; different sessions are compared, so this cannot separate the method from the sessions' own interference. The S-band cadence gives 0.31.

The checks called every one of the 1,187 candidates at S/N 10 and over interference (119 shown, 1,068 likely), and left none unexplained. Most of those verdicts rest on signs short of proof, chiefly zero or very low drift, look-alikes in the cadence and crowding. The tables and verdict counts were made again from the packaged records alone (identical), and the checks run again from nothing changed 0 of 1,187 verdicts; both on the same computer, so not an independent verification.

Would the checks keep a real signal? A zero means little if they would also explain away a signal. Following a protocol fixed beforehand, 3,520 signals were planted in the ON looks of 32 blocks of these cadences, half the blocks at the hit limit and half complete, drawn by their completeness alone, and put through the search, the cadence test and the checks as they are:

blocks	signals counted	a hit in an ON look at S/N 30	candidates	left unexplained	unexplained among candidates (95% over blocks)
complete	1,760	646	605	447	0.74 (0.56 to 0.90)
at the hit limit	1,760	489	421	86	0.20 (0.05 to 0.38)
all	3,520	1,135	1,026	533	0.52 (0.36 to 0.67)

Against the planted candidates they did not keep, the check cited most was known interference bands (401 times), then crowding (80), zero or very low drift (56), recurrence at other targets and times (56). A signal in a listed interference band is marked likely interference whatever it is, and the crowded blocks lie mostly in those bands; so there the checks cannot keep a real signal, by design.

By the reading fixed in the protocol (informative where the lower end of the interval for unexplained among candidates is at least 0.5), the checks kept planted signals in complete blocks, and were not shown to keep them in blocks at the hit limit. So control B's zero is informative in complete blocks, and says little in blocks at the hit limit. Planted tones are clean and do not behave like every real transmitter.

What the code audits found that bears on these results

A line-by-line audit of the code (docs/audits.md in the repository) found the following. None changes a frozen verdict; each is stated here so the numbers above are read as what they are.

- **The 2016 reads' DC bins.** The 2016 session's coarse channels are 1,033,216 channels wide and its files' storage chunks 2^20; the reads blanked the receiver's DC spike at each chunk's middle, so its real DC spikes (tens of thousands of times the noise) stayed in and one ordinary channel a chunk was smoothed. None of the 4,016 false events at S/N 30 over the 19 wrong choices lies within 50 Hz of a DC channel (a spike in every look is vetoed by its OFF looks); step 10's records and verdict stand as made, and the same blocks read with the corrected rule are being compared. 2017 and 2019 are not affected.
- **False candidates a block** count the distinct candidates over all 19 wrong choices of ON looks together, not what one cadence would give; the union can be up to about 19 times one choice's

rate. It is a control for comparing pipelines on the same blocks, not a calibrated false-alarm rate.
- **The false-candidate intervals** are a bootstrap over blocks within each class (a Poisson upper limit where a class has none). Where interference correlates between blocks they hold the true rate far less often than 95%: about 41% in a simulation like 2016, and no method tried fixes that. The 2016 rate's interval is therefore not a 95% interval.
- **The recovery intervals** with whole windows of neighbouring blocks resampled are a supplementary check; they too fall short of 95% when blocks correlate strongly (about 46 to 60% at a correlation of 0.95 in simulation, 80 to 90% at 0.5 to 0.8).
- **turboSETI** as run (FindDoppler's defaults) never searches zero drift, and at fast drifts its reported frequency and drift can fall outside the ±6 Hz, ±0.05 Hz/s tolerances: of 60 strong signals at 2 to 4 Hz/s it detected all, and the tolerances credit 5. The comparison is of whole pipelines, so detection, event association, the reported frequency and drift, and the matching rule all enter the difference.
- **BLISS** was run from its source at commit 2b98afe on its GPU, with its event stage ported line for line (and checked against the original on the hits it was given, not under the whole production chain). Its event stage keeps events with hits in two of the three ON looks; counting only three-ON events, its rate is still far above 0.1 a block at every cut (1.6 at S/N 100).

Data and code availability

The observations are from the Breakthrough Listen Open Data Archive (Lebofsky et al. 2019, PASP 131, 124505). Breakthrough Listen is managed by the Breakthrough Initiatives. The code is at https://github.com/Ryan-Clinton/SignalSearchYoke (this version was built from the commit named on its page). The code and the result records behind every number here are to be archived on Zenodo; the DOI will be added here when they are.

References

- Lebofsky, M. et al. 2019, PASP 131, 124505: The Breakthrough Listen search for intelligent life: public data, formats, reduction and archiving.
- Margot, J.-L. et al. 2023, AJ 166, 206: A search for technosignatures around 11,680 stars with the Green Bank Telescope at 1.15-1.73 GHz. doi:10.3847/1538-3881/acfda4 (the ±6 Hz, ±0.05 Hz/s tolerances).
- Enriquez, J. E. et al. 2017, ApJ 849, 104: The Breakthrough Listen search for intelligent life: 1.1-1.9 GHz observations of 692 nearby stars (turboSETI's method).
- Isaacson, H. et al. 2017, PASP 129, 054501: The Breakthrough Listen search for intelligent life: target selection of nearby stars and galaxies (the list step 12 chose from).
- turboSETI, version 2.3.2: https://github.com/UCBerkeleySETI/turbo_seti
- BLISS, commit 2b98afe: https://github.com/UCBerkeleySETI/bliss

Revisions

- 0.1 (2026-11-18): first draft.
- 1.0 (2026-11-19): first release: the signals BLISS finds and this search misses traced stage by stage, and the detector's misses measured.
- 1.1 (2026-11-19): the code's repository, now public, linked.

- 1.2 (2026-11-19): the third session, chosen and fixed in advance (NGC 6946 at C band): every criterion held.
- 1.3 (2026-11-19): recovery at every tested cut for all three sessions, with the paired difference and blocks resampled in runs; step 12 described as meeting its criteria at the point estimate; what the injections, the false-candidate control and the comparison of whole pipelines do and do not measure.
- 1.4 (2026-11-19): recovery weighted to each session's share of crowded and clean blocks beside the sample's; the neighbouring-block check made over consecutive blocks of the session; the matched cut by one rule throughout; the blocks of each session stated.
- 1.5 (2026-11-19): checks of the planted signals: a tone in the voltages, finer sampling, the DC bin and a fixed flux, none of which moves recovery.
- 1.6 (2026-11-20): what the code audits found, stated where it bears on the results: the 2016 reads' DC bins, the false-candidate count and its intervals, turboSETI's zero drift and fast-drift reporting, how BLISS was run; the planted signals counted per session; the injection checks' small differences stated as such; references completed.
- 1.7 (2026-11-24): a false-candidate control on four real cadences of other targets, chosen by a rule fixed beforehand and searched whole, with the checks' verdicts on its candidates; and signals planted in those cadences, to show how many real-looking signals the checks keep.