Mitigating Mining-Induced Hazards

A Continuous Surface Monitoring System using Kriging

Trigteq DSS
Unubold Erdenebeleg, Azjargal Ochirbat

ANSO-MAS 2026 · Ulaanbaatar · 29 July 2026

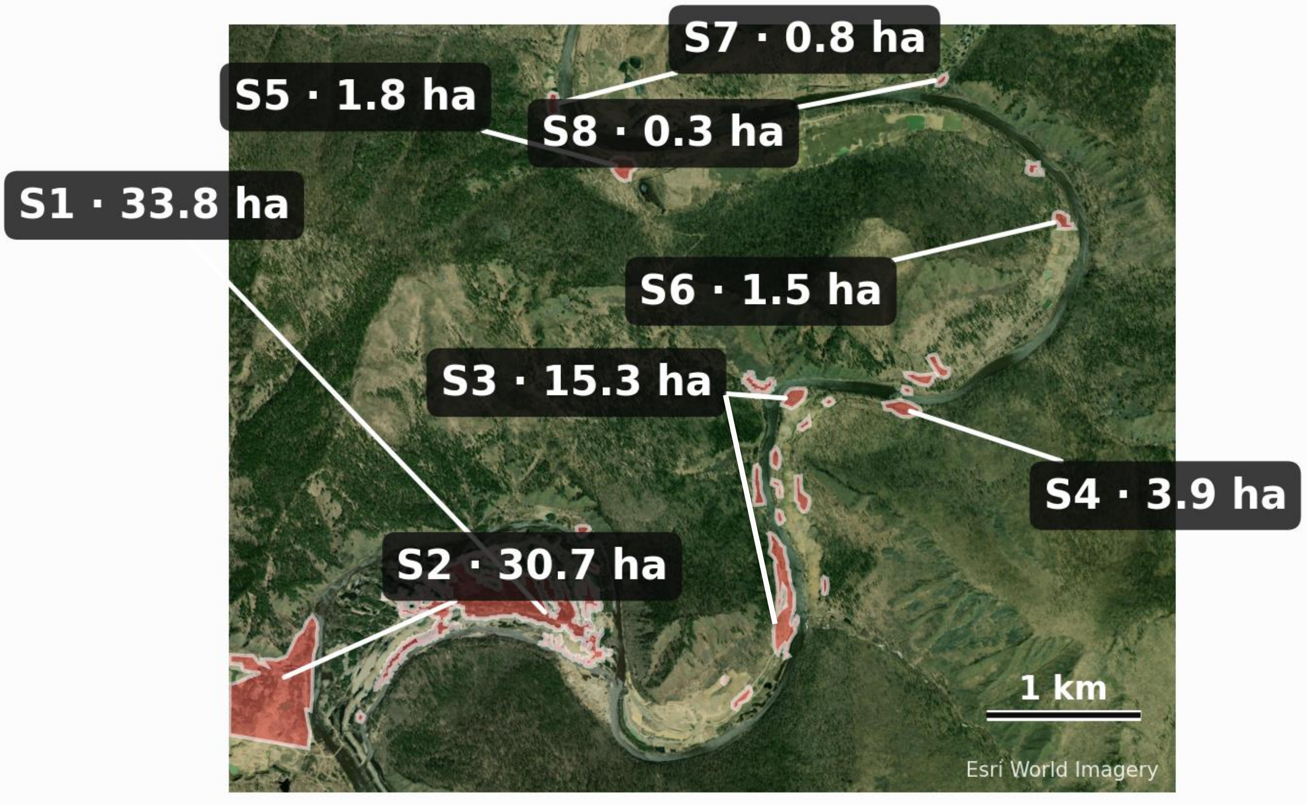
Extraction outpaces oversight

- Exploration and extraction routinely proceed without the required inspections or ecological assessment.
- Not only unauthorized operations, **licensed activity outruns its checks too**.
- What is left behind **migrates into rivers and pastures long after work stops**. [1][2]
- Remediation costs exceed prevention **by orders of magnitude**.

What a continuous monitor has to do

- Run on **free, routinely updated imagery**, or it cannot be continuous.
- Catch **new, small** disturbance **early**, and separate it from **natural seasonal change**.
- Build a **longitudinal record**: early warning, dated evidence, and later verification of rehabilitation.

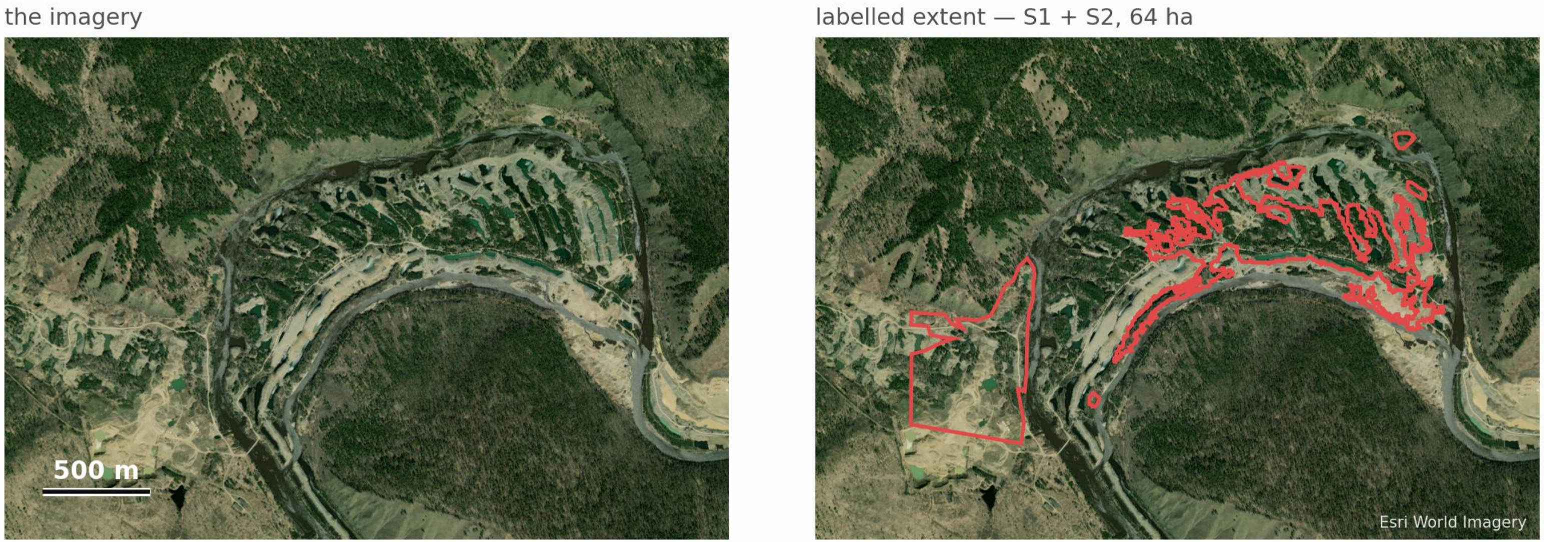
The demonstration area: 30.6 km² of the Eroo, eight sites



What the disturbance looks like

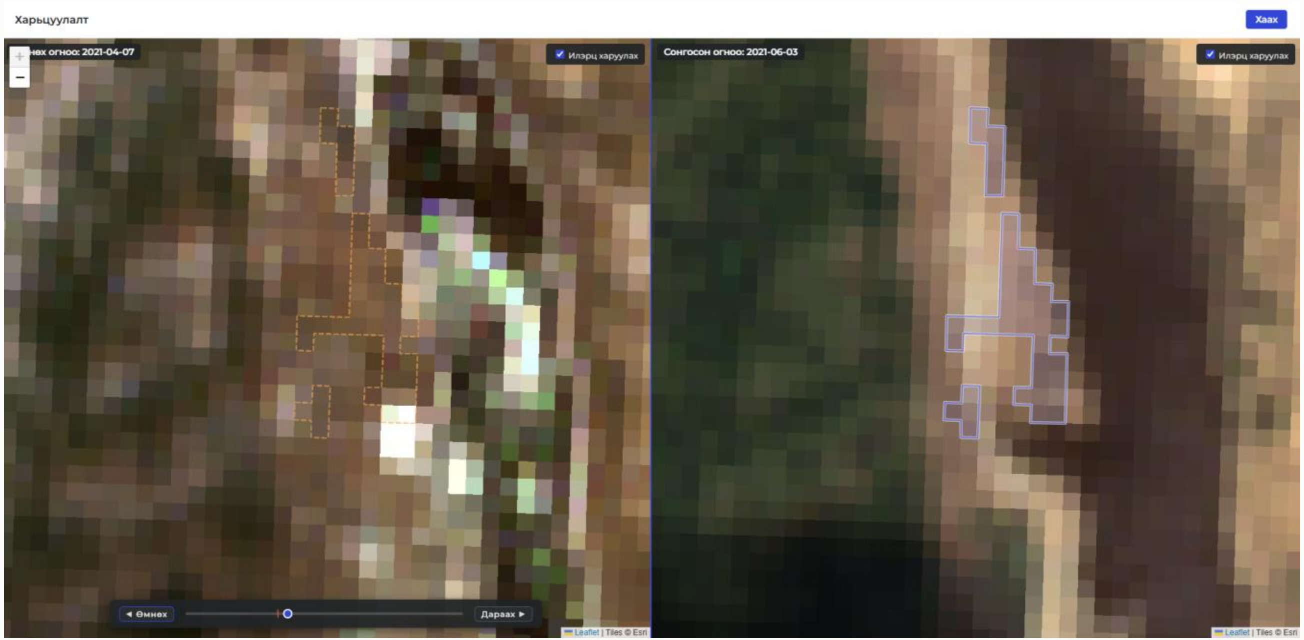
Braided pits and spoil along the channel, unmistakable once large.

The southwest complex — braided pits and spoil along the channel, the shape the monitor has to find while it is still small



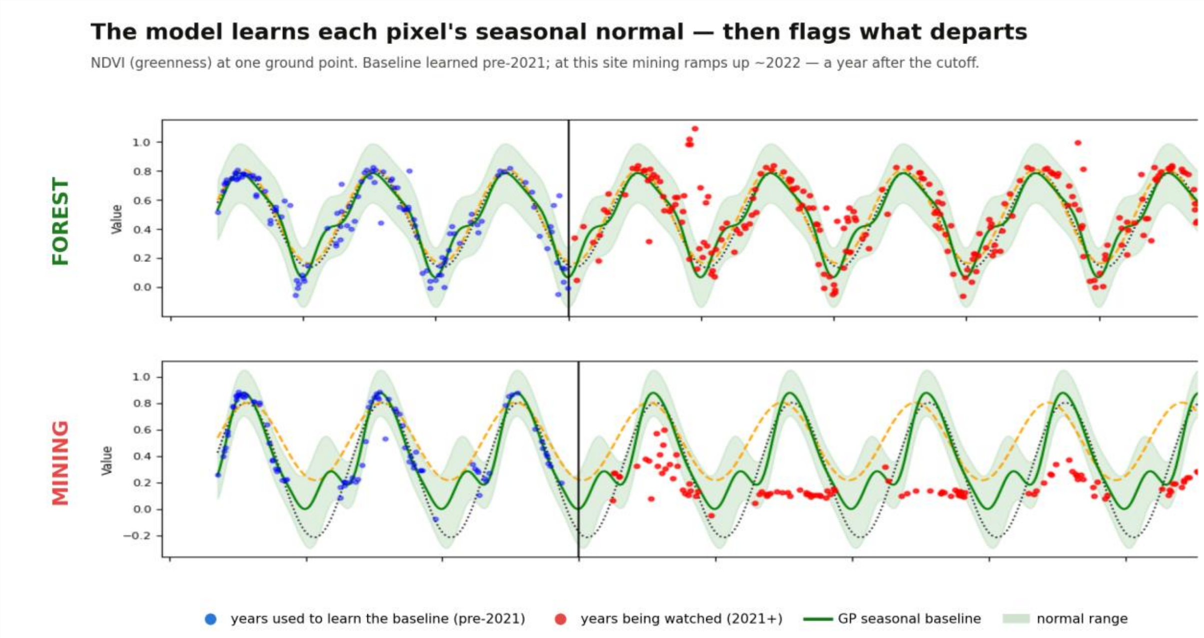
Why this is hard to predict

At 10 m, a fresh cut is **a handful of pixels**.
In any single image it looks like dry grass or river gravel.
The valley greens and browns each year, **“change” alone is not disturbance**.



A self-learned baseline per pixel

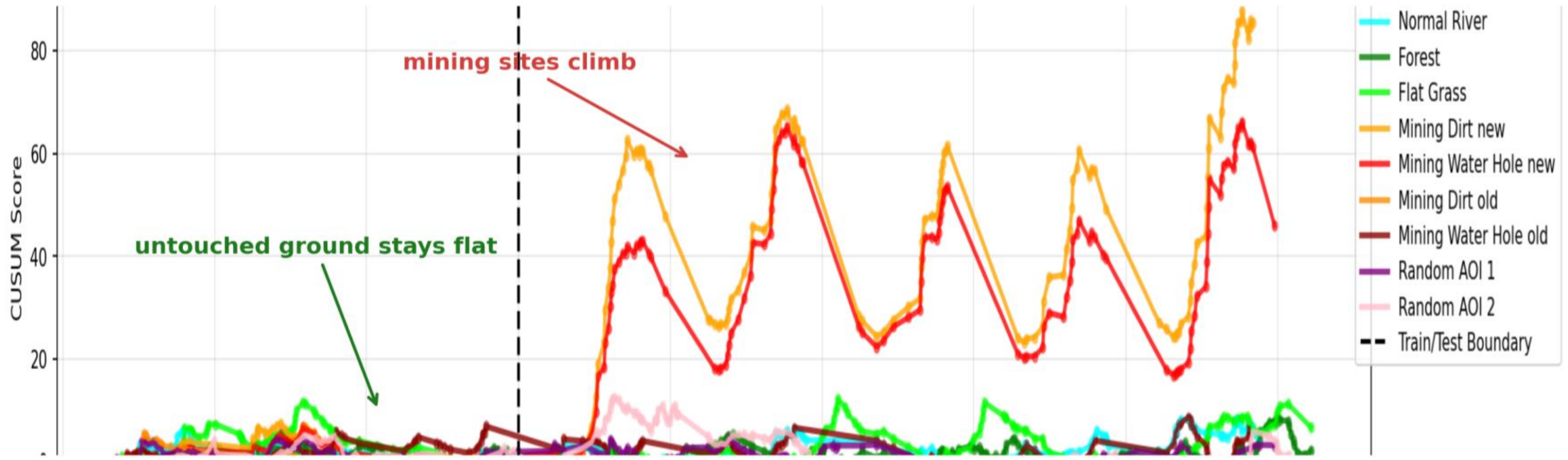
Every pixel learns **its own seasonal rhythm [4][5]**, a kriging baseline with a normal range.
The kernel is locked: **one-year** periodic, **five-year** trend. A step change cannot be absorbed.
Unsupervised, nothing to train, nothing to port to a new region. [3]



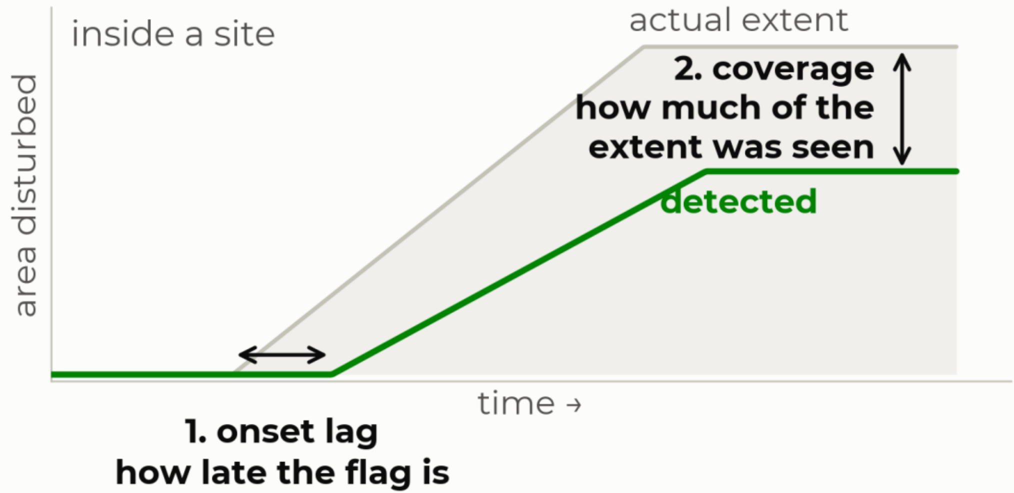
Telling disturbance apart from nature

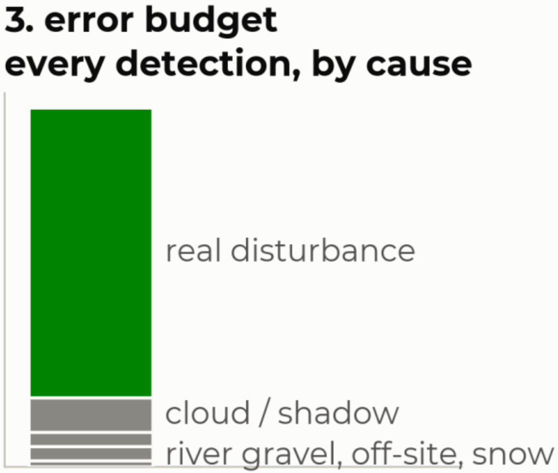
One off-band observation is not a strong evidence evidence, cloud and haze do that too.
Cumulative disturbance score climbs away; forest, grass and river stay flat, **with no per-site tuning**.[6]

Accumulated anomaly score (NDVI). Each line is one landmark; higher = more sustained departure from its baseline.



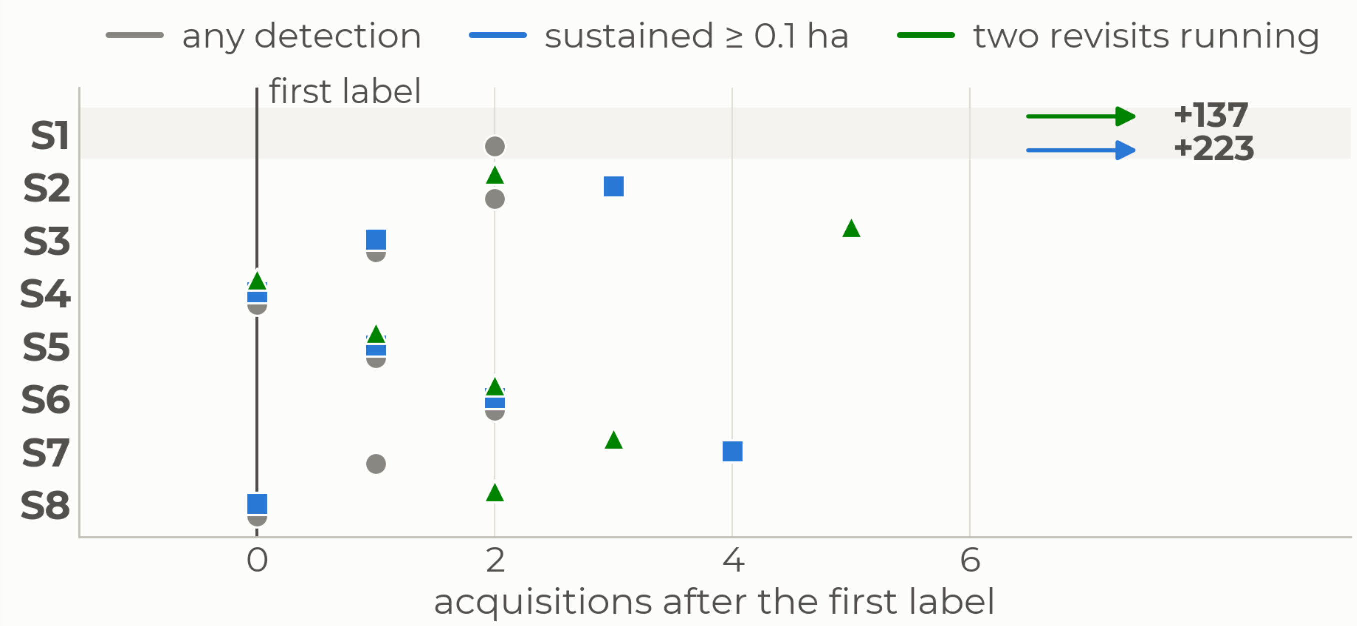
How do you grade a monitor?



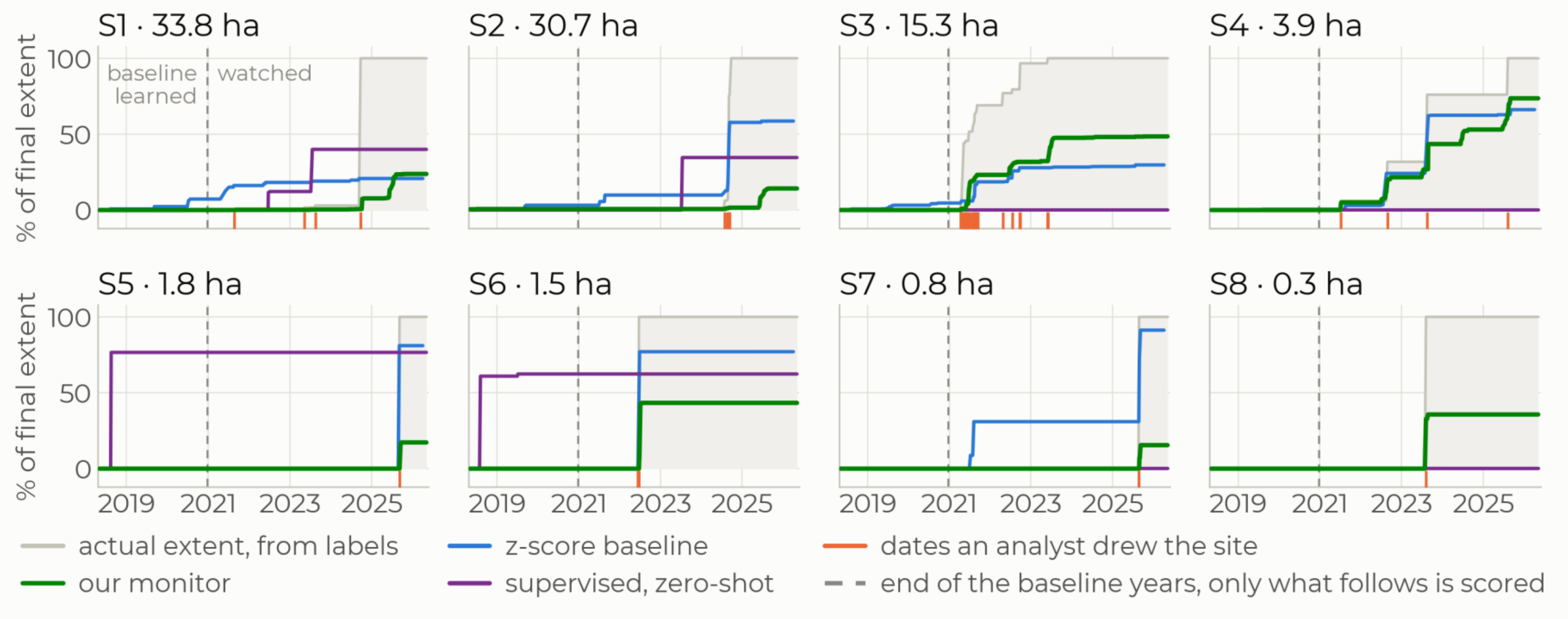


Is it on time?

Strictly **out of sample**: baseline learned pre-2021, everything scored 2021–2026.
Every site flagged within ~two timesteps of data, median one.

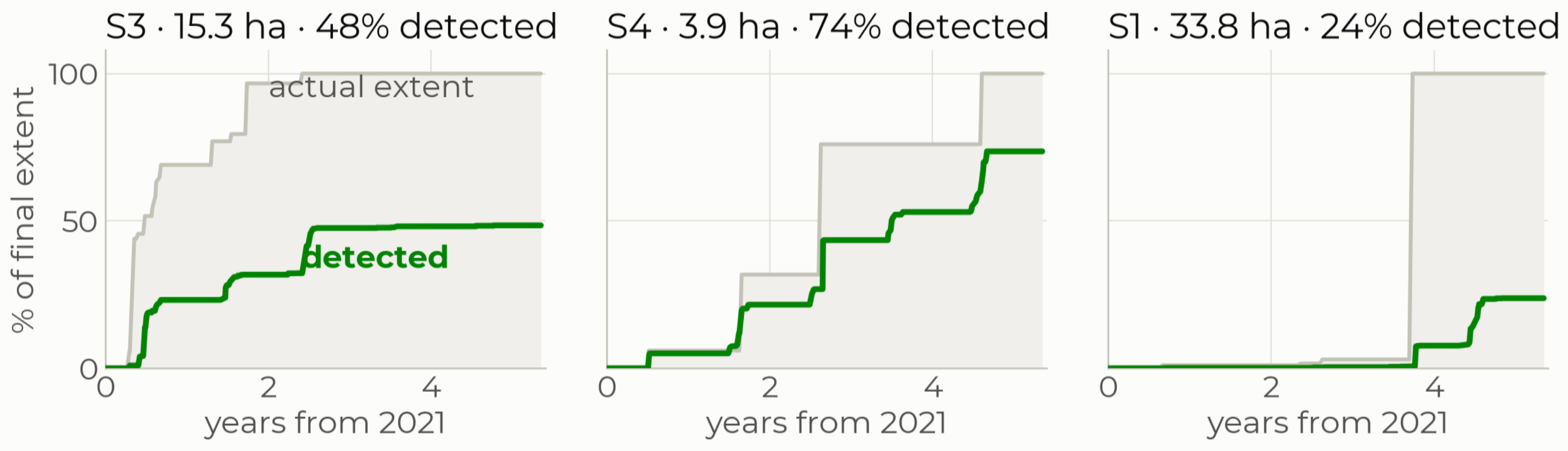


Every site, every detector



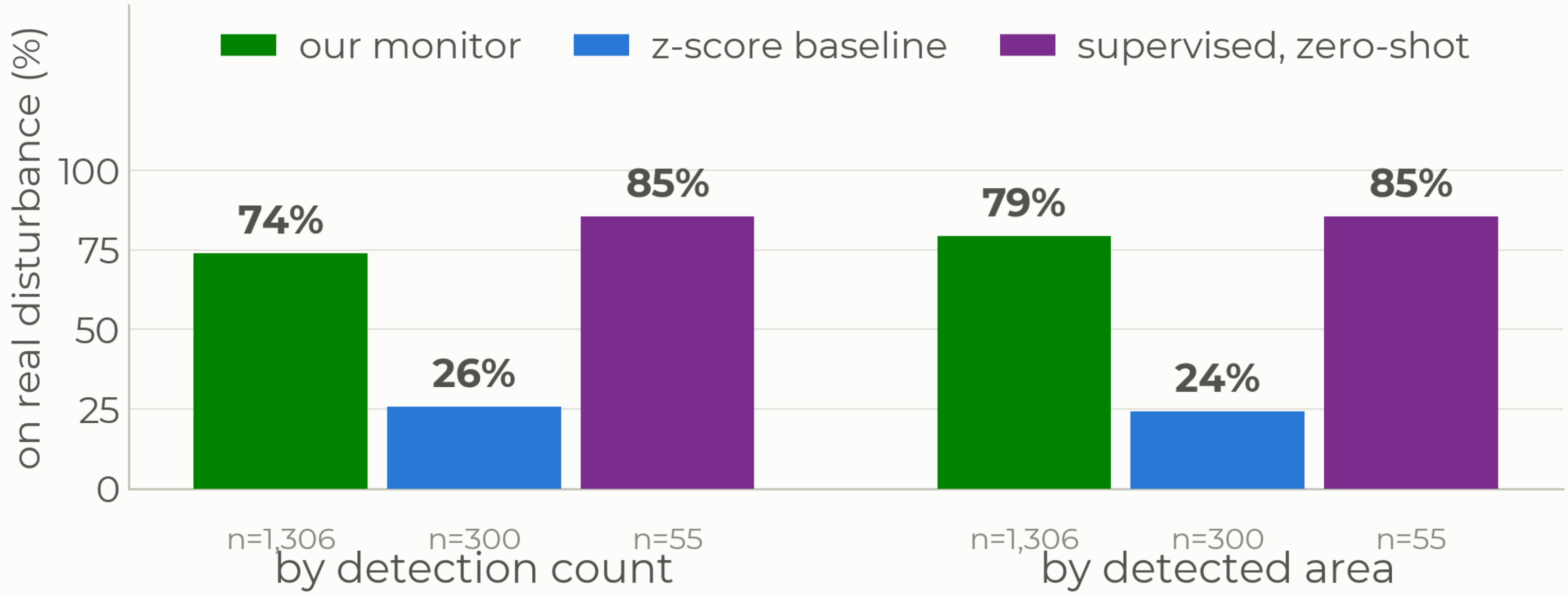
Does it trace the extent?

Partial by design: the monitor pings the active face, it does not trace the whole pit.
Acceptable for an alert, but only if we **say so** rather than quietly scoring ourselves on presence.

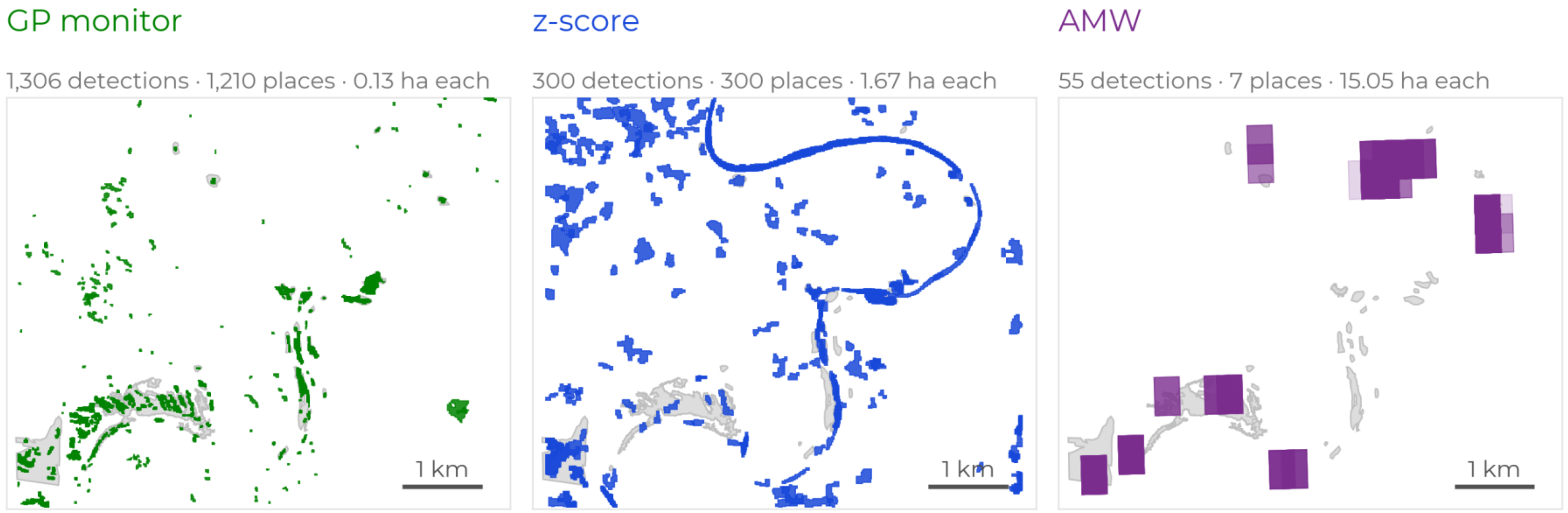


Three detectors, one basis

3× more of ours lands on real disturbance than the z-score's.
The supervised model AMW [7] leads on precision. Coverage is the next slide.



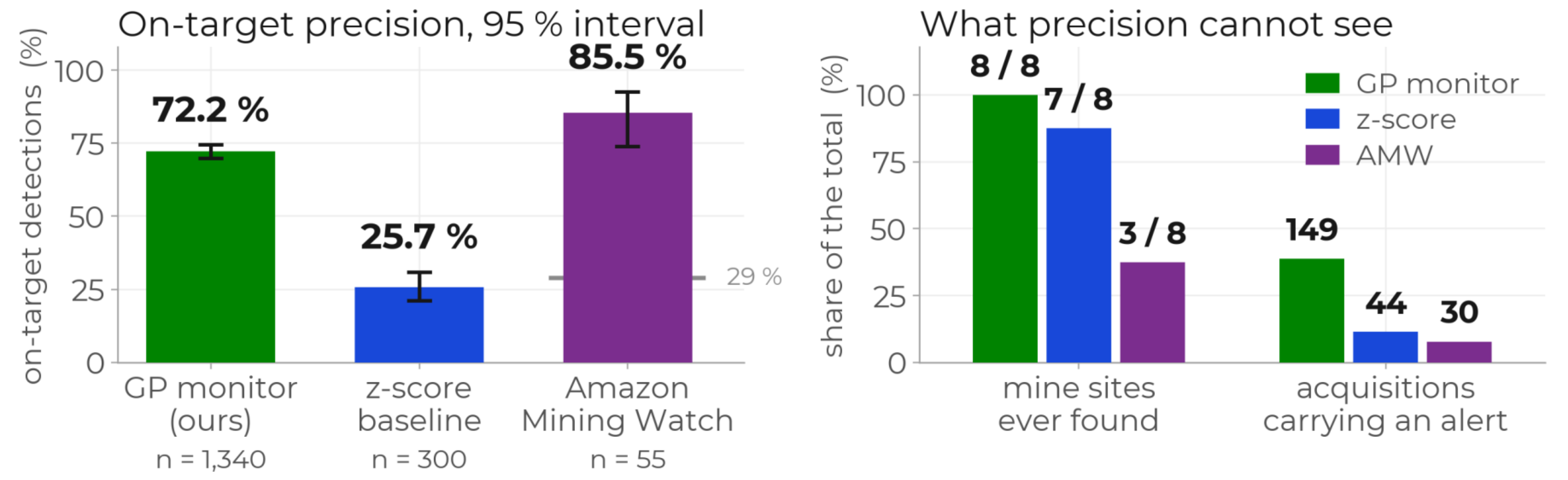
Map comparison view.



Grey = analyst-labelled mining; pale purple = AMW inside the training window, not scored. Over the whole record AMW returned 269 patches on only 22 distinct 480 m cells: repeat firings on the same ground, not new places.

Against a supervised model

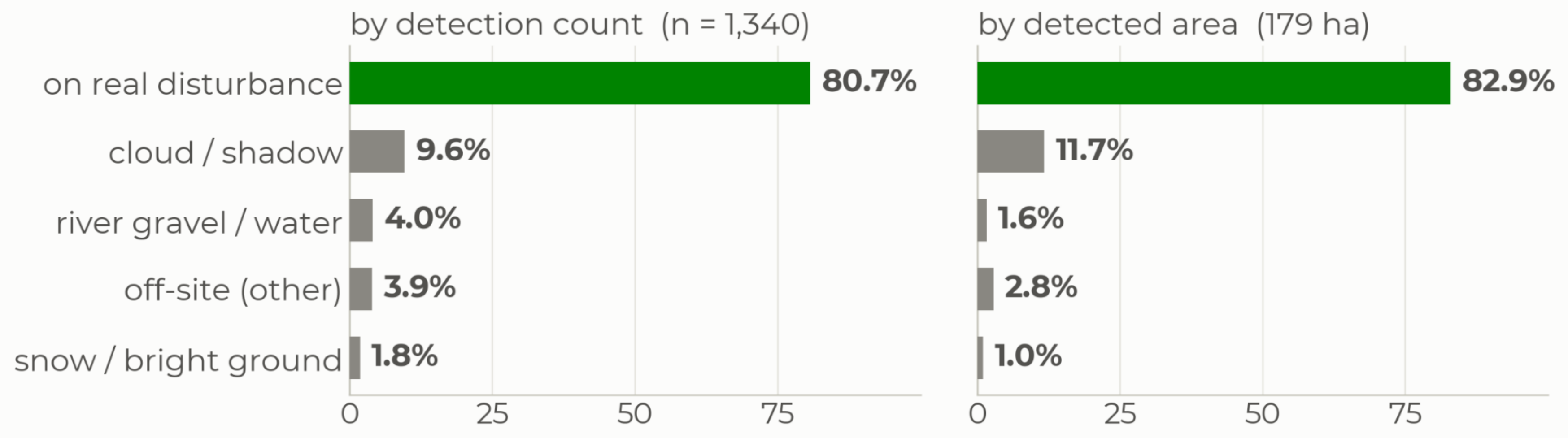
Amazon Mining Watch [7], a supervised CNN, re-run on **our grid**.
The most precise detector here: 85.5 % against our 57.4 %.
It reaches **three of eight sites** and paints **828 ha** doing it. We flag 176.



Reporting precision without coverage inverts the ranking.
On-target = touching labelled mining or ground a reviewer confirmed as mining.
n includes the 34 detections a late audit reclassified as errors.

What it gets wrong

Every off-label detection was clustered and **reviewed in context**, then given a cause.
Cloud and snow are 11 % with a known fix; river gravel adds 4 %. Unexplained residual **≈ 4 %**.
Honest caveat: only **53 % of detected *area*** sits strictly inside a mapped site.



Where this goes next

- **Filter the known noise.** Cloud, snow and river gravel are named, recurring and maskable, 15 % of detections with a known remedy.
- **Make it cheap to scale.** Cost per km^2 is a research constraint, not an engineering afterthought.
- **More observations along the time axis.** The binding limit: the alert needs ~2 *usable* timesteps of data acquisition.

What we cannot yet claim

- **One basin, eight sites.** The design argues the method transfers; not yet *demonstrated* on a second basin.
- **Ground disturbed before the baseline window is invisible by construction**, new ruptures, not history. [8]
- **The metric inherits its labels' dating.**
- Ground truth was cross-checked by **four labellers** only.

In short

- **Challenge**, small, new disturbance, caught early, told apart from seasonal change, on imagery that stays free.
- **Technique**, an unsupervised per-pixel kriging baseline plus accumulated, multi-index evidence.
- **Metric**, onset, coverage, and an error budget with named causes.

Early warning before the damage is obvious.

Trigteq DSS · hello@trigteq.com

References

1. Stubblefield, A., Chandra, S., Eagan, S., et al. (2005). Impacts of gold mining and land use alterations on the water quality of central Mongolian rivers. Integrated Environmental Assessment and Management, 1(4), 365-373. https://doi.org/10.1002/ieam.5630010406
2. Pecina, V., Juricka, D., Hedbavny, J., et al. (2023). The impacts of mining on soil pollution with metal(loid)s in resource-rich Mongolia. Scientific Reports, 13, 2763. https://doi.org/10.1038/s41598-023-29370-w
3. Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian processes for machine learning. MIT Press. https://gaussianprocess.org/gpml/chapters/RW.pdf
4. Zhu, Z., & Woodcock, C. E. (2014). Continuous change detection and classification of land cover using all available Landsat data. Remote Sensing of Environment, 144, 152-171. https://doi.org/10.1016/j.rse.2014.01.011
5. Verbesselt, J., Zeileis, A., & Herold, M. (2012). Near real-time disturbance detection using satellite image time series. Remote Sensing of Environment, 123, 98-108. https://doi.org/10.1016/j.rse.2012.02.022
6. Page, E. S. (1954). Continuous inspection schemes. Biometrika, 41(1-2), 100-115. https://doi.org/10.1093/biomet/41.1-2.100
7. Earth Genome. (2026). Mining detector [Computer software]. GitHub. https://github.com/earthrise-media/mining-detector
8. VanderPlas, J. T. (2018). Understanding the Lomb-Scargle periodogram. The Astrophysical Journal Supplement Series, 236(1), 16. https://doi.org/10.3847/1538-4365/aab766

Backup: the three detectors, every number

	GP monitor	z-score	AMW
detections in the scored window	1,306	300	55
acquisitions carrying an alert	**149 of 385**	44 of 202	30 of 386
mine sites ever found	**8 of 8**	7 of 8	3 of 8
distinct places flagged	**1,210**	300	7
on-target, + audited mining	74.0 %	25.7 %	**85.5 %**
on-target by area, labels only	61.0 %	21.9 %	**85.5 %**
mean detection footprint	0.13 ha	1.67 ha	15.05 ha
total area flagged	176 ha	501 ha	828 ha

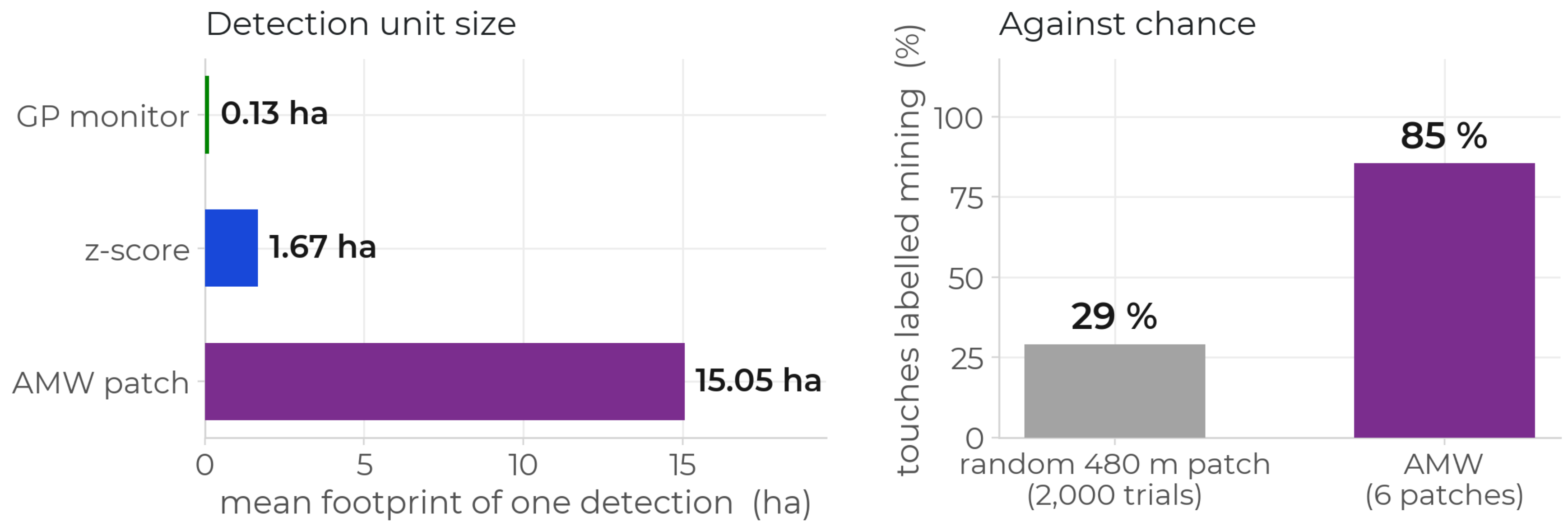
AMW leads every on-target row and finds 3 sites in eight, on 7 distinct places, while flagging 828 ha. That is the case for reporting coverage beside precision, not instead of it.

Backup: what was actually run

MODEL **48px_v3.7-ensemble** 9-member CNN, Earth Genome 2024-01-17	INPUT **48 × 48 px @ 10 m** 480 m patches, non-overlapping grid
BANDS **12-band S2 L1C TOA** B1–B9, B8A, B11, B12 (B10 dropped)	COMPOSITE **± 45 d median** S2_HARMONIZED, CLOUD_SCORE_PLUS ≥ 0.6
TIMESTEPS **597** our own acquisition grid, 2018–2026	THRESHOLD **0.30** shipped default is 0.50
OUTPUT **269 patches** max confidence 0.670, whole record	SCORED **55 detections** post-training, 2020-12-30 → 2026-05-06

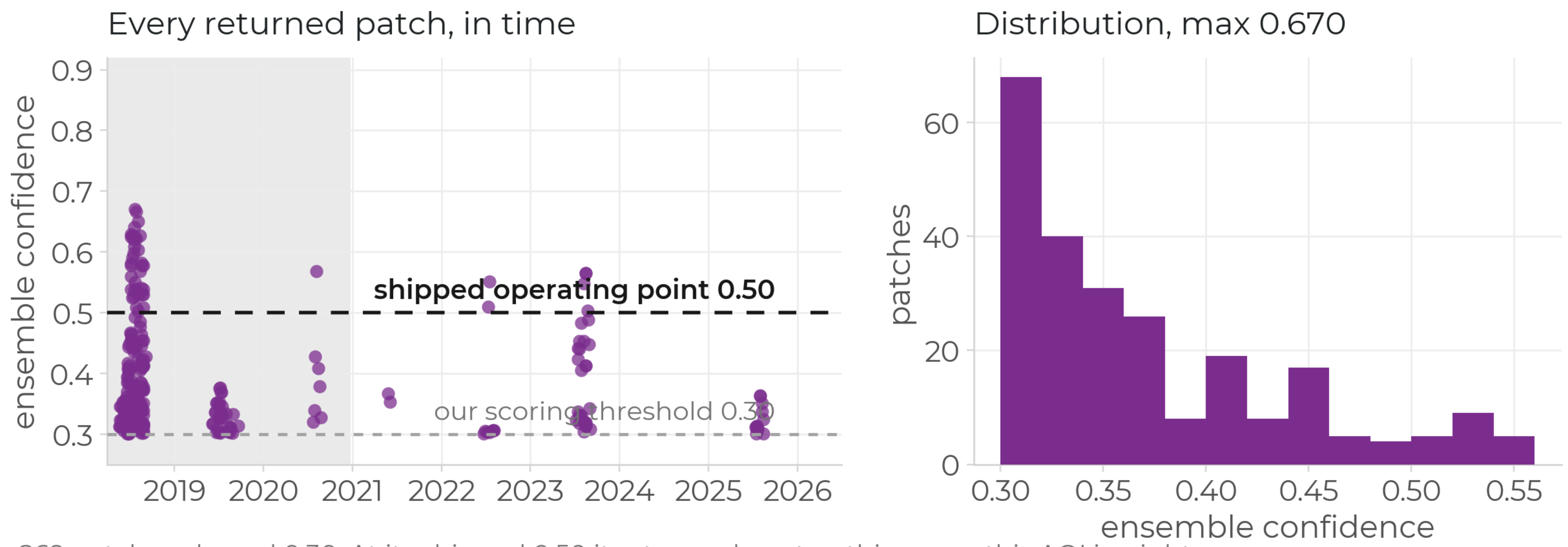
Zero-shot by design: does an Amazon-trained detector transfer to Mongolian steppe placer workings?

Backup: is 85 % good?



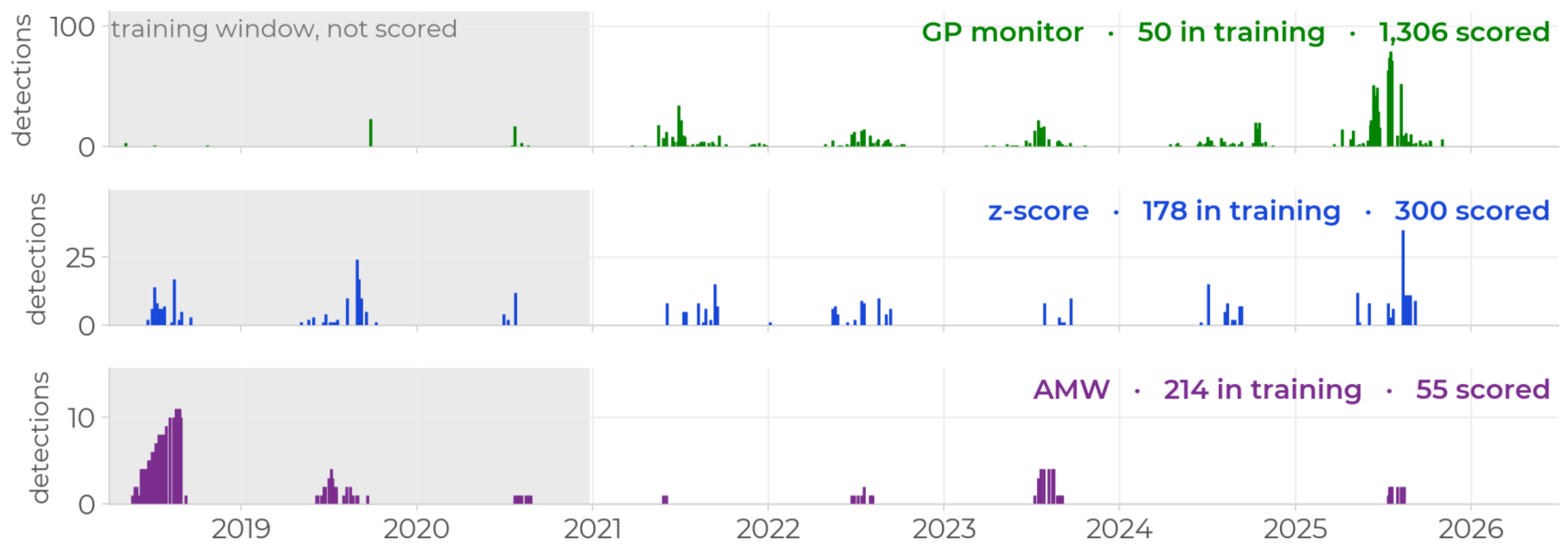
AMW beats chance, with a detection unit 115x larger than ours and 4.7x more ground flagged in total.

Backup: it never reaches its own operating point



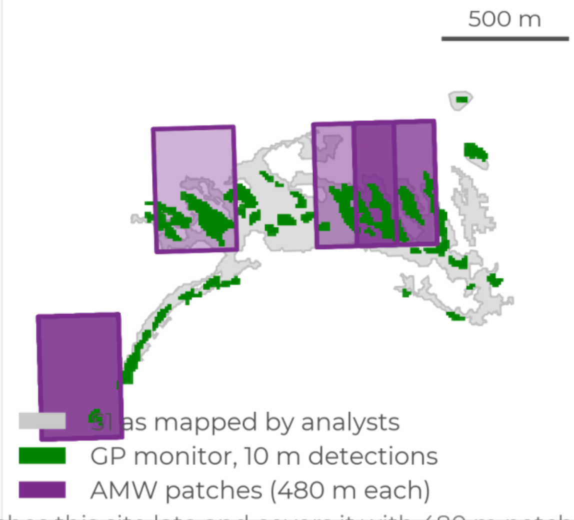
269 patches cleared 0.30. At its shipped 0.50 it returns almost nothing over this AOI in eight years.

Backup: when each detector fired



AMW: 70 patches total, 49 in 2018, 15 in 2019, 6 in 2023, none in 2020–22 or 2024–26. The mines grew throughout.

Backup: the largest site, and when each detector reached it



ONSET LAG AT S1, BY TRIGGER RULE

	GP	AMW
any detection	+17 d	+304 d
≥ 0.1 ha on a date	+1,144 d	+304 d

Under the weak rule we are 48 days early; under the 0.1 ha rule AMW is 452 days earlier. Neither is a fair contest, a 23 ha patch clears a 0.1 ha bar by construction, and our weak onset rests on one small pixel.

S1 EXTENT

33.8 ha

largest of the eight

AMW AT S2–S8

never

no patch intersects them

AMW reaches this site late and covers it with 480 m patches; the comparison of trigger rules above is why we do not claim a timing win here.

Backup: why not IoU?

- **Polygon or spatiotemporal IoU punishes under-coverage**, and under-coverage is the behaviour we deliberately accept.
- An alert needs the **right place, early**. A pretty outline is a different product.
- IoU also **hides the error budget**: a detector can improve its score by growing footprints, which is the opposite of what we want.
- So we score onset, coverage and errors separately.

Backup: how the error budget was audited

- Off-label detections were **clustered**, and a reviewer gave each cluster a cause on the imagery of the date it fired.
- The audit covers **176 clusters**, not 1,391 individual detections. Off-label detections **inherit the nearest cluster's verdict**.
- That assignment is tight, median **19 m**, p90 **100 m**, max **346 m**, which is why the totals stand.
- Labelling and adjudication were cross-checked by **four labellers**.

Backup: reproducing the comparison

- Model: `48px_v3.7-ensemble_2024-01-17.h5`, published weights, unmodified.
- Imagery: **downloaded and cached locally**, 597 twelve-band composites, 2.2 GB, not re-queried per run.
- Predictions: one GeoJSON per acquisition, `metric/detections_amw/<date>.geojson`, patch polygons plus confidence.
- Scoring: `metric/scripts/score_amw.py` reuses the same loaders, CRS and cutoff as our own detector; figures from `make_amw_figs.py`.
- Everything above ships in the project bundle.

Backup: every site, every number

Site	Extent	Label dates	Onset lag	Coverage
S1	33.8 ha	5	+17 d / +2 ts	24 %
S2	30.7 ha	4	+7 d / +2 ts	14 %
S3	15.3 ha	31	+9 d / +1 ts	48 %
S4	3.9 ha	4	+0 d / +0 ts	74 %
S5	1.8 ha	1	+11 d / +1 ts	17 %
S6	1.5 ha	2	+14 d / +2 ts	43 %
S7	0.8 ha	1	+6 d / +1 ts	15 %
S8	0.3 ha	1	+0 d / +0 ts	36 %

Backup: why the alert needs two acquisitions

- **596 acquisitions**, 2018–2026. Median gap **4 days**, p90 **10 days**, worst **25 days**.
- But **49 % fall Nov–Apr**, snow and low sun, little usable signal for bare ground.
- Per-pixel cloud masking removes more. Two *usable* observations can span weeks.
- By construction: the accumulator must cross threshold with **all four indices agreeing**, one observation cannot trigger it.

Backup: how the sites were grouped

- 56 dated label snapshots, single-link clustered at **300 m**, named S1–S8 by final extent.
- Geometry only, the grouping does not depend on how anyone named the sites.
- Supersedes an earlier grouping that mixed analyst names with a 500 m rule and produced sites that interleaved along the same river reach.
- Detection-level results, precision, error budget, extent precision, are **unaffected** by the grouping.

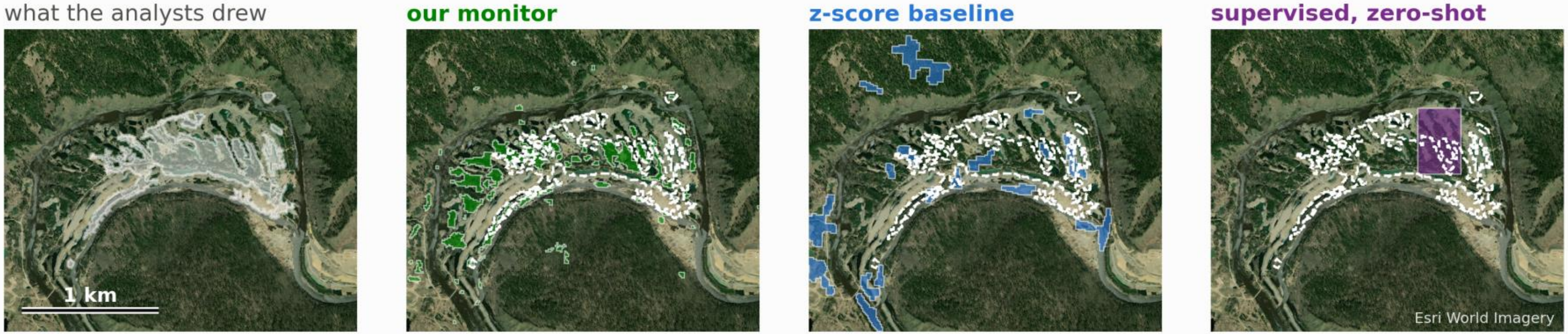
Backup: site by site, overlaid

The next eight slides put the analysts’ extent, our detections and the z-score’s detections on the same imagery, one site at a time. White dashed outline is the labelled extent in every panel.

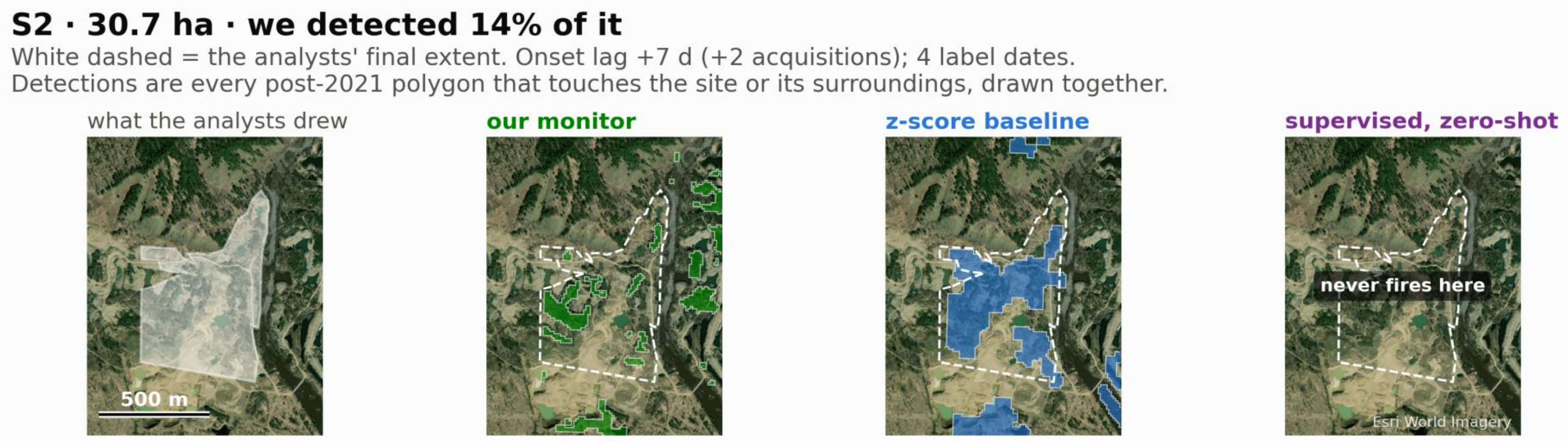
Backup: S1 · 33.8 ha · 24% detected

S1 · 33.8 ha · we detected 24% of it

White dashed = the analysts' final extent. Onset lag -48 d (-11 acquisitions); 5 label dates.
Detections are every post-2021 polygon that touches the site or its surroundings, drawn together.



Backup: S2 · 30.7 ha · 14% detected



Backup: S3 · 15.3 ha · 48% detected

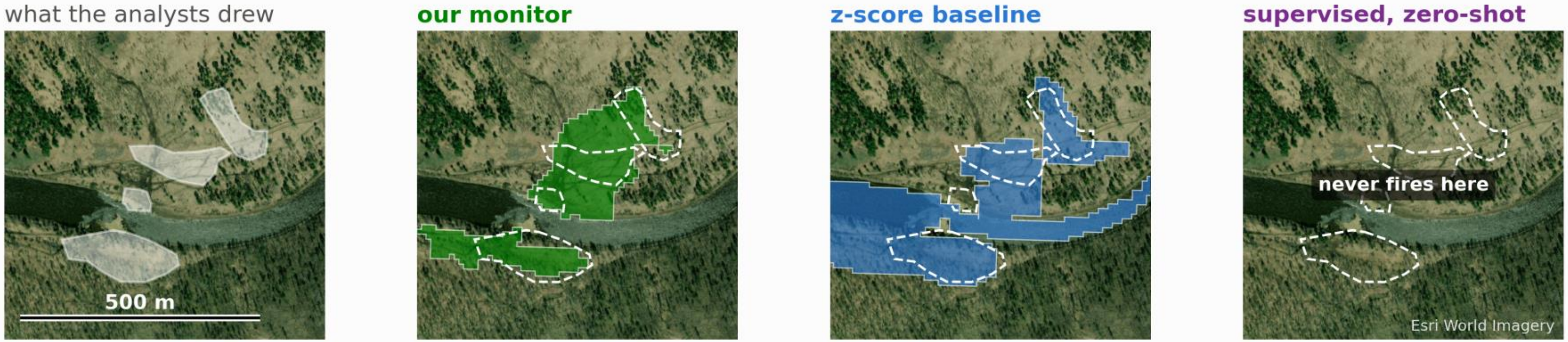
S3 · 15.3 ha · we detected 48% of it

White dashed = the analysts' final extent. Onset lag +9 d (+1 acquisitions); 31 label dates.
Detections are every post-2021 polygon that touches the site or its surroundings, drawn together.

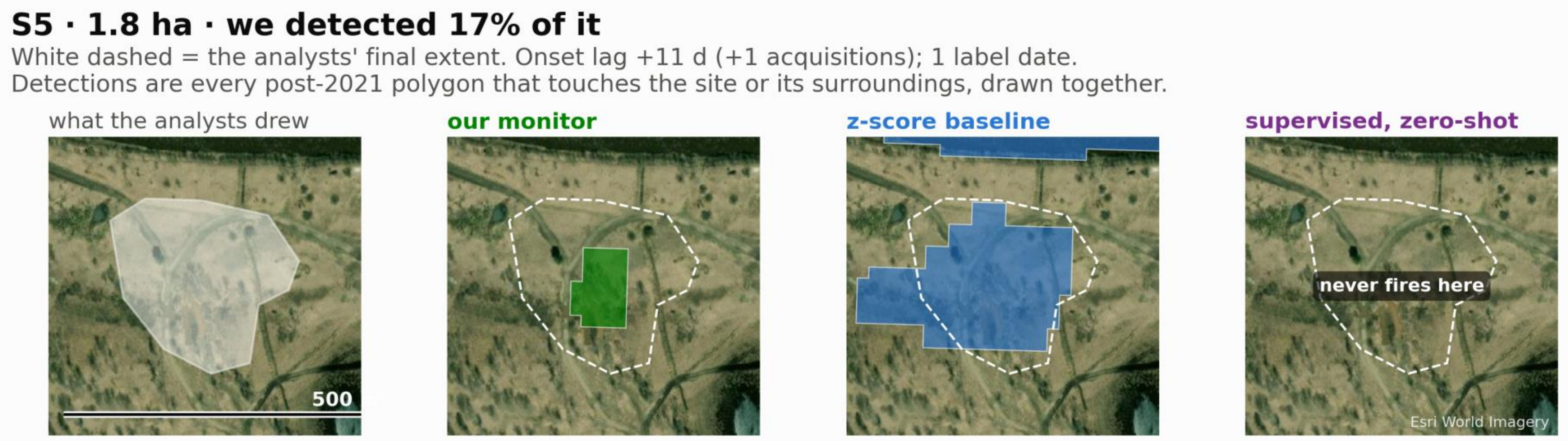
Backup: S4 · 3.9 ha · 74% detected

S4 · 3.9 ha · we detected 74% of it

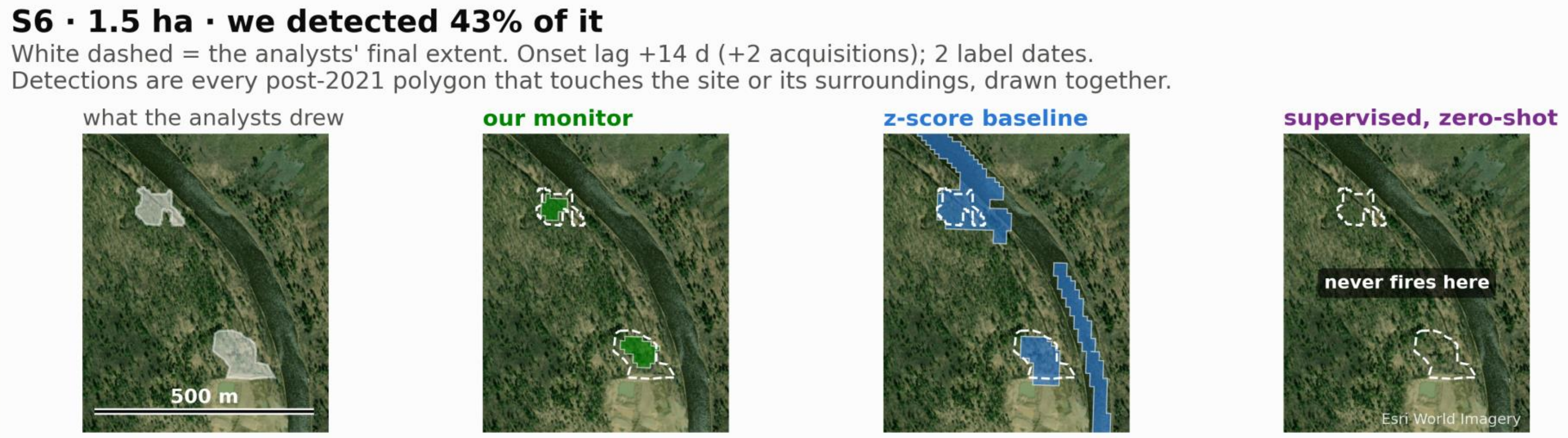
White dashed = the analysts' final extent. Onset lag +0 d (+0 acquisitions); 4 label dates.
Detections are every post-2021 polygon that touches the site or its surroundings, drawn together.



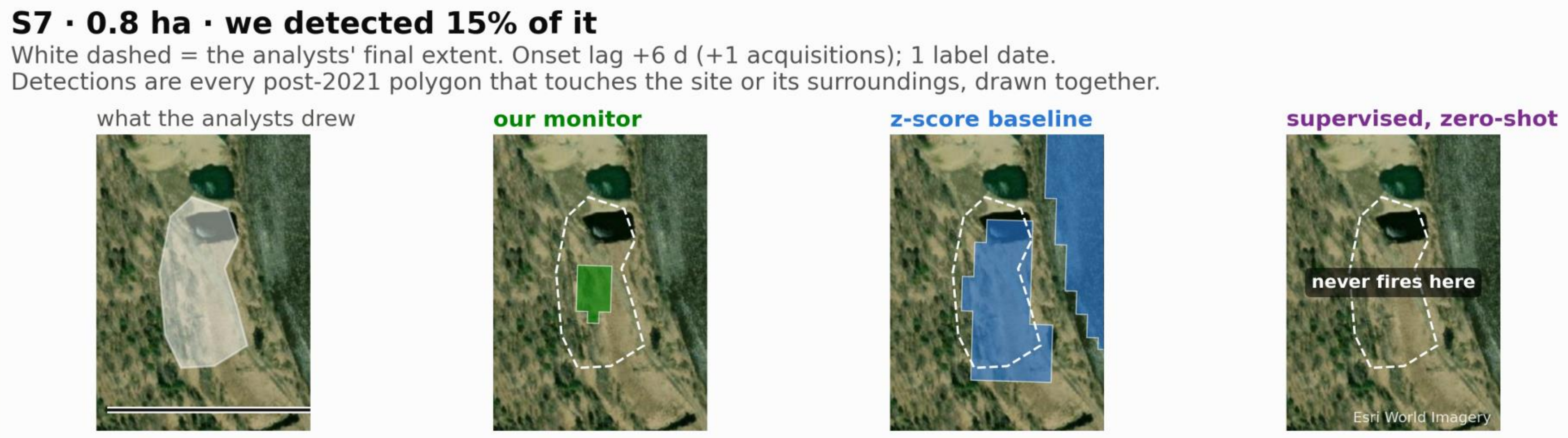
Backup: S5 · 1.8 ha · 17% detected



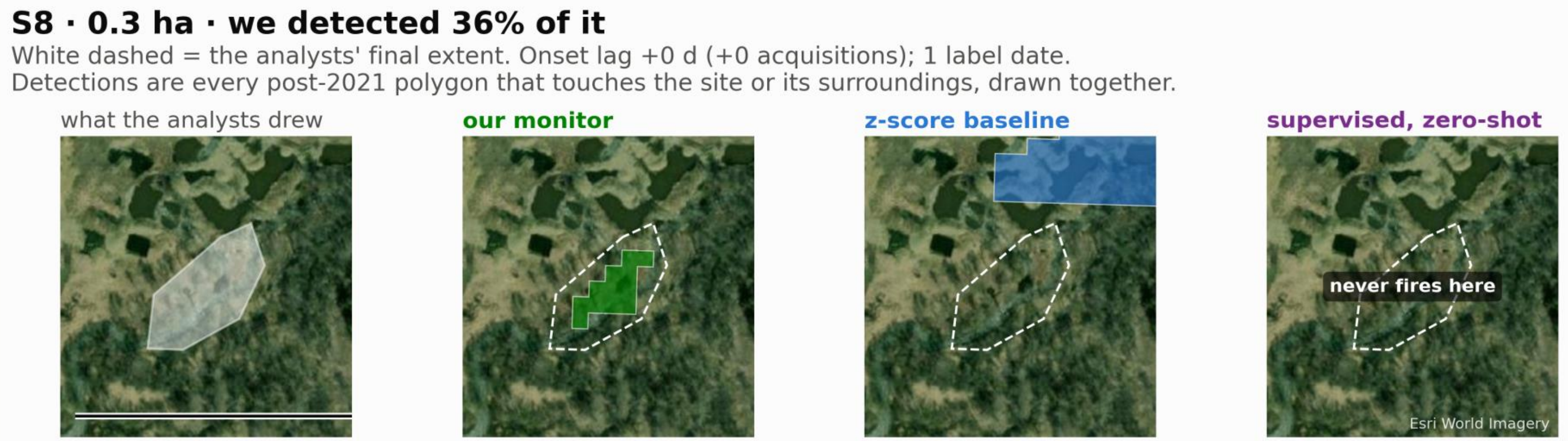
Backup: S6 · 1.5 ha · 43% detected



Backup: S7 · 0.8 ha · 15% detected



Backup: S8 · 0.3 ha · 36% detected



Backup: what the false detections actually are

Every off-label detection was clustered and reviewed in context. The next slides show the largest clusters for each cause the reviewers assigned.

Backup: Cloud and shadow

Cloud and shadow — 9.6 % of detections

Thin cirrus and haze pass the scene classification mask. A stricter or probabilistic mask, with buffered shadows, is the remedy.

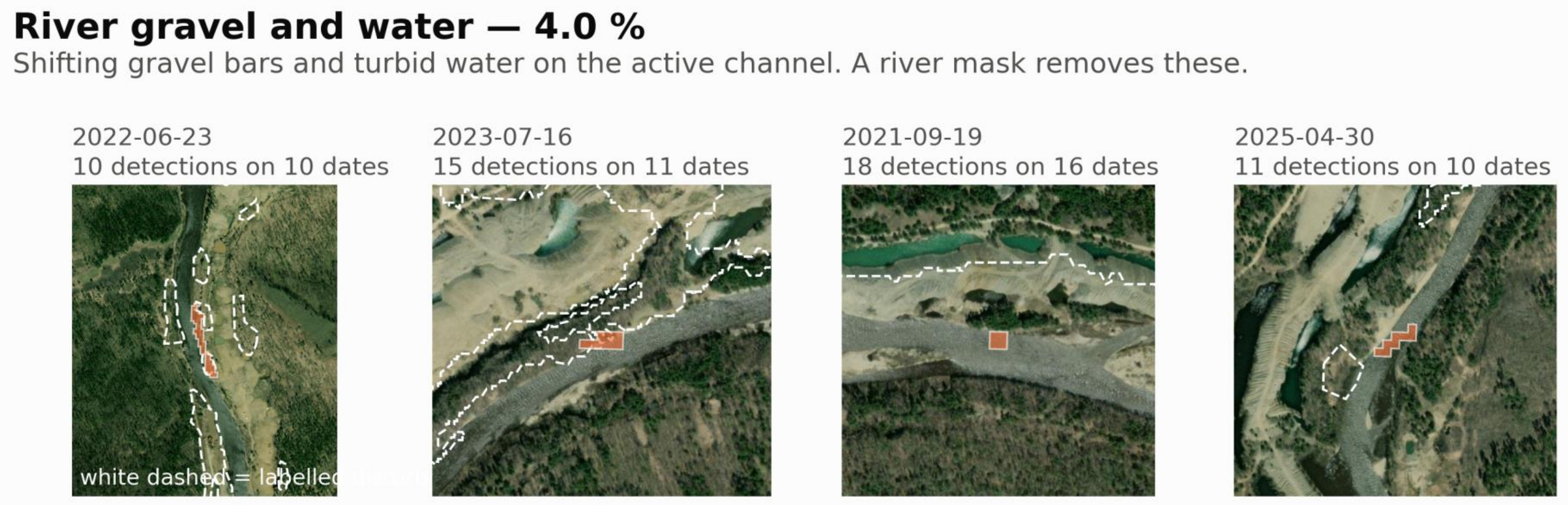
2021-06-28
3 detections on 2 dates

2021-05-17
1 detection on 1 date

2021-06-18
1 detection on 1 date

2020-07-23
1 detection on 1 date

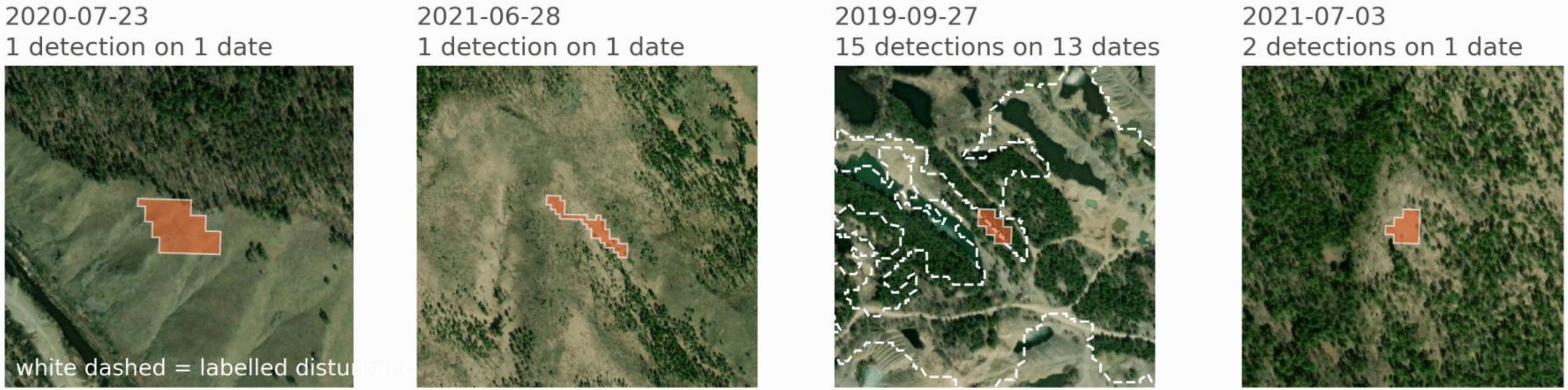
Backup: River gravel and water



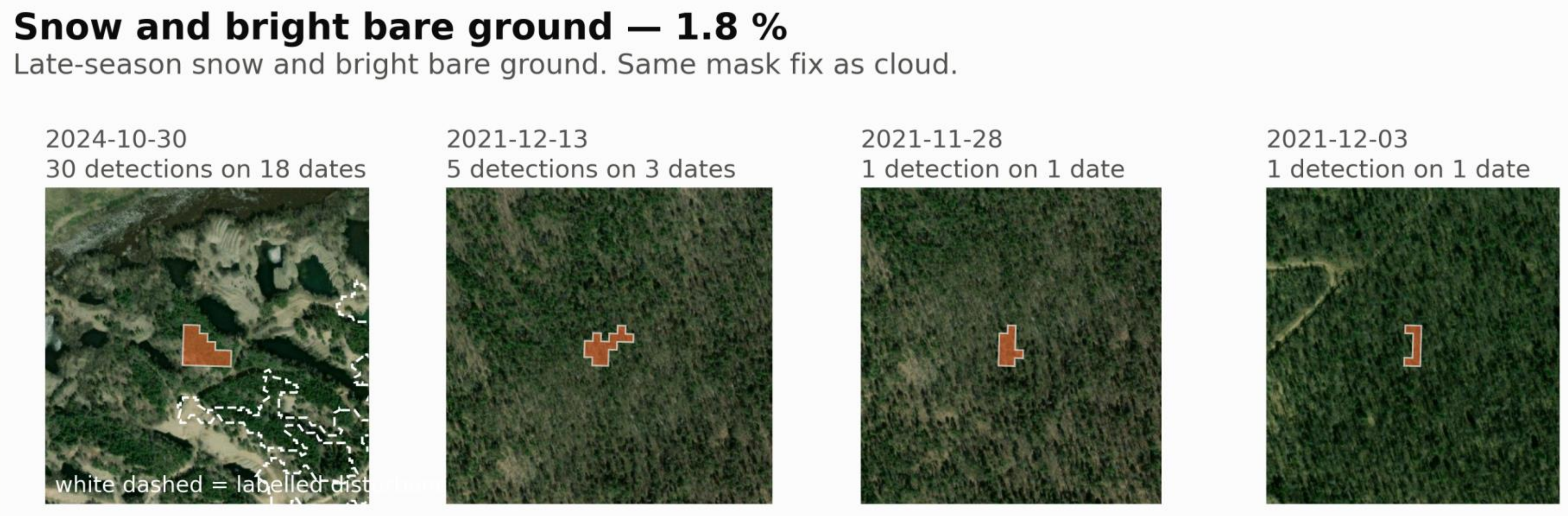
Backup: Off-site, unresolved

Off-site, cause not resolved — 3.9 %

Off-river, recurring, and not attributable — this is the genuine residual, about 4 % of detections.



Backup: Snow and bright ground



Backup: Confirmed disturbance outside the labels

Called real disturbance by the reviewer, outside the labels

These are why the truth set grows: ground the labels missed that a reviewer confirmed as disturbance in context.