

EEFIT Colombia 2026: Remote Sensing Support Report

Draft remote sensing chapter for the UK Earthquake Engineering Field Investigation Team (EEFIT), Institution of Structural Engineers (IStructE), mission to the Mw 7.4 Colombia earthquake of 10 August 2026, prepared ahead of the 15 to 26 September 2026 field window.

Pietro Milillo (University of Houston), Riccardo Vitale, Brandon Voelker

5 September 2026

Provisional notice. This report is assembled ahead of the field mission. NISAR products cited throughout are provisional version 1 products and are labelled as such wherever they appear, and results resting on 41 to 46 labelled buildings in one city are labelled provisional where they are stated. Parts of this report were drafted and checked with AI assistance (PeerPrism agent suite, Claude models) under the supervision of the remote sensing group; every number was verified against the project files named after it (see section 9).

U N I V E R S I T Y of *The Institution*
HOUSTON of *Structural*
Engineers

Table of contents

Right-click and choose Update Field to build the table of contents.

1. Summary for the field team

We support the EEFIT (Earthquake Engineering Field Investigation Team) mission to the Mw 7.4 Colombia earthquake of 10 August 2026, 108 km deep beneath Chocó. The field window is 15 to 26 September 2026 and the cities are Quibdó, Armenia, Pereira, Manizales, Quimbaya, Sevilla and Cali.

Two radar satellites imaged all seven cities before and after the earthquake: Sentinel-1, the European C-band synthetic aperture radar (SAR), and NISAR, the provisional L-band radar of NASA (National Aeronautics and Space Administration) and ISRO (Indian Space Research Organisation). Coherence measures how alike two radar images of the same ground are; collapse destroys it, and so do rain and vegetation, so a drop is damage evidence only after calibration. We compute it only where a footprint carries a valid radar pixel, a minority everywhere (Table 1); the Sentinel-1 screen is a first pass on one co-event pair, 8 to 14 August.

Pereira and Dosquebradas is the calibration city, the only one with released building-level products and reviewer labels. From each labeled building we subtract the median drop of its at least five unlabeled neighbors within 200 m, so a neighborhood-wide drop cancels. On Microsoft's 'ground_truth' class, damaged at a desk from imagery and not field verified, the NISAR excess is +0.135 against +0.061 for Sentinel-1; divided by each sensor's own unflagged interquartile range, 0.96 against 0.33. On the 27 buildings carrying both sensors NISAR exceeds Sentinel-1 by a median of +0.168, and on the independent CEMS grades Sentinel-1 has no excess (Table 1). The Sentinel-1 drop is a provisional neighborhood-scale signal, tied to damage in Pereira and nowhere else.

Cali needs its own layer: in its two graded areas the released Microsoft product flags none of the 12 Copernicus Emergency Management Service (CEMS) points on a Microsoft footprint, and has no footprint within 15 m of the other 9 of 21. In Sevilla 57.6 percent of 250 m cells exceed a mean drop of 0.15 on Sentinel-1 and none on NISAR; the same cut is applied to both sensors and NISAR flags fewer cells in every city, so Sevilla is the extreme of a systematic offset.

Carry three products: the provisional NISAR per-building drop for visit lists; the Sentinel-1 250 m cell tables for the affected core, never for per-building triage; and the Pereira label pack on 91,266 building outlines as the sampling frame.

Calibrate the NISAR drop in Pereira (sections 5 and 8), build and field-check our own layer in Cali (sections 4, 5 and 8), and record georeferenced damage and no-damage observations in the four cities with no released third party product, Armenia, Manizales, Quimbaya and Sevilla (sections 3 and 8).

City	Released third party products	Buildings with a Sentinel-1 value	Buildings with a NISAR value (track, frame)	Product to calibrate	Action	File	Section
Pereira and Dosquebradas	Microsoft reviewed, Vantor, CEMS AOIO	22,731 of 91,266	9,948 (T047A, f004)	NISAR per-building drop (provisional)	survey whole 250 m cells by label class, 100 to 200 building	catalogue/s1_screen/out_nisar/Pereira_Dosquebradas_T047A_f004_buildings_dpm.gpkg; catalogue/s1_screen/out_experiments/pereira_features.gpkg	4, 5, 6, 8

	2				s per class where the class is large enough, with unflagged controls		
Cali	Micro soft, CEMS AOIO1 and AOIO3	48,126 of 158,489	26,385 (T148 A, f003)	our Sentine l-1 co-event drop at cell level, and the CEMS points	field-check the CEMS AOIO3 destroyed points; use the co-event drop columns only, not the post-event or recovery columns, and read the layer by cell rather than building by building	catalogue/s1_screen/out20/final/Cali_buildings_s1_final.gpkg (columns 's1_dpm_late', 'source_burst'); catalogue/s1_screen/out20/final/Cali_cells250.csv	4, 5, 8
Sevilla	none released	2,165 of 6,726	762 (T047 A, f004)	Sentine l-1 250 m cells	walk three high and three low cells of the 66 that hold at least five buildings	catalogue/s1_screen/out20/final/Sevilla_cells250.csv	5, 8
Quibdó	CEMS AOIO4 graded points	8,660 of 34,066	1,848 (T011 D, f086)	Sentine l-1 250 m cells	record demolition dates and rainfall; use the 63 CEMS graded points as labels	catalogue/s1_screen/out20/final/Quibdo_cells250.csv	5, 8
Armenia	none released	24,406 of 106,040	7,947 (T047 A, f004)	none released	georeferenced damage and no-damage observat	catalogue/s1_screen/out20/final/Armenia_buildings_s1_final.gpkg	3, 8

					ions, includin g the 695 of 4,976 two- burst overlap building s assigned to burst 303907		
Manizal es	none releas ed	16,70 8 of 67,01 4	6,049 (T047 A, f004)	none release d	as Armeni a, includin g the 753 of 3,842 two- burst overlap building s assigned to burst 303905	catalogue/s1_screen/out20/final/Manizales_buildings_ s1_final.gpkg	3, 8
Quimba ya	none releas ed	2,122 of 7,707	953 (T047 A, f004)	none release d	georefer enced observat ions stratifie d by building height class	catalogue/s1_screen/out_heights/buildings_heights_Qu imbaya.gpkg; catalogue/s1_screen/out20/final/Quimbaya_cells250.csv	5, 7, 8

Table 1. What to survey in each of the seven cities and the file that carries it. The message of the table is that NISAR is the only product worth calibrating building by building, only Pereira can calibrate it, and in the other six cities the product the mission delivers is the georeferenced observation. **Coverage.** The third and fourth columns give the buildings that carry a value against the open building outlines of the city, 23 to 32 percent for Sentinel-1 in every city, because only a footprint whose centroid falls on a valid radar pixel receives one. A sample drawn from the full outline layer will therefore draw mostly buildings with no drop at all. **Label gradient behind the Pereira result.** Against the same 200 m baseline the NISAR excess is +0.141 on CEMS destroyed, +0.085 on CEMS damaged and +0.048 on Microsoft confirmed, while the Sentinel-1 excess is +0.023 on Microsoft confirmed and between -0.046 and +0.004 on all three CEMS classes, none of them significant. **Restrictions.** Every NISAR entry is a provisional product; the Pereira per-building calibration rests on 41 to 46 labeled buildings; the Sentinel-1 screen is a first pass on one co-event pair, and adding the wet 21 to 27 July pair would lower the city drop median by 0.043 to 0.083, of the same order as the 0.15 cell cut; and in Cali the two Sentinel-1 bursts agree on the per-building drop at Spearman 0.138 over their 7,555 building overlap, which is the repeatability floor for any statement about a single Cali building. **Action column sources.** Sevilla three high and three low cells of 66 with at least five buildings, Armenia 695 of 4,976 and Manizales 753 of 3,842 overlap buildings on the second burst; Sevilla cell count also in catalogue/s1_screen/out20/final/s1_nisar_cells_250m.csv, row "Sevilla", column "n_cells_ge5" (66); Quibdó 63 graded CEMS points; Pereira 100 to 200 buildings per class. **Glossary.** AOI is a Copernicus area of interest; T and f are the satellite track and frame that identify a NISAR interferogram; Vantor, formerly Maxar, released an automatic damage assessment over Pereira that flagged 613 of 75,262 footprints. Data providers: NASA, ISRO and the Alaska Satellite Facility (ASF) for the provisional NISAR products, Copernicus Sentinel-1, Microsoft AI for Good, Vantor, the Copernicus

Emergency Management Service, Overture Maps and the Technical University of Munich (TUM) GlobalBuildingAtlas. Source: the section named in the last column.

Sources: `AGREEMENT\_FRAMEWORK.md`, `COLOMBIA\_RECON\_CATALOGUE.md`,
`Emails/RE: Remote team support to Colombia.pdf`, `SAR\_SCREEN\_RESULTS.md`,
`catalogue/agreement/agreement\_log.txt`,
`catalogue/s1\_screen/out20/final/Armenia\_buildings\_s1\_final.gpkg`,
`catalogue/s1\_screen/out20/final/Cali\_buildings\_s1\_final.gpkg`,
`catalogue/s1\_screen/out20/final/Cali\_cells250.csv`,
`catalogue/s1\_screen/out20/final/Manizales\_buildings\_s1\_final.gpkg`,
`catalogue/s1\_screen/out20/final/Pereira\_Dosquebradas\_buildings\_s1\_final.gpkg`,
`catalogue/s1\_screen/out20/final/Quibdo\_cells250.csv`,
`catalogue/s1\_screen/out20/final/Quimbaya\_cells250.csv`,
`catalogue/s1\_screen/out20/final/Sevilla\_cells250.csv`,
`catalogue/s1\_screen/out20/final/build\_final\_log.json`,
`catalogue/s1\_screen/out20/final/fig12\_pereira\_label\_test\_stats.json`,
`catalogue/s1\_screen/out20/final/s1\_nisar\_cells\_250m.csv`,
`catalogue/s1\_screen/out20/qa2/Cali\_qa2\_results.json`,
`catalogue/s1\_screen/out\_experiments/pereira\_features.gpkg`,
`catalogue/s1\_screen/out\_experiments/pereira\_features\_README.md`,
`catalogue/s1\_screen/out\_experiments/pereira\_features\_summary.json`,
`catalogue/s1\_screen/out\_heights/buildings\_heights\_Quimbaya.gpkg`,
`catalogue/s1\_screen/out\_nisar/Pereira\_Dosquebradas\_T047A\_f004\_buildings\_dpm.gpkg`.

2. The event: source, shaking, loss and ground failure estimates, and ShakeMap at the seven cities

2.1 Source

The earthquake occurred on 10 August 2026 at 12:34:28 coordinated universal time (UTC) with moment magnitude 7.4, 2 km southeast of San José del Palmar, in the department of Chocó. The epicenter is at 4.8836 N and 76.2182 W and the hypocentral depth used by ShakeMap is 108.2 km. The Servicio Geológico Colombiano (SGC) locates the same event 14.3 km away, at 4.99 N and 76.29 W, and 103 km deep, within a depth range of 96 to 103 km. Both agencies place the rupture inside the subducting Nazca slab, and the reconnaissance catalogue characterizes it as an intraslab strike-slip event on the evidence of the USGS finite fault model.

The USGS finite fault inversion was run on a single plane of 150 km along strike by 66 km down dip, discretized into 25 by 11 cells of 6.00 km, with corners at 83.2 km and 144.0 km depth. That plane is the domain of the inversion and not the rupture. In the slip column of the same 275 cells, peak slip is 397.9 cm, 27 cells reach one fifth of the peak and 46 reach one tenth, which is 972 km² and 1,656 km² of the 9,900 km² plane, and all of them lie between 88.9 km and 106.0 km depth. That slip bearing band contains the SGC depth range of 96 to 103 km, so the two agency depths are consistent with one rupture reported at two reference points rather than with a disagreement. The two moment tensor nodal planes are strike 123.99, dip 58.63, rake 158.52 and strike 225.57, dip 71.78, rake 33.24; the second is the plane used by the finite fault inversion, and both rakes describe predominantly strike-slip motion.

The shallowest cell of that significant slip band sits 88.9 km below the surface, and the inversion grid itself carries no slip shallower than 83.2 km. The reconnaissance catalogue records that the event produced no measurable surface deformation. Consistently with that, the synthetic aperture radar (SAR) work reported

in sections 5 to 7 uses only wrapped-grid coherence magnitude and does not use the unwrapped interferometric grid at all. We therefore make no interferometric SAR (InSAR) displacement claim anywhere in this report. Every remote sensing layer we hand to the field team is a shaking damage proxy that has to be calibrated against what the team observes on the ground, not a deformation map.

2.2 Shaking

ShakeMap version 8 is the shaking product of record for this report. It is constrained by 67 seismic stations and 434 intensity observations. Across the whole grid the maxima are Modified Mercalli Intensity (MMI) 8.012, peak ground acceleration (PGA) 0.396 g, peak ground velocity (PGV) 32.225 cm/s, and spectral acceleration (SA) 0.704 g at 0.3 s, 0.558 g at 0.6 s, 0.485 g at 1.0 s and 0.172 g at 3.0 s. We report the short period bands beside 1.0 s because the stock the team will survey is expected to be low rise, which is our own engineering judgment and rests on no file under the data directory.

The catalogue records broad shaking of 90 s to 2 min over the Cauca valley and the coffee axis. Duration matters independently of the peak values: a minute and a half of cyclic demand accumulates damage in brittle wall systems that a single peak acceleration does not describe, so a building surveyed at the same PGA as one in a short crustal event may present a different damage state. That is engineering judgment; no file under the data directory maps a duration to a typology or a period band.

Four properties of version 8 travel with it. It carries an internal quality grade of C and a mean uncertainty ratio of 0.981, with 357 of the intensity observations and 66 of the ground motion records flagged. It is automatic and not reviewed; the SGC stations and all Did You Feel It (DYFI) reports entered at version 7 on 22 August. Its ground motion model is an event specific custom module with an empty reference field, and its intensity field comes from a virtual intensity prediction equation (IPE) rather than a calibrated regional relation, so a disagreement between this grid and a field observation is not evidence against the observation. And site amplification is thin: the model applies the ground motion prediction equation (GMPE) native site term on a default site condition of 760 m/s, and the only site velocities anywhere under the data directory are per station, spanning 221.27 to 900.0 m/s with six of the 67 stations sitting exactly on that default. No site velocity grid exists.

The largest recorded peak ground acceleration is 43.10 percent g at FLND, Filandia, with JAMC second at 42.69 percent g. FLND sits 69.99 km from the epicenter and 128.87 km from the hypocenter. Other measures have other leaders. The largest recorded peak ground velocity is 32.84 cm/s at CRU and 32.58 cm/s at CIRS, against 23.38 cm/s at FLND, and the largest instrumental intensity is 7.3, reached at both CRU and CIRS against 6.8 at FLND. CIRS is the near field leader on both, at 130.49 km hypocentral and 15.2 km from the Armenia city center. CRU is not: it lies 392.74 km from the hypocenter and 212.5 km from the Cali city center, on the 760.0 default site velocity assigned where no site datum exists. A station carrying the network maximum at that distance deserves the same instrument query the catalogue raises for the anomalously low CBOCA reading near Pereira.

The rule for reading the grid against the instruments differs by measure. Gridded PGA of 0.396 g sits below the recorded maximum of 0.431 g and gridded PGV of 32.225 cm/s just below the recorded 32.84 cm/s, so both accelerometric grids are smoothed envelopes and the station record is the local point constraint. Gridded MMI of 8.012 sits above every instrumental intensity, the largest of which is 7.3, because the intensity field is driven through the virtual IPE by the 434 macroseismic observations and not by the accelerographs; that the three do not move together is quantified by the 357 of 434 and 66 of 67 flagged records above. In the field, correct the PGA and PGV grids by the nearest record, and check the MMI grid against what the team sees rather than against the accelerographs.

A second ShakeMap exists and we do not hold it. SGC runs the USGS ShakeMap code with 119 stations, against the 67 here, and collected more than 20,000 Sismo Sentido felt reports, against 434 intensity

observations; no SGC ShakeMap product is present under the data directory. The mission lesson is that the remote team compares both ShakeMaps at each surveyed building, and the acquisition path sits on the same catalogue line: the SGC strong motion record bulletins, which the agency names RNAC, at bdrsnc.sgc.gov.co, recent records requested at l_rnac@sgc.gov.co. That request has to be made before 15 September for the comparison to be possible in the field, and until it arrives the team works from an automatic grade C product on an unpublished ground motion model, so instrumental disagreement should be recorded rather than resolved in favor of the map.

2.3 Loss estimates and reported impact

The USGS Prompt Assessment of Global Earthquakes for Response (PAGER) system issued a red alert for economic loss and an orange alert for fatalities. The empirical model gives a median estimate of 921 fatalities and USD 6.08 billion of direct economic loss. The fatality distribution is wide and belongs in the report whole: 0.326 on the 100 to 1,000 band, 0.317 on 1,000 to 10,000, 0.140 on 10,000 to 100,000 and 0.030 above that, so 0.487 on a thousand fatalities or more. Population exposure is 1,079,497 people at MMI VIII, 5,550,959 at MMI VII and 4,047,436 at MMI VI.

Reported impact figures, transcribed by us from public sources rather than derived from a product file and to be treated as reported rather than verified, are the balance of 31 August from the national disaster risk management agency (Unidad Nacional para la Gestión del Riesgo de Desastres, UNGRD): 331 dead, 4,505 injured, 136 missing, 198,720 homes damaged and 37,255 destroyed, 5,849 buildings damaged and 646 collapsed across 16 departments and 497 municipalities. The reported 331 falls inside the model's most probable band, the 100 to 1,000 band at 0.326, which is agreement rather than reconciliation. The economic side carries the opposite tension: the United Nations Development Programme (UNDP) rapid damage assessment RAPIDA gives 42.1 trillion Colombian pesos (COP), about USD 13.3 billion, against the PAGER figure of USD 6.08 billion. Neither model figure should be carried into the field as a target.

Pereira and Dosquebradas is the calibration city for the remote sensing layers, and the strongest supports for that choice are building counts. The Charter Pléiades Neo mapping gives Pereira 143 buildings destroyed and 585 damaged, against Manizales 15 and 170 and Cali 27 and 124, and the same RAPIDA departmental split puts Risaralda, the department Pereira sits in, at COP 14.4 trillion, the largest of any department, against Valle at 9.7 trillion. The collapse counts of 11 August compiled by Asocapitales, Pereira 92, Armenia 69, Cali 56, Manizales 17 and Quibdó 9, cover only the five departmental capitals among our seven cities and are absolute rather than normalized by building stock, so they rank Pereira first among those five and support no ranking over all seven. All three sources agree that Pereira carries the largest reported building damage on our list, which is why the released products and reviewer labels that make calibration possible exist there.

2.4 Ground failure

The USGS ground failure products, version 8 and computed on 2026-09-01, put liquefaction at a red alert and landsliding at an orange alert. Both headline quantities are defined at a threshold the alert label omits: the aggregate hazard is the area over which the model exceeds 0.1 g and the exposure is the population inside it. The preferred liquefaction model, Zhu and others 2017 general, gives 2,700 km² with a standard deviation of 65.4 km² and 460,000 people with a standard deviation of 58,042, against limiting values of 12,133 km² and 2,200,184 people. The preferred landslide model, Nowicki Jessee and others 2018, gives 24.0 km² with a standard deviation of 9.83 km², 41 percent of the value, and 1,700 people with a standard deviation of 52.8, against a limiting 1,491 km². The run carries a point source flag and a rupture warning, so the spatial pattern of both models follows the ShakeMap and not a finite rupture geometry.

Liquefaction outranks landsliding by 112 to 1 on hazard area and 270 to 1 on exposure, both taken on the alert values at the same 0.1 g threshold and computed by us; the two are different model formulations, so the ratio is indicative rather than a measured contrast, and the priority holds on either quantity. Sampling

the two model rasters inside the seven city boxes locates it. The liquefaction model reaches a maximum probability of 0.44 and a box mean of 0.21 over Quibdó, 0.36 and 0.11 over Cali, 0.21 and 0.008 over Pereira and Dosquebradas, and at most 0.07 over the four remaining cities, while the landslide model reaches at most 0.015 in any of the seven boxes. The geotechnical sub-team should carry the overlay named in the product file, zhu_2017_general.png, and document liquefaction first in Quibdó, second in Cali and third in Pereira, where the model reaches 0.21 only locally. The catalogue attaches this alert regionally to Cauca valley alluvium, but the highest modeled probability among the seven is at Quibdó, on the Atrato, so the survey should follow the raster and not the regional phrase.

2.5 ShakeMap at the seven cities

Table 2.1 gives the ShakeMap version 8 mean and standard deviation at each city center and the nearest of the 67 strong motion stations. City center intensities span MMI 6.49 at Sevilla to 7.95 at Pereira and Dosquebradas.

City	MMI (std)	PGA g (ln std)	PGV cm/s (ln std)	SA 0.3 s g (ln std)	Nearest station and distance	Station PGA percent g	Station PGV cm/s	Station instrumental MMI
Pereira and Dosquebradas	7.95 (0.54)	0.239 (0.46)	24.4 (0.38)	0.414 (0.51)	CBOCA, 5.8 km	1.83	2.41	3.9
Quibdó	7.77 (0.63)	0.315 (0.58)	24.7 (0.49)	0.539 (0.63)	SOL, 102.1 km	3.40	5.64	4.8
Armenia	7.41 (0.65)	0.260 (0.50)	20.7 (0.41)	0.532 (0.54)	ARMEC, 4.1 km	22.62	17.58	6.4
Quimbaya	7.38 (0.87)	0.307 (0.59)	24.4 (0.49)	0.592 (0.62)	ARMEC, 13.2 km	22.62	17.58	6.4
Manizales	7.24 (0.60)	0.174 (0.45)	14.1 (0.36)	0.281 (0.50)	MAN1C, 2.4 km	28.86	17.28	6.4
Cali	6.82 (0.68)	0.161 (0.52)	23.7 (0.43)	0.336 (0.56)	RAC2C, 6.5 km	11.56	25.35	6.9
Sevilla	6.49 (0.94)	0.210 (0.64)	13.4 (0.53)	0.371 (0.67)	CTRUJ, 43.2 km	13.24	16.31	6.3

Table 2.1. ShakeMap version 8 at the seven city centers with its own uncertainty, and the nearest strong motion station in the whole network. The message of the table is that the seven cities fall inside 1.46 MMI units of each other while the model's uncertainty at a single city center reaches 0.94 MMI units and the nearest instrumental constraint ranges from 2.4 km to 102.1 km, so the difference between cities is small relative to what is known about any one of them. Ground motion columns are the ShakeMap grids sampled by us at the bounding box centroid of each city, the centroid being the midpoint of the minimum and maximum coordinates of the box. PGA, PGV and SA mean grids store the natural logarithm of the quantity and are exponentiated here, so their standard deviation columns are in natural log units while the MMI standard deviation is in intensity units. The sampled MMI values reproduce the seven city center intensities recorded independently in SAR_SCREEN_RESULTS.md section 6b to the stated precision, and the sampled global maxima reproduce the grid maxima of usgs_shakemap_info.json exactly. Station columns come from catalogue/strong_motion_stations_shakemap_v8.csv, rows selected as the minimum great circle distance from the city centroid over all 67 stations, computed by us; that distance is the "and distance" figure in the sixth column and differs from the file's own dist_to_city_km column, which is computed only inside each station's assigned city partition. Of the 67 rows, 66 carry SGC as their source and one, RUS, has an empty source field. Source: the files named in this caption.

The network wide recomputation changes which station is nearest to which city, with three consequences for survey planning. Quimbaya, which has no station assigned to it in the file's own partition, is 13.2 km from ARMEC, consistent with the 14 km in the catalogue, so it is instrumentally constrained. Quibdó at 102.1 km and Sevilla at 43.2 km are not, on any definition, so there the grid rests on macroseismic observations alone and every field observation carries more weight. And the stations near Pereira do not agree: CBOCA at 5.8 km reads 1.83 percent g and instrumental MMI 3.9, FLND at 15.9 km reads 43.10

percent g and 6.8, and CIRS at 20.8 km reads 39.55 percent g and 7.3. With CBOCA flagged for an instrument check, the MMI of 7.95 at the Pereira center rests on intensity observations more than on the three nearby records.

Fig. 1 maps the two panels behind that table, the mean intensity and its uncertainty, with the seven city boxes and the whole station network on top. The grid cannot rank buildings inside any city, and the reason is its construction rather than the size of the spread it shows. It holds 685 by 685 points spaced 0.0090259 by 0.0089932 degrees, which at the latitude of the coffee axis is 1.00 by 0.99 km, about one square kilometer per cell. In Pereira and Dosquebradas the 91,266 building footprints of our feature table fall into 346 distinct ShakeMap cells, about 264 footprints per cell, so most neighbors receive an identical modeled value by construction. Like for like, the modeled MMI standard deviation across those footprints is 0.403 against 0.47 across the seven city center values of Table 2.1, computed by us, so the within-city and between-city spreads are of the same order. Both are smaller than the ShakeMap's own median MMI standard deviation at those same footprints, 0.649. A field whose stated uncertainty exceeds its own variation cannot separate one building from its neighbor, which is the gap the coherence products of sections 5 to 7 exist to fill.

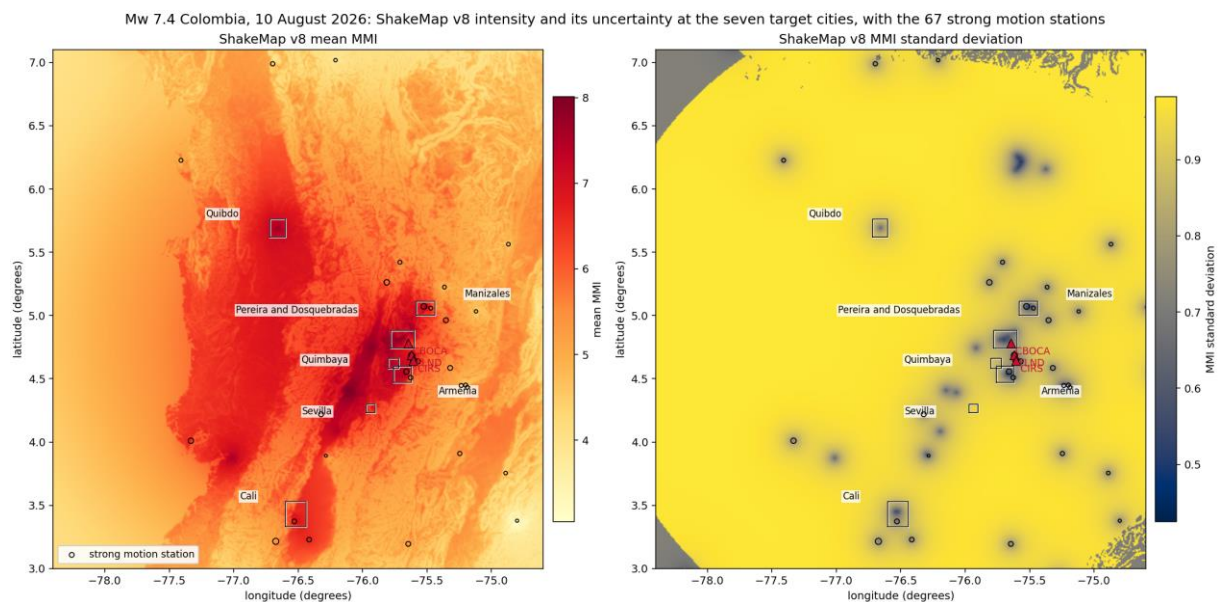


Fig. 1. two panels, ShakeMap version 8 mean MMI and its standard deviation, with the seven target city boxes and the 67 strong motion stations, symbols scaled by recorded PGA and CBOCA, FLND and CIRS labeled. The message of the figure is that all seven cities lie between MMI 6.49 and 7.95 at their centers while the model's own intensity uncertainty is largest at Sevilla and Quimbaya, and that the three stations near Pereira span 1.83 to 43.10 percent g. Data providers: USGS ShakeMap version 8, SGC station network, city boxes from our footprint table. Source: catalogue/usgs_products/shakemap_raster/mmi_mean.tif and mmi_std.flt with its.hdr, catalogue/strong_motion_stations_shakemap_v8.csv, catalogue/footprints/cities_bbox.csv.

2.6 Aftershocks during the field window

The USGS aftershock forecast issued on 2026-09-01 at 00:37 UTC covers the field window in its one month bin, 2026-08-31 at 16:00 UTC to 2026-10-01 at 16:00 UTC. For that bin the model gives probability 0.229 of at least one magnitude 5.0 or larger, with a 95 percent range of 0 to 2 events, 0.025 for magnitude 6.0 or larger with 0 to 1, and 0.003 for magnitude 7.0 or larger. Nine aftershocks of magnitude 3.0 or larger and one of 5.0 or larger had been recorded when it was issued.

Two facts bound that probability and have to travel with it. The forecast applies only inside a circle of radius 125.4 km centered on 4.6997 N and 76.5317 W; distances from that center to the seven city centroids, computed by us as great circles from catalogue/footprints/cities_bbox.csv, are Sevilla 81.9 km, Quimbaya 86.0 km, Pereira and Dosquebradas 94.1 km, Armenia 95.1 km, Quibdó 111.0 km, Manizales 119.9 km and Cali 141.2 km, so Cali lies outside the region and five of the six inside it sit beyond two thirds of the radius. And the sequence is intraslab at about 100 km depth, as section 2.1 records, so a magnitude 5.0 inside that circle is not a magnitude 5.0 felt as damaging shaking at a surveyed building. The forecast is a reason to agree an evacuation rule before entering a damaged structure, not a reason to expect a specific ground motion.

The hazard products place all seven cities inside a narrow band of modeled intensity and put Pereira and Quibdó at its upper end, but the grid resolves about one square kilometer, its own intensity uncertainty at a city center exceeds the spread between cities, and two cities have no station closer than 43 km, so section 2 can tell the field team which cities carry the strongest modeled and reported damage and cannot rank buildings inside any of them, which is why the coherence products of sections 5 to 7 and the calibration sampling frames of section 8 exist.

Sources: `COLOMBIA\_RECON\_CATALOGUE.md`, `LESSONS\_LEARNED.md`,
`SAR\_SCREEN\_RESULTS.md`, `catalogue/footprints/cities\_bbox.csv`,
`catalogue/s1\_screen/out\_experiments/E4\_shakemap\_variation.csv`,
`catalogue/s1\_screen/out\_experiments/E8\_shakemap\_variance.csv`,
`catalogue/s1\_screen/out\_experiments/pereira\_features\_summary.json`,
`catalogue/s1\_screen/out\_zerohour/layers/jessee\_2018\_model.tif`,
`catalogue/s1\_screen/out\_zerohour/layers/zhu\_2017\_general\_model.tif`,
`catalogue/strong\_motion\_stations\_shakemap\_v8.csv`,
`catalogue/usgs\_products/shakemap\_raster/mmi\_mean.tif`, `catalogue/usgs\_products/usgs\_event.json`,
`catalogue/usgs\_products/usgs\_finite-fault\_basic\_inversion.param`,
`catalogue/usgs\_products/usgs\_ground-failure\_info.json`,
`catalogue/usgs\_products/usgs\_losspager\_alerts.json`,
`catalogue/usgs\_products/usgs\_losspager\_exposures.json`,
`catalogue/usgs\_products/usgs\_losspager\_losses.json`, `catalogue/usgs\_products/usgs\_moment-tensor\_quakeml.xml`,
`catalogue/usgs\_products/usgs\_oaf\_forecast.json`,
`catalogue/usgs\_products/usgs\_shakemap\_info.json`,
`catalogue/usgs\_products/usgs\_shakemap\_stationlist.json`, `figures/fig0\_shakemap\_cities\_stations.png`.

3. What was acquired: imagery and synthetic aperture radar timeline by provider and date

This section states which sensor saw which city on which date, and which acquisitions we turned into a product. The seven target cities are Quibdó, Pereira and Dosquebradas, Armenia, Manizales, Quimbaya, Sevilla and Cali. The archive is large, and the constraint is which acquisitions form usable interferometric pairs and which cities carry a released damage product (3.2, 3.3, 3.6). The two synthetic aperture radar (SAR) stacks behind section 5 do not share a temporal baseline. Copernicus Sentinel-1, distributed through the Copernicus Data Space Ecosystem (CDSE) and the Alaska Satellite Facility (ASF), brackets the earthquake with a 6-day pair, the provisional products of the National Aeronautics and Space Administration (NASA) and Indian Space Research Organisation (ISRO) Synthetic Aperture Radar mission (NISAR) only with a 24-day pair, and three of the seven cities, Armenia, Quimbaya and Sevilla, received no building-level damage product.

3.1 The unified acquisition timeline

We harvested nine acquisition programs into a single table of 1,023 acquisitions covering 9 November 2025 to 4 September 2026. Of these, 357 are pre-event, 613 are post-event and 53 were acquired on 10 August itself, the day of the earthquake. Table 3.1 gives the split by program, with the post-event column reporting the two post-event classes together and the same-day count in parentheses, so its total of 666 is 613 plus 53. On the same time base we record 28 dated product releases between 10 August and 3 September across five platform groups holding 34 sensor lanes, tracked by a feed of 329 items from 24 sources. Fig. 2 draws those 34 lanes against the earthquake and the field window.

The two remaining agency acronyms in the program names are the United States Geological Survey (USGS) and the Copernicus Emergency Management Service (CEMS), the European Union's satellite mapping service, which the acquisition file names Copernicus EMS.

Table 3.1. Catalogued acquisitions by program. Program names are the strings the source file carries, so the access route stays visible and each row can be traced back by name. Source: ``catalogue/imagery_acquisitions_timeline.csv`` (columns ``program``, ``phase``, ``date_utc``).

Program (as named in the source file)	Acquisitions	Pre-event	Post-event (of which same day)	First to last date
International Charter act. 1048 (call 1202)	624	175	449 (34)	2025-11-09 to 2026-09-01
Copernicus Sentinel-2 (CDSE)	118	43	75 (15)	2026-07-21 to 2026-09-04
Copernicus Sentinel-1 (ASF/CDSE)	82	56	26 (0)	2026-07-02 to 2026-09-01
NASA-ISRO NISAR (ASF, provisional)	78	49	29 (0)	2026-06-17 to 2026-09-01
USGS Landsat C2 L2 (Planetary Computer)	46	24	22 (0)	2026-07-19 to 2026-08-29
OpenAerialMap	36	5	31 (1)	2026-07-16 to 2026-09-03
Vantor (Maxar) Open Data Program	22	4	18 (1)	2026-07-16 to 2026-09-03
Copernicus EMS EMSR916	9	0	9 (0)	2026-08-11 to 2026-08-14
Planet Crisis Response (source.coop)	8	1	7 (2)	2026-04-01 to 2026-08-18
Total	1,023	357	666 (53)	2025-11-09 to 2026-09-04

Take-away: the imagery record is dense and open, so the constraint is which acquisitions form usable interferometric pairs and which cities carry a released damage product.

3.2 Sentinel-1: C-band with a 6-day co-event pair

The Copernicus Sentinel-1 archive holds 82 single look complex (SLC) scenes over the target area between 2 July and 1 September 2026 on five tracks. Four of them acquire repeatedly: 142 and 69 descending, 48 and 121 ascending. The fifth, track 40 descending, made a single pre-event pass on 2 July 2026 over two frames and never returned, so it forms no pair and enters no product. The first post-event pass was on 12 August at 23:29 Coordinated Universal Time (UTC) on track 121; two tracks give a 6-day pair bracketing the earthquake, 142 from 8 to 14 August and 48 from 7 to 13 August.

We processed 13 burst and subswath products, drawn from 11 distinct bursts of track 142 because bursts 303901 and 303902 appear once in interferometric wide swath subswath 2 (IW2) and once in IW3, at seven dates, 21 and 27 July and 2, 8, 14, 20 and 26 August 2026, giving three pre-event pairs, one co-event pair of 8 to 14 August and two post-event pairs. Twelve of the 13 pass the geolocation and signal check and one carries a warning: burst 303901 in IW2 covers only 15 percent of the Quibdó bounding box,

contributes 36 valid building pixels and enters 35 buildings, so it is a marginal contributor rather than an equal one.

Coherence was estimated in a 7 by 7 window and geocoded to 20 m, which puts the estimation window at 140 m by 140 m on the ground. That window delivers fewer independent samples than its 49 cells suggest: effective looks are 25 to 27, because bicubic geocoding correlates neighbouring pixels and the 22 m azimuth resolution exceeds the 20 m posting. Both the 20 m products of section 5 and the 10 m products of section 6 carry a geolocation offset of about 40 m east and 5 m south, because the Copernicus 30 m global digital elevation model (DEM), product GLO-30, supplies orthometric heights that were used as ellipsoidal ones, and the 10 m geocoding path uses the same uncorrected DEM. City and neighbourhood statistics are unaffected, because the offset is small against the 140 m window, but a single building's position on a map is not, so field teams should treat a flagged pixel as locating a block and not an address.

Two further limits set the scale at which the Sentinel-1 layers can be read. Per-building drops are noise-dominated, with adjacent-burst repeatability of Spearman 0.14 to 0.37, which is why we report Sentinel-1 at 250 m cell or neighbourhood-excess level rather than per building. Coverage is also biased toward large buildings: Pereira footprints have a median area of 61 square metres, 93 percent of them are smaller than one pixel, and only 25 percent contain a pixel centre. The same bias sets how many buildings carry a co-event Sentinel-1 value in each city, in Table 3.2.

Table 3.2. Sentinel-1 co-event coverage per city: buildings with a valid co-event coherence value against all Overture footprints in the city. The share is set by footprint size and by how much of each city its bursts cover, not by damage. Source: `catalogue/s1_screen/out20/final/s1_city_table.csv` (columns `'city'`, `'bursts'`, `'n_bldg'`, `'n_footprints'`); burst coverage shares from `catalogue/s1_screen/out20/final/SECTION_7.md` ("QA verdict per burst" table).

City	Bursts combined	Buildings with a co-event value	All footprints	Share
Cali	303914 IW2, 303915 IW2	48,126	158,489	30 %
Armenia	303907 IW1, 303908 IW1	24,406	106,040	23 %
Pereira and Dosquebradas	303906 IW1	22,731	91,266	25 %
Manizales	303904 IW1, 303905 IW1	16,708	67,014	25 %
Quibdó	303901 and 303902, IW2 and IW3	8,660	34,066	25 %
Sevilla	303909 IW1, 303910 IW1	2,165	6,726	32 %
Quimbaya	303907 IW1	2,122	7,707	28 %
All seven	13 burst and subswath products	124,918	471,308	27 %

Take-away: Sentinel-1 gives the only 6-day co-event pair over all seven cities (see 3.1), so its drop is the measurement closest in time to the shaking, but it reaches roughly a quarter of the footprints in each city and it is a neighbourhood measurement, not a per-building one.

3.3 NISAR: L-band with a 24-day co-event pair

NISAR, the NASA and ISRO L-band radar, delivered 78 frame-level acquisitions on 27 passes over the area between 17 June and 1 September 2026 on five tracks, T011, T083 and T112 descending and T047 and T148 ascending. We use the track and frame convention of the products, T011D and T047A, throughout; the timeline file writes the same tracks in short form as T11, T83, T112, T47 and T148, so the two can be matched. There was no acquisition anywhere in the archive between 25 July and 11 August: the last pre-event pass was on 25 July and the first post-event pass on 11 August at 10:46 UTC on track 148. That 17-day gap swallows exactly one 12-day repeat cycle on each track, which is why the bracketing pair

is 24 days and not 12. Four of the five tracks bracket the earthquake at 24 days; track 112 lost two cycles rather than one and brackets it at 36 days, from 15 July to 20 August. Every pair we use is 24 days long, against 12 days for the pre-event and post-event pairs.

Which frames we use is set by whether a frame covers any of the seven cities. In the granule inventory, both T083D frames (086 and 087), both T112D frames (085 and 086) and the T011D 085, T047A 003 and 005 and T148A 005 frames carry an empty city field, while T011D 086 and 087, T047A 004 and T148A 003 and 004 carry city names. T083 and T112 therefore contribute no city and are out on coverage alone. One city-covering combination remains unused: T148A frame 004 spans Quibdó and Sevilla and holds an 18 July to 11 August co-event pair, but it was not among the granules selected for download, all three of which are T148A frame 003 over Cali. It is an available and provisional co-event pair over the two cities where the Sentinel-1 and NISAR readings diverge most, and processing it is the cheapest addition available before the field window. We used the provisional geocoded unwrapped interferogram (GUNW) products of Table 3.3; all NISAR products here are provisional version 1.

Table 3.3. NISAR provisional GUNW pairs used, by track and frame. Source:

`catalogue/s1\_screen/out\_nisar/gunw\_city\_stats.csv` (columns `track`, `frame`, `ref`, `sec`, `phase`, `days`, `city`). All values provisional.

Track and frame	Pre-event pair (12 d)	Co-event pair (24 d)	Post-event pair (12 d)	Cities with a value
T011D frame 086	8 to 20 Jul	20 Jul to 13 Aug	13 to 25 Aug	Quibdó, Armenia, Pereira and Dosquebradas, Manizales, Quimbaya, Sevilla
T011D frame 087	8 to 20 Jul	20 Jul to 13 Aug	13 to 25 Aug	Cali
T047A frame 004	11 to 23 Jul	23 Jul to 16 Aug	16 to 28 Aug	Armenia, Pereira and Dosquebradas, Manizales, Quimbaya, Sevilla
T148A frame 003	6 to 18 Jul	18 Jul to 11 Aug	11 to 23 Aug	Cali

Fig. 3 maps the co-event pair of Table 3.3 over the seven cities.

Take-away: the provisional L-band against C-band comparison here is 24 days against 6 days (see 3.2), not equal baselines, so part of every NISAR co-event drop is ordinary temporal decorrelation, which section 5 bounds with a matched 24-day pre-event baseline.

3.4 OPERA (Observational Products for End-Users from Remote Sensing Analysis) products over the area

We downloaded 96 GeoTIFFs of the OPERA radiometric terrain corrected (RTC) Sentinel-1 product (RTC-S1), covering 19 bursts on 11 dates: nine bursts of track 48 at six dates each and 10 bursts of track 142 at four or five dates each, 54 files plus 42. The two tracks repeat on different days, so the file set is 96 burst-and-date combinations rather than a complete grid of one date list applied to every burst.

Over this area OPERA produced RTC-S1 and the surface water extent product (DSWx-S1) through 31 August and the land surface disturbance alert built on harmonized Landsat and Sentinel-2 data (DIST-ALERT-HLS) through 2 September, the coregistered single look complex product (CSLC-S1) only sporadically (one burst, 14 July), and neither the surface displacement product (DISP-S1) nor DIST-ALERT-S1; nothing event-specific was released, because NASA's event support for this earthquake went through its disasters response coordination system (DRCS) to the Colombian national disaster agency (UNGRD) rather than through public OPERA products.

Take-away: no ready-made OPERA change or displacement product exists for this earthquake and none is coming through the public route, so the backscatter change layer of section 5 is one we computed from RTC-S1 and keep only as a weak third proxy.

3.5 Optical, rapid mapping and aerial acquisitions

The International Charter on Space and Major Disasters was activated on 10 August at 16:23 UTC as activation 1048, requested by UNGRD; it accumulated 624 acquisitions and 23 published products. CEMS activated EMSR916 on 10 August at 17:13 UTC over seven areas of interest. Six were delivered as grading products between 12 and 14 August; the seventh, area 00 Western Colombia, was a wide-area Sentinel-1 product CEMS declared not feasible on 17 August at 10:24 UTC, reason "Because of remote sensing limitations".

The open optical and aerial record holds 118 Sentinel-2 Level 2A tiles, 46 Landsat Collection 2 Level 2 scenes, 22 Vantor (Maxar) Open Data items and eight Planet Crisis Response collections under Creative Commons Attribution NonCommercial 4.0 (CC BY-NC 4.0), and 36 OpenAerialMap images under Creative Commons Attribution 4.0 (CC BY 4.0) that carry the 10 cm emergency orthophotos of the Instituto Geográfico Agustín Codazzi (IGAC), Sevilla included. The Sentinel-2 tile list was built with a scene cloud filter of 60 percent. We set the threshold high rather than tight because the reported cloud fraction is a whole-tile figure while our targets are seven city footprints inside those tiles, so a tile at 50 percent cloud can still be clear over a city, and discarding it would have cost usable scenes.

Take-away: optical and aerial coverage limits no city, but the rapid-mapping products built on it concentrate on Pereira and Dosquebradas, Cali, Quibdó and Manizales, which is what section 4 examines.

3.6 What the archive does not contain

Three gaps matter.

First, no building-level damage product was released for Armenia, Quimbaya or Sevilla. None of the 23 Charter product titles names them and no CEMS area of interest covers them. The other four cities all carry one. Manizales is covered by Charter products only, titled "Colombia, Manizales Impact Map", whose released title carries a dash between the city and the product name that we do not reproduce here, and "Mapa de identificación de viviendas afectadas en Manizales, Caldas", with counts of 15 destroyed, 170 damaged and 124 possibly damaged buildings. Pereira and Cali carry both Charter and CEMS products, and Quibdó carries the CEMS "Quibdo Centre" grading product. Some Charter and CEMS products cover places outside our seven, Buenaventura and Istmina among them; they are useful context for the regional damage pattern and they are not part of the seven-city calibration frame of section 8.

Second, there is no open very high resolution (VHR) SAR over the area. Of the 31 radar acquisitions in the Charter collect, exactly four are very high resolution, all from TanDEM-X. Nine RADARSAT Constellation Mission (RCM) RCM-2, four RCM-3 and seven Gaofen-3 (GF3) acquisitions are classed high resolution, five EOS-04 acquisitions medium and two Sentinel-1C low. All 31 carry the same access string, "Charter authorised users only (UNGRD); products public", so the only very high resolution radar over these cities is four TanDEM-X acquisitions behind Charter authorization, and no part of the radar collect is openly redistributable.

Third, no damage proxy map (DPM) from NASA's Advanced Rapid Imaging and Analysis (ARIA) team was published for this event. With no ARIA product and no OPERA event product (3.4), the only L-band change layer over these cities is the provisional NISAR one we computed.

Take-away: Armenia, Quimbaya and Sevilla have imagery but no released product, so prioritization there rests on the radar layers we computed and their validation rests on the field team.

Figures

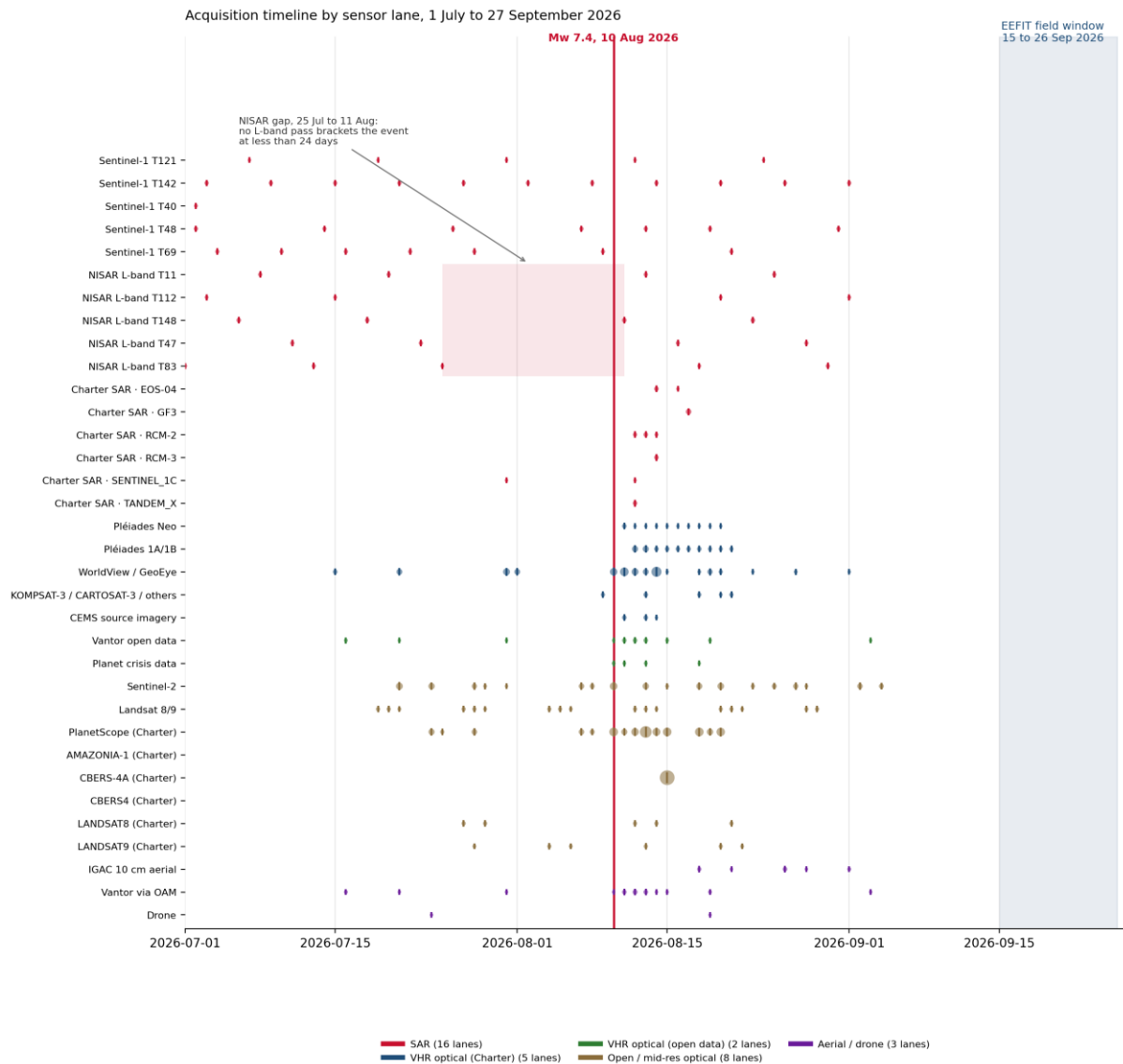


Fig. 2. acquisition timeline for the 34 sensor lanes of the five platform groups, with the 10 August 2026 earthquake and the 15 to 26 September 2026 field window marked; it shows the NISAR gap of 25 July to 11 August (see 3.3) against the uninterrupted 6-day Sentinel-1 cadence on track 142 (see 3.2). Two Charter optical lanes appear empty because their only acquisitions fall before the 1 July start of the plotted window. Data providers: Copernicus Sentinel-1 and Sentinel-2, NASA and ISRO NISAR (provisional), Charter 1048, CEMS EMSR916, Vantor, Planet, USGS Landsat, OpenAerialMap and IGAC. Source: `catalogue/timeline\_data.json` (keys `groups`, `event`, `field`, `start`, `end`).

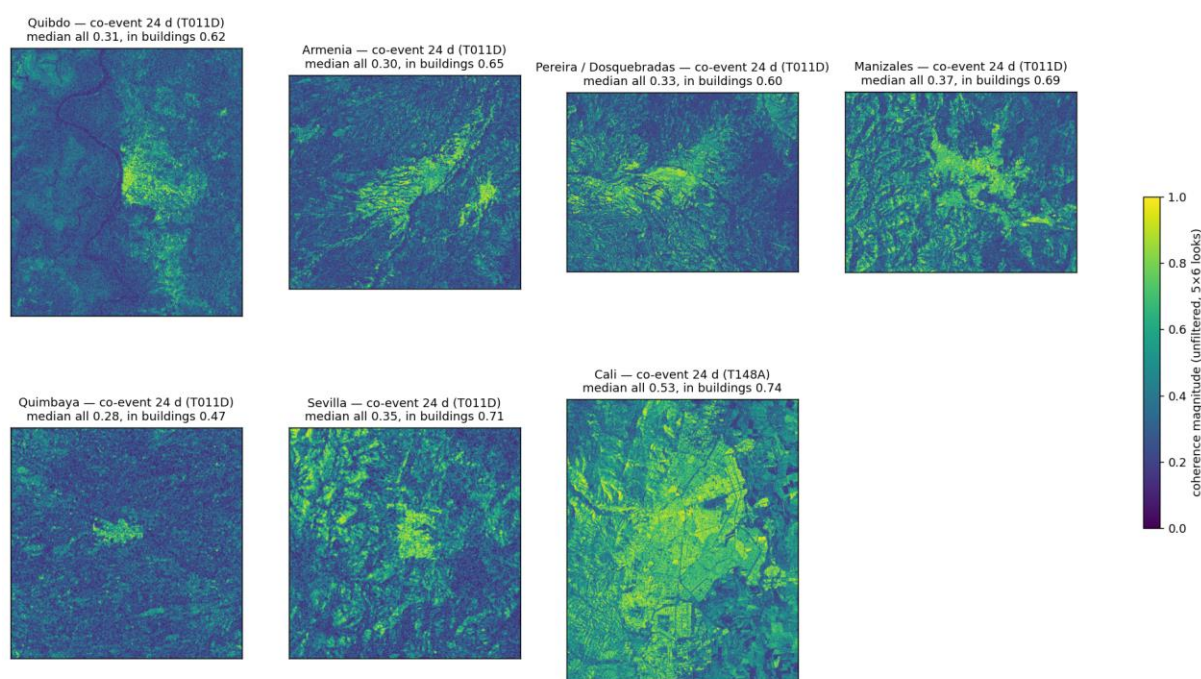


Fig. 3. the provisional NISAR 24-day co-event coherence pair for the seven cities, 24 days rather than 12 because NISAR did not acquire between 25 July and 11 August (see 3.3), so part of the drop is temporal decorrelation, not damage. Data providers: NASA and ISRO NISAR via ASF (provisional), Overture footprints. Source: `catalogue/s1\_screen/out\_nisar/\*\_co\_coh.tif`.

The full record is also browsable as the interactive atlas `choco\_recon\_atlas.html` at the root of the project data directory.

Section take-away

Sentinel-1 covers all seven cities with a 6-day co-event pair and the provisional NISAR products cover them with a 24-day one, but each Sentinel-1 layer carries a co-event value for only about a quarter of the buildings in a city and is read at neighbourhood scale. Released damage products cover four cities, Pereira and Dosquebradas, Cali, Quibdó and Manizales, while Armenia, Quimbaya and Sevilla have imagery and no product, so the field team is the only source of labels there.

Sources: `COLOMBIA\_RECON\_CATALOGUE.md`, `PROCESSING\_LOG.md`,
`SAR\_SCREEN\_RESULTS.md`, `catalogue/aoi/city\_aoi.csv`, `catalogue/asf\_s1.csv`,
`catalogue/cems\_api.json`, `catalogue/charter1048\_acquisitions.json`,
`catalogue/charter1048\_products.json`, `catalogue/imagery\_acquisitions\_timeline.csv`,
`catalogue/oam.json`, `catalogue/s1\_screen/nisar/gunw\_granules.csv`,
`catalogue/s1\_screen/nisar/gunw\_selected3.csv`, `catalogue/s1\_screen/out20/final/SECTION\_7.md`,
`catalogue/s1\_screen/out20/final/s1\_city\_table.csv`,
`catalogue/s1\_screen/out20/final/test\_burst\_geolocation\_table\_rerun.csv`,
`catalogue/s1\_screen/out\_nisar/\*\_co\_coh.tif`, `catalogue/s1\_screen/out\_nisar/gunw\_city\_stats.csv`,
`catalogue/s1\_screen/rtc`, `catalogue/timeline\_data.json`, `choco\_recon\_atlas.html`,
`figures/fig0b\_acquisition\_timeline.png`, `figures/fig8\_coevent\_24day\_seven\_cities.png`,
`report/site/feed.json`.

4. Damage products released and how they agree

Every released optical damage product covers only the ground its producer actually assessed, and over Pereira and Dosquebradas the three producers assessed grossly different ground: 254.28 km² for Microsoft, 77.14 km² for Vantor and 9.75 km² for the Copernicus area of interest AOI02, sharing 5.92 km². Every product comparison in this section is therefore reported twice, first inside that common valid area, where all three producers had an opinion about every building, and second over all footprints of the city, which is the comparison normally reported and which we carry alongside to show what the correction costs. The two views disagree in both directions. Inside the common valid area the two optical products share a Jaccard index of 0.445 and a Cohen kappa of 0.611 on 12,023 footprints, while over all 91,266 footprints the same pair returns 0.353 and 0.520, because the 56,224 footprints Vantor never looked at are counted as Vantor silence. The consequence for the field window of 15 to 26 September 2026 is unchanged in kind and sharper in degree: the released products are usable to choose which 250 m cells to visit, they are not usable to choose which building to knock on, they say nothing whatever about the ground outside their own masks, and in Cali they are not usable as a sampling frame at all.

The products that name individual buildings or points in the seven target cities were released by Microsoft AI for Good and by Vantor, both of them artificial intelligence (AI) models applied to very high resolution optical imagery, by the Copernicus Emergency Management Service (CEMS, the satellite mapping service of the European Union) under activation EMSR916, whose products are graded points inside numbered areas of interest (AOI), by the International Charter Space and Major Disasters under activation 1048 and its value adding partners ICube-SERTIT and the Operational Satellite Applications Programme (UNOSAT) of the United Nations Institute for Training and Research (UNITAR), and by the International Organization for Migration (IOM), whose Displacement Tracking Matrix (DTM) database merges the CEMS and Microsoft layers onto one footprint base in Cali and Pereira and carries Copernicus grades alone in Quibdó. Table 4.1 lists every one of them that names individual buildings or points in a target city, whatever form it is delivered in, because delivery form is what decides whether a product can be compared or loaded. Only two cities, Pereira and Cali, carry more than one product in a form we can compare, that is as per building or per point vectors rather than as a published map, so the question this section can answer is narrow: where two vector products describe the same buildings, do they flag the same ones. Three of the six comparison angles of our framework produce numbers here, footprint agreement on a shared footprint base, point to footprint matching within 15 m with its class conditional table, and density agreement on a 250 m grid. Two more, aggregation against official counts and validation against field grades, are what the mission supplies. Angle 5, ordinal consistency of continuous damage scores against the CEMS grades, is computable today because both delivered Vantor layers and the delivered Microsoft layer carry `damage\_pct\_0m`, `damage\_pct\_10m` and `damage\_pct\_20m` as continuous fields, and we defer it to the post-mission benchmark card, where it can be reported for the field grades and the released grades on the same axis.

Three of the seven target cities of the mission, Armenia, Quimbaya and Sevilla, have no building level product at all. Two further CEMS areas of interest, AOI06 Buenaventura with 322 graded residential buildings and AOI05 Istmina with 10, lie outside the field itinerary and are therefore absent from Table 4.1.

City	Product	Delivery form and footprint base	Size
Pereira and Dosquebradas	Microsoft AI for Good, reviewed, from Legion imagery of 13 August	per building vectors, Overture footprints	446 flagged of 77,589
Pereira and Dosquebradas	Vantor, from Legion imagery of 12 August	per building vectors, Overture and Google footprints	309 of 35,760 on Overture; 613 of 75,262 on Google
Pereira and Dosquebradas	CEMS EMSR916 AOI02, version 1	graded points	182 points: 76 Destroyed, 85 Damaged, 21 Possibly

			damaged
Pereira and Dosquebradas	Charter 1048 Pléiades Neo affected housing map	published map, vectors to request	143 destroyed, 585 damaged, 36 possibly
Pereira and Dosquebradas	Charter 1048 Sentinel-1 DPM	published map	count not published
Cali	Microsoft AI for Good, from Pléiades imagery of 10 August	per building vectors, Overture footprints	266 flagged of 97,351
Cali	CEMS EMSR916 AOI01 version 2 and AOI03 version 1	graded points	7 points in AOI01, 14 in AOI03
Cali	Charter 1048 Pléiades Neo affected housing map	published map, vectors to request	27 destroyed, 124 damaged
Cali	Charter 1048 Sentinel-1 DPM	published map	count not published
Cali	ICube-SERTIT impact map	vectors on request	count not published
Cali and Pereira (merged CEMS and Microsoft), Quibdó (Copernicus grades only)	IOM-DTM harmonised damage database	dashboard, export on request	count not published
Quibdó	CEMS EMSR916 AOI04, version 1	graded points	74 points: 18 Destroyed, 8 Damaged, 48 Possibly damaged
Manizales	ICube-SERTIT impact map	vectors on request	count not published
Manizales	Charter 1048 Pléiades Neo affected housing map	published map	15 destroyed, 170 damaged, 124 possibly
Armenia, Quimbaya, Sevilla	none	none	none

Table 4.1. Every released product that names individual buildings or points in one of the seven target cities, with the form it is delivered in and its size. Only Pereira and Cali carry more than one product in vector form on the same ground, so they are the only two cities in which the products can be cross-checked against each other before the mission, and Quibdó carries 74 graded buildings that nobody has cross-checked at all. The CEMS products in hand are of mixed versions, AOI01 and AOI06 at version 2 and AOI02, AOI03 and AOI04 at version 1, and a version 2 grading can revise the point counts used as a reference below. Data providers: Microsoft AI for Good, Vantor, Copernicus Emergency Management Service (EMSR916), International Charter Space and Major Disasters (activation 1048), ICube-SERTIT, IOM, Overture Maps.

4.1 Valid areas: the ground each producer actually assessed

A valid-area mask is the polygon a producer publishes beside its damage layer to record where it looked. Inside the mask the producer examined the buildings it could see and returned a verdict on each, damaged or not damaged. Outside the mask the producer returned nothing, and nothing is not the same as not damaged. A building outside a producer's mask is a building that producer never assessed, so it can neither agree nor disagree with any other producer about it. The common valid area of a set of producers is the intersection of their masks: the ground on which every one of them holds an opinion about every building, and therefore the only ground on which their verdicts can be set against each other building by building. Treating the ground outside a mask as ground the producer called undamaged manufactures agreement where there was only silence, and it manufactures a spatial pattern with no damage content in it, which is what section 4.2 measures.

Table 4.2. Valid areas of the optical damage products over the seven target cities, the footprints each mask contains, and the common valid area where more than one producer looked. The share column is the share of the city's Overture footprints the mask contains. The city core is the densest cell of building-footprint centroids, defined per city in the "core_definition" column of the source file, and the reach column is the greatest distance from that core reached by a footprint inside the mask, which is what tells a reader whether a mask covers the city or only its center. Four cities carry no building-level optical product, so for them no valid-area mask and no common valid area exist and no product against product comparison is possible.

Source: catalogue/s1_screen/out_validarea/valid_area_summary.csv. Data providers: Microsoft AI for Good, Vantor, Copernicus Emergency Management Service (EMSR916), Overture Maps.

City	Producer	Valid area (km2)	Footprints assessed	Share of city	Contains core	Reach from core (km)
Pereira and Dosquebradas	Microsoft	254.28	75,033	0.822	yes	10.89
Pereira and Dosquebradas	Vantor	77.14	35,042	0.384	no	10.85
Pereira and Dosquebradas	CEMS AOI02	9.75	17,188	0.188	yes	4.19
Pereira and Dosquebradas	common valid area	5.92	12,023	0.132	no	4.19
Cali	Microsoft	256.03	92,876	0.586	yes	14.70
Cali	CEMS AOI01	18.79	12,003	0.076	no	7.21
Cali	CEMS AOI03	3.74	1,549	0.010	no	4.17
Cali	common valid area	22.52	13,552	0.086	no	7.21
Quibdó	CEMS AOI04	5.53	15,310	0.449	yes	2.25
Quibdó	common valid area, one producer only	5.53	15,310	0.449	yes	2.25
Armenia	none	not applicable	not applicable	not applicable	not applicable	not applicable
Quimbaya	none	not applicable	not applicable	not applicable	not applicable	not applicable
Manizales	none	not applicable	not applicable	not applicable	not applicable	not applicable
Sevilla	none	not applicable	not applicable	not applicable	not applicable	not applicable

The three Pereira masks are mismatched in size and in position. The Microsoft mask contains the city core and reaches 10.89 km from it; the Vantor mask reaches almost as far, 10.85 km, and does not contain the core at all; the CEMS AOI02 mask contains the core and stops at 4.19 km. Their intersection is 5.92 km² holding 12,023 footprints, 13.2 percent of the 91,266 footprints of the city, and it is the central 4.19 km and nothing beyond.

Cali is arranged differently and reaches a similar result. Both Copernicus areas lie wholly inside the Microsoft mask, so the common valid area is their union, 22.52 km² and 13,552 footprints, 8.6 percent of the 158,489 footprints of the city, and it does not contain the city core. Quibdó carries one producer only, CEMS AOI04, 5.53 km² and 15,310 footprints, 44.9 percent of the city and containing the core, so its common valid area is that single mask and no product against product agreement exists there. Armenia with 106,040 footprints, Quimbaya with 7,707, Manizales with 67,014 and Sevilla with 6,726 carry no building-level optical damage product at all, so for those four cities there is no valid-area mask, no common valid area and no product comparison to report, and that absence is stated rather than left blank.

Take-away: the ground on which the three Pereira producers can be compared at all is 5.92 km² holding 13.2 percent of the city's buildings, the ground on which the two Cali producers can be compared is 8.6 percent of that city's buildings and excludes its core, and four of the seven cities carry no optical product to compare (Table 4.2).

4.2 Where the labels come from, and why the correction matters

The label classes used throughout this report and in section 5 are drawn from different geographies, and that is what makes the uncorrected comparison misleading rather than merely wide. Of the 368 positive labels in Pereira, buildings that Microsoft reviewers marked 'ground_truth' or 'confirmed' or that carry a CEMS Destroyed or Damaged point, 159 lie inside the common valid area, 43.2 percent. Of the 358 reviewer-rejected negatives, 37 lie inside it, 10.3 percent. Taking the Microsoft field 'review_status' equal to 'rejected' before the one CEMS override that moves a rejected building into the positive class gives 359 negatives and 38 inside, 10.6 percent, and the asymmetry is the same. A classifier scored on those two classes over all footprints is therefore scored partly on where a building sits rather than on what happened to it.

That contamination is measured and not asserted. A geometry-only null that carries no damage information at all, drawing pseudo-positives from the footprint of whichever producer generated each positive and pseudo-negatives from the Microsoft area, reaches a receiver operating characteristic area under the curve (AUC) of 0.6447. The best model of the attribution run reaches 0.8739 against the same labels, so the geometry-only null covers 38.7 percent of the span from chance to that value. The control that draws both classes from a single common footprint returns 0.5050, which is chance, so the machinery is unbiased and the 0.6447 is footprint mismatch and nothing else.

The correction is expensive and the price is stated wherever it is paid. At building level the Pereira product pair falls from 91,266 footprints to 12,023, and the labelled comparison of section 5 falls from 726 labelled buildings to 196, of which only 37 are negatives. The corrected numbers are noisier than the uncorrected ones and they describe the city center only, out to 4.19 km, so they say nothing about the periphery. Neither limitation makes the uncorrected view the safer one, because the uncorrected view answers a different question, one in which a producer's silence is scored as a verdict.

Three limits bind every corrected number in this section and in section 5, and they are stated once here rather than repeated. The corrected comparison rests on 196 labelled buildings of which only 37 are negatives, so it is noisy and its confidence intervals are wide. The common valid area is the city center and reaches 4.19 km from it, so the corrected numbers describe the core and say nothing about the periphery, which is precisely where section 8 sends the field teams. And the Microsoft reviewer classes and the CEMS grades are both photointerpretation of imagery rather than field observation, so everything reported here is agreement among remote sensing methods and not accuracy against truth; only the mission closes that gap.

Take-away: 43.2 percent of the positive labels and 10.3 percent of the rejected negatives lie inside the common valid area, and a null carrying no damage information at all reaches an AUC of 0.6447 on that asymmetry against 0.5050 on a single common footprint, so any comparison that includes an optical product has to be read inside the intersection.

4.3 Two optical products on the same footprints, Pereira

The agreement statistics in this subsection and the two that follow are first numbers computed on the products as delivered on 5 September 2026, and they change if a CEMS area of interest is reissued at a higher version.

Microsoft and Vantor both published per building damage flags on Overture footprints from Legion imagery acquired one day apart, on 13 August for Microsoft and 12 August for Vantor. On the Overture base Microsoft flags 446 of 77,589 footprints citywide and Vantor 309 of 35,760. Both counts are for the whole delivered layer, that is with no cloud restriction: the Microsoft layer carries an `unknown\_pct` cloud or unknown cover field and no footprint in it reaches 50 percent, so the 50 percent exclusion the framework asks for removes nothing, while the stricter rule of `unknown\_pct` below 1 percent, which is the rule behind the clear footprint count carried in our own inventory, leaves 71,007 footprints and 441 flags in the same file. The Vantor 35,760 is the Overture subset of a delivery that also carries 75,262 Google footprints with 613 flags, so a reader planning against 309 flags is planning against the thinner of the two bases. We compare the two products on the Overture base because it is the only base both of them share.

Both products are compared as binary flags because both are delivered as binary flags: neither Vantor layer carries a severity class, only the binary `damaged` field beside the continuous `damage\_pct` fields. The severity split of 169 severe, 199 moderate and 245 light recorded for this product in our inventory is a property of the 613 flags on the Google base and comes from the product description rather than from either delivered layer, and it is the only severity information anyone holds on it. The Microsoft layer name

records a decision threshold of 0.2 on its continuous score, which is the choice that sets the product's precision and recall and which neither the producer nor this report interrogates.

The valid areas of the two products, after their own cloud masks, overlap on 55 km² out of 254 km² for Microsoft and 77 km² for Vantor, and inside that overlap Microsoft flags 318 of 29,479 footprints and Vantor 300 of 29,529. The contingency table is built on the 29,460 Vantor footprints matched to a Microsoft footprint within 3 m of centroid distance, a tolerance set to absorb the offset between two versions of the same Overture geometry without joining neighboring buildings, and on that base the Microsoft flag count is 325. The match is many to one, so a flagged Microsoft footprint that receives more than one Vantor footprint is counted more than once and the row totals of the contingency table count matched pairs rather than distinct Microsoft footprints; that accounting, and not a change of base, is why the Microsoft total is 325 on the matched base against 318 counted once each inside the overlap. On that base the products flag 197 buildings in common, 128 Microsoft only and 103 Vantor only, giving a Jaccard index of 0.460 and a Cohen kappa of 0.626, and the 50 percent cloud or unknown cover exclusion changes neither value. Cohen kappa is the agreement that survives after removing what two products would reach by chance given how rarely either of them flags a building, so the kappa values of 0.256 and 0.258 that the two products reach against the CEMS grading in Table 4.4 mean that most of the apparent agreement there is what that rarity produces on its own.

The Jaccard index of 0.460 is not the independent agreement. The Microsoft layer carries a provenance field, 'damage_from', set to 'vantor' on 263 footprints, of which 238 also carry a damage flag, which means Microsoft merged Vantor detections into its own reviewed layer. Restricted to the 29,191 footprints where the Microsoft flag was not imported from Vantor, the two products flag 81 and 130 buildings, share 45, and the Jaccard index falls to 0.271. Cohen kappa is not available for that subset, so the like for like comparison is Jaccard 0.460 against 0.271. Two optical models reading near identical imagery agree independently on about a quarter of the buildings either of them flags, which is the same order of product disagreement EEFIT met in Turkey in 2023, where the ARIA and CEMS products correlated at 0.31 (Voelker and others 2024, section IV; see the reference list in section 9). The two missions for which we hold both a pair of released products and a published comparison, Turkey 2023 and Colombia 2026, disagree at building level to the same order, so a field team should expect this and the field pack should carry the provenance fields beside the flags.

Both published values, the Jaccard index of 0.460 with its Cohen kappa of 0.626 and the independent Jaccard index of 0.271, were computed inside the 55 km² Microsoft and Vantor overlap, which is a two-producer overlap and not the common valid area of section 4.1. Recomputing the same two comparisons in the two views of this section places them in context. Inside the common valid area the raw pair reaches a Jaccard index of 0.445 and a Cohen kappa of 0.611, close to the published pair; over all footprints the same comparison falls to 0.353 and 0.520, because a footprint Vantor never assessed is counted as a footprint Vantor declined to flag. The independent subset moves the same way and further, 0.192 inside the common valid area against 0.095 over all footprints, against the published 0.271 inside the two-producer overlap. Section 4 as published in earlier versions therefore overstated the agreement between the two optical products, not understated it: the published Jaccard indices of 0.460 and 0.271 sit above the corrected common-valid-area values of 0.445 and 0.192 rather than below them, and it is the all-footprints view, which no earlier version reported, that understates the agreement further still, to 0.353 and 0.095.

Table 4.3. Agreement between the two Pereira optical products, corrected first and uncorrected second, with the published value and the ground it was computed on carried beside each pair. "Corrected" is the common valid area of Table 4.2, 12,023 footprints on which Microsoft, Vantor and CEMS AOI02 all hold an opinion; "uncorrected" is all 91,266 Overture footprints in the city bounding box, on which a footprint outside a producer's mask is counted as unflagged by that producer. The correction costs 79,243 footprints of the 91,266, that is 86.8 percent of the comparison base, and buys a base on which both producers

actually looked. The flagged columns are the counts of footprints each product flags inside the view; the independent subset is the Microsoft flags whose 'damage_from' field is not 'vantor'. The directional precision and recall of each row are carried in the same source file and are not repeated here, because for two products of equal standing the symmetric Jaccard index and Cohen kappa are the comparable statistics. Source: catalogue/s1_screen/out_validarea/product_agreement_corrected.csv. Data providers: Microsoft AI for Good, Vantor, Copernicus Emergency Management Service (EMSR916), Overture Maps.

Comparison	View	Footprints	Microsoft flags	Vantor flags	Both	Jaccard	Kappa	Published value and its scope
Microsoft against Vantor	corrected, common valid area	12,023	164	151	97	0.445	0.611	0.460 and 0.626, inside the 55 km2 Microsoft and Vantor overlap
Microsoft against Vantor	uncorrected, all footprints	91,266	446	309	197	0.353	0.520	0.460 and 0.626, inside the 55 km2 Microsoft and Vantor overlap
Microsoft independent flags against Vantor	corrected, common valid area	12,023	54	151	33	0.192	0.317	0.271, Microsoft flags with 'damage_from' not 'vantor'
Microsoft independent flags against Vantor	uncorrected, all footprints	91,266	208	309	45	0.095	0.172	0.271, Microsoft flags with 'damage_from' not 'vantor'

Take-away: on the ground where both optical producers looked, they agree independently on about a fifth of the buildings either of them flags, a Jaccard index of 0.192, and the uncorrected version of the same comparison halves that to 0.095 by scoring Vantor silence as a Vantor negative.

4.4 The optical products against the CEMS grading, Pereira AOI02

CEMS graded 182 points inside AOI02, 9.7 km2 of Pereira: 76 Destroyed, 85 Damaged and 21 Possibly damaged. Of these, 160 fall within 15 m of a Microsoft footprint and 102 within 15 m of a Vantor footprint, so Vantor is the thinner sampling frame; the 15 m tolerance is the framework's own rule for matching a graded point to a footprint of a different base, chosen to absorb the offset between a photointerpreted point and a footprint centroid. Several graded points fall on the same building, so those 160 points resolve to 134 distinct Microsoft footprints and those 102 points to 83 distinct Vantor footprints, which are the two reference set sizes behind the true positive and false negative columns of Table 4.4. In that table TP is true positives, buildings both the product and CEMS mark, FP is false positives, buildings the product marks and CEMS did not, and FN is false negatives, buildings CEMS graded and the product did not mark.

Reference set	Product	TP	FP	FN	Precision	Recall	Kappa
CEMS any class, AOI02	Microsoft	51	202	83	0.20	0.38	0.256
CEMS any class, AOI02	Vantor	31	120	52	0.21	0.37	0.258
CEMS Destroyed only, AOI02	Microsoft	32	221	22	0.13	0.59	0.204

Table 4.4. Both released optical products are poor in precision and good in recall on collapses: they find most of what CEMS graded as Destroyed and mark about four buildings for every one CEMS also marked, so a field team following them without controls will spend most of its time on buildings no other product

calls damaged. Source: catalogue/agreement/agreement_log.txt, block "CEMS (any class) as reference inside AOI02". Data providers: Microsoft AI for Good, Vantor, Copernicus Emergency Management Service (EMSR916), Overture Maps.

Table 4.4 is computed inside CEMS AOI02, which is one producer's mask and not the common valid area, so the same three comparisons are recomputed in both views of this section. Restricting to the common valid area raises most of these comparisons over the published, inside-AOI02 view, but not all of them: the CEMS any class comparison against Vantor falls slightly on correction, from precision 0.210, recall 0.370 and kappa 0.258 as published to precision 0.205, recall 0.369 and kappa 0.257 inside the common valid area. Nor is the uncorrected, all-footprints view the lowest of the three throughout: its recall of 0.593 on CEMS Destroyed against Microsoft exceeds the published recall of 0.590 for the same pair, even though the corrected recall of 0.719 is higher than both. What holds without exception is that scoring the ground outside a producer's mask as a negative moves precision, recall and kappa by a few points in either direction and never reverses which product finds more collapses. Table 4.5 reports the pairs.

Table 4.5. The optical products against the CEMS grading in three views, the published one first and then corrected and uncorrected. "As published" is inside CEMS AOI02, 17,194 Microsoft footprints and 12,048 Vantor footprints; "corrected" is the 12,023 footprints of the common valid area of Table 4.2; "uncorrected" is all 91,266 city footprints. Precision is the share of product flags that carry a CEMS point and recall is the share of CEMS-pointed buildings the product flags, both computed at footprint level with a CEMS point joined to a footprint within 15 m. The corrected and uncorrected rows are computed on the same footprint join and are comparable with each other; the published row uses the point to distinct-footprint reference sets of Table 4.4 and is carried for continuity rather than as a fourth measurement, and no Jaccard index was published for those rows. The published kappa of 0.204 for CEMS Destroyed only against Microsoft is the value of Table 4.4 and is not held in the published_value column of the corrected and uncorrected source file, which records published precision and recall for that pair but not kappa. Source: catalogue/s1_screen/out_validarea/product_agreement_corrected.csv. Data providers: Microsoft AI for Good, Vantor, Copernicus Emergency Management Service (EMSR916), Overture Maps.

Reference set	Product	View	Footprints	Precision	Recall	Kappa	Jaccard
CEMS any class	Microsoft	as published, inside AOI02	17,194	0.200	0.380	0.256	not published
CEMS any class	Microsoft	corrected, common valid area	12,023	0.232	0.452	0.300	0.181
CEMS any class	Microsoft	uncorrected, all footprints	91,266	0.114	0.378	0.174	0.096
CEMS any class	Vantor	as published, inside AOI02	12,048	0.210	0.370	0.258	not published
CEMS any class	Vantor	corrected, common valid area	12,023	0.205	0.369	0.257	0.152
CEMS any class	Vantor	uncorrected, all footprints	91,266	0.100	0.230	0.138	0.075
CEMS Destroyed only	Microsoft	as published, inside AOI02	17,194	0.130	0.590	0.204	not published
CEMS Destroyed only	Microsoft	corrected, common valid area	12,023	0.140	0.719	0.231	0.133
CEMS Destroyed only	Microsoft	uncorrected, all footprints	91,266	0.072	0.593	0.127	0.068
CEMS Destroyed	Vantor	corrected, common	12,023	0.146	0.688	0.237	0.137

only		valid area					
CEMS Destroyed only	Vantor	uncorrected, all footprints	91,266	0.071	0.407	0.120	0.065

The corrected recall on collapses is the number a field team should plan against: inside the common valid area Microsoft flags 23 of the 32 buildings carrying a CEMS Destroyed point, a recall of 0.719, and Vantor 22 of the same 32, a recall of 0.688. Precision stays low in every view, 0.140 and 0.146 on that same class, so the four-to-one ratio of unconfirmed flags to confirmed ones survives the correction.

The class conditional table explains the pattern. Of the 76 Destroyed points, 71 have a Microsoft footprint and 44 of those are flagged; of the 85 Damaged points, 70 have a footprint and 20 are flagged; of the 21 Possibly damaged, 19 have a footprint and four are flagged. Vantor behaves the same way on the points it can see, flagging 34 of 44 matched Destroyed points and six of 44 matched Damaged points. Both products find most collapses and miss most damaged roofs, and for every building both Microsoft and CEMS mark, Microsoft marks about four more than CEMS did not.

One property of the reference bears on every row of Table 4.4. CEMS grading is bi-temporal photointerpretation, and for AOI02 the pre-event reference is ESRI World Imagery of 12 May 2025 against a post-event Pléiades scene of 11 August 2026, so the reference records everything that changed on a roof in more than a year, while Microsoft and Vantor infer damage from a single post-event scene. The Microsoft precision of 0.20 in Table 4.4, measured against such a reference, is a statement about two different measurement designs as much as about the products, and the mission is the only way to separate them.

The human review labels in the Microsoft product converge with CEMS on the same collapses. The Microsoft reviewer class named 'ground_truth' is a desk review of imagery, not a field observation, and we call it reviewer-confirmed (MS ground_truth) hereafter; 'confirmed', 'unsure' and 'rejected' are the reviewers' verdicts on machine flags. Citywide the counts are 367 rejected, 156 confirmed, 154 unsure and 136 reviewer-confirmed, and of the 76 CEMS Destroyed points 35 sit on a reviewer-confirmed building and six on a 'confirmed' one, against one on a 'rejected' building. Nothing in either product has been checked on the ground, so the reviewer classes are survey strata and not a validation set.

Take-away: inside the common valid area both optical products find about seven of every ten buildings CEMS graded as Destroyed and about two of every ten of their own flags carry a CEMS point of any grade, and the uncorrected version of the same comparison halves both precision figures by counting CEMS silence outside AOI02 as a CEMS negative.

4.5 The same comparison at 250 m

On a 250 m grid restricted to cells with at least five footprints, a floor chosen so that a per cell rate is not driven by one or two buildings, AOI02 contains 169 cells, of which 68 carry a Microsoft flag, 56 a Vantor flag and 54 a CEMS point; the Spearman correlation of flagged counts per cell is 0.69 between Microsoft and Vantor, 0.52 between Microsoft and CEMS and 0.36 between Vantor and CEMS, and the Jaccard index of the top decile of cells is 0.44, 0.52 and 0.30 for the same three pairs. Fig. 4 places the building level pair of 4.3 beside these grid level statistics.

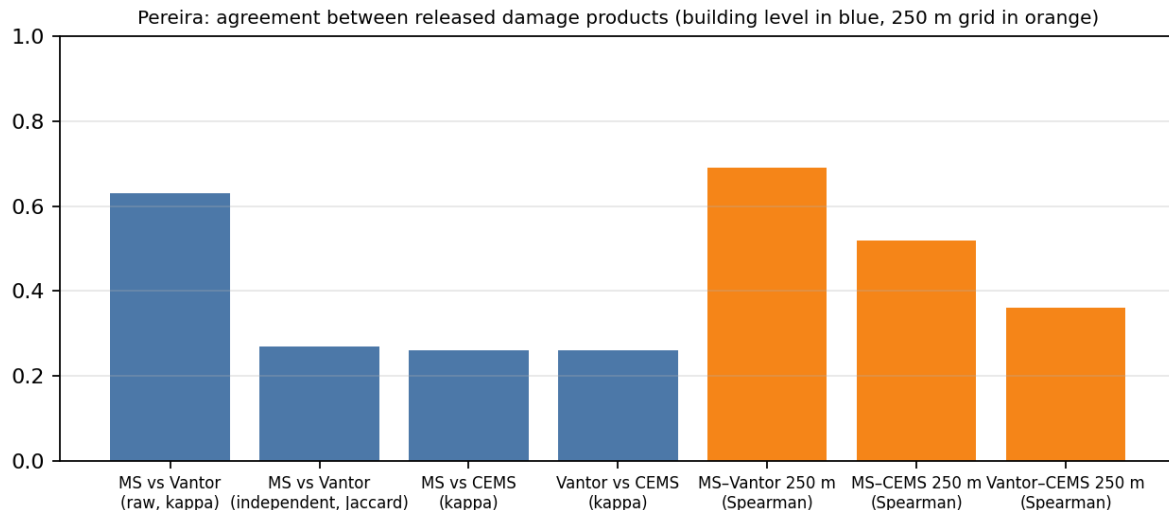


Fig. 4. Microsoft, Vantor and CEMS agree at 250 m in Pereira, with Spearman correlations from 0.36 to 0.69, and disagree at building scale, where the like for like pair is a Jaccard index of 0.460 raw against 0.271 once the Microsoft flags imported from Vantor are removed. The two scales are drawn in separate panels, each with its own labelled axis, because a Jaccard index, a Cohen kappa and a Spearman correlation are not commensurable. Data providers: Microsoft AI for Good, Vantor, Copernicus Emergency Management Service (EMSR916) and Overture Maps. Source: catalogue/agreement/agreement_log.txt.

The grid comparison is computed inside CEMS AOI02, which is one producer's mask, and 12,023 of the 17,188 footprints inside that mask also lie inside the common valid area, so 70 percent of its cells rest on ground all three producers assessed and the rest rest on ground Vantor did not. The Vantor column of the grid therefore carries the same silence-as-negative defect as the building-level comparison over the part of AOI02 outside the Vantor mask, and the Microsoft against CEMS correlation of 0.52 does not, because both of those producers assessed all of AOI02.

Take-away: the products point at the same neighborhoods and disagree about which building inside them, so a ranking carried out at 250 m is stable while the same ranking carried out per building is not, and the Vantor column of that grid is the one to read inside the common valid area alone.

4.6 Cali, where the released products flag none of the graded points

Cali is the counterexample. Microsoft covers 97,351 Overture footprints there and flags 266 of them; our own inventory records 97,085 footprints for the same product, a base count we have not reconciled, while the flag count is 266 in both. That 266 is the count in the released Microsoft layer. Sections 7, 8 and 11 report 256, which is the number of those flags that join the 158,489 footprint Cali building layer the field team carries; the other 10 fall outside it, and every statement made on the delivered layer is therefore made on 256. CEMS graded 21 points across the two Cali AOIs: seven in AOI01, covering 18.8 km² of northern Cali, where Microsoft flagged 25 of 12,007 footprints, and 14 in AOI03, covering 3.7 km² of the city center, where Microsoft flagged two of 1,413 footprints. Six points in each AOI fall within 15 m of a footprint, and in both AOIs the number of CEMS points Microsoft also flags is zero, giving kappa values of -0.001 and -0.002.

Cali gets the same valid-area treatment as Pereira and the zero survives it. The Microsoft mask covers 256.03 km² holding 92,876 footprints and contains the city core; CEMS AOI01 covers 18.79 km² holding 12,003 footprints and CEMS AOI03 covers 3.74 km² holding 1,549, and both Copernicus areas lie wholly inside the Microsoft mask, so the common valid area is their union, 22.52 km² and 13,552 footprints, 8.6

percent of the 158,489 footprints of the city, and it does not contain the core. Inside that common valid area Microsoft flags 27 footprints and CEMS marks 11 of any grade, and the two sets do not intersect, a Jaccard index of 0.000 and a Cohen kappa of -0.0012; over all 158,489 footprints, where Microsoft flags 256 and CEMS still marks 11, the intersection is again empty. The Cali zero is therefore a property of the two products and not of the ground either of them declined to look at, which is what distinguishes it from the Pereira agreement figures. One further caution attaches to it: the Microsoft Cali layer is an unreviewed prediction layer, with no reviewer verdicts of the kind the Pereira layer carries, so Cali supports no positive against reviewer-rejected label set and no technique ranking of the kind section 5 reports for Pereira.

The two AOIs reach that zero by different routes, and the difference decides what the field team does in each. In AOI01 six of the seven graded points sit on a Microsoft footprint and none is flagged, a result recorded as TP 0, FP 25, FN 5; the six points resolve to five distinct footprints, because two of them fall on the same building, which is why the false negative count is one lower than the matched point count. Microsoft holds those buildings and does not call them damaged, which is disagreement, and the response is to sample them. In AOI03 eight of the 14 graded points have no Overture footprint at all and only six match, a result recorded as TP 0, FP 2, FN 6. That is non intersection of two thin data sets rather than disagreement, and the response is a field sketch of buildings the base map does not carry. Across the two AOIs 12 of the 21 graded points fall within 15 m of a Microsoft footprint and the other 9 fall on none; those 12 points resolve to 11 distinct footprints, and the released product flags none of the 11. The count of 12 is points matched to the Microsoft Cali layer within 15 m; section 8 reports 10, which is the smaller number of the same 21 points that carry an Overture identifier in the delivered building layer the field team takes into the field, and the two joins are not interchangeable.

The acquisition timing is not the explanation. The Microsoft Cali product was inferred from Pléiades imagery acquired at 15:37 UTC on 10 August, 3 hours 2 minutes and 32 seconds after the origin time of 12:34:28 UTC, and the CEMS post-event scene for both Cali AOIs is Pléiades of 11 August at 15:31 UTC, 23 hours and 54 minutes later. The confound that matters is the design of the two references. CEMS grades bi-temporally, against a pre-event Pléiades scene of 28 July 2026 in AOI01 and against ESRI World Imagery of 24 March 2025 in AOI03, more than a year before the event, while Microsoft infers from a single post-event scene. A change detection reference set against single scene detectors explains the Cali zero, the Microsoft precision of 0.20 in Table 4.4 and the low recall on the Damaged class better than any timing argument does. On top of that, the damage reported in Cali is dominated by mid rise structural failures, 56 collapses on 11 August, which nadir optical products under see and which neither released layer carries information about.

No released product is a usable sampling frame in Cali today. The one that could become one is the Charter activation 1048 Pléiades Neo affected housing map, 27 destroyed and 124 damaged, which exists as a published map and has to be requested as vectors from the Colombian value adding partner; section 8 sends the team to those vectors, to the CEMS graded points and to a layer we generate ourselves.

Take-away: on the 8.6 percent of Cali footprints where Microsoft and CEMS both looked, the two products share no flagged building at all, and the correction changes neither the zero nor the conclusion that no released product is a usable sampling frame in that city.

4.7 What the field pack carries

Four GeoPackages in 'catalogue/agreement/' hold the harmonized comparison and are the files to load on the tablet. 'pereira_aoi02_footprints_products.gpkg' carries 17,194 footprints inside AOI02 with 253 Microsoft flags, 134 footprints holding a CEMS point of any grade and 54 holding a CEMS Destroyed point, 118 flags imported from Vantor, and reviewer classes of 127 reviewer-confirmed, 73 'confirmed', 53 'unsure' and 58 'rejected'. The 'cems_destroyed' column is the one that separates the class both products detect well from the class they miss, so it is the column the strata of section 8 are cut on.

`pereira\_grid250\_agreement.gpkg` carries the 169 grid cells with the per cell footprint count `n`, which holds the five footprint filter and ranges from 5 to 598, and the per cell flag counts `ms`, `va` and `ce`; `va` is stored as a real but holds only whole values from 0 to 21, so it is a plain count and the Spearman correlation of 0.69 in 4.5 is a correlation of counts. `cali\_AOI01\_footprints\_products.gpkg` and `cali\_AOI03\_footprints\_products.gpkg` carry 12,007 and 1,413 footprints with `ms\_d`, `cems` and `unknown\_pct`.

The Pereira pack is built on the Microsoft footprint base and carries no Vantor column, and the Microsoft only, Vantor only and shared strata that section 8 sends the team to survey in full cannot be located on the tablet without one. We add a `va\_d` column to `pereira\_aoi02\_footprints\_products.gpkg` before the pack is frozen. Its coverage is bounded: inside AOI02 Microsoft holds 17,194 footprints against Vantor's 12,048, so at most 12,048 of the pack's rows can carry a real Vantor value and 5,146 cannot. The column therefore has to distinguish 0, meaning inside the Vantor base and not flagged, from null, meaning outside the Vantor base, and the section 8 strata are defined on the rows where `va\_d` is not null, because a field engineer reading an empty cell cannot otherwise tell a negative from an absence. No Vantor severity column is added, because the delivered Vantor layers carry none.

The same argument applies to the valid areas themselves, and it is the reason the three Pereira masks and the four Cali and Quibdó masks travel with the pack. `MS\_Pereira\_valid\_area\_mask.geojson`, `Vantor\_Pereira\_valid\_area\_mask.geojson` and the CEMS AOI02 area-of-interest layer are the polygons that define the 254.28, 77.14 and 9.75 km² of Table 4.2, and `MS\_Cali\_valid\_area\_mask.geojson` with the AOI01 and AOI03 areas define the Cali equivalents. Three boolean columns are added to the Pereira pack, one per producer, plus a fourth marking the 12,023 rows of the common valid area, so a field engineer standing in front of a building can tell which producers had an opinion about it before reading their verdicts. A building outside a producer's mask is not evidence that the producer disagreed with the others about it.

The field pack is loadable on a tablet only if every column tells the field team what an empty cell means, which is why the coverage bound, the null convention and the four valid-area columns travel with it.

Take-away: the Pereira field pack carries a `va\_d` column bounded to the 12,048 footprints inside the Vantor mask, a null convention that separates 0 from missing, and four new boolean valid-area columns, so a field engineer can tell which producers had an opinion about a building before reading their verdicts.

4.8 What to carry into the mission products

Intersect the producers' valid areas before computing any agreement statistic, and report the uncorrected version beside it as the demonstration of what the correction costs. Report agreement at both scales in every mission product, building level with Cohen kappa and the Jaccard index and 250 m with the Spearman correlation and the hotspot Jaccard index, and carry the provenance fields into every harmonized layer. For Pereira and Dosquebradas, publish those four statistics for `catalogue/agreement/pereira\_aoi02\_footprints\_products.gpkg` and `catalogue/agreement/pereira\_grid250\_agreement.gpkg` in the nightly field pack, with `damage\_from`, `review\_status`, the new `va\_d` and the four valid-area columns retained in the first file. Two numbers give the reason. Without `damage\_from` the Pereira building level agreement would have been reported as a Jaccard index of 0.460 instead of the independent 0.271. And without the valid-area intersection the same independent comparison reads 0.095 rather than 0.192, a factor of two in the opposite direction.

One distinction decides which of these rules binds which product. The radar screens of section 5, Sentinel-1, NISAR, OPERA RTC and the Balz coherence products, carry no valid-area problem: each of them covers the whole city and returns a value for every footprint it can resolve, so a radar screen can be compared against field observations anywhere the field team goes. Only the optical products are masked.

Any comparison that involves an optical product, including the use of optical labels to score a radar screen, is valid only inside the common valid area of the optical producers concerned. That is why section 5 reports its technique ranking twice and why the field itinerary of section 8 is not confined to the 5.92 km² of Table 4.2.

Take-away: every agreement statistic in this report is computed inside the common valid area and reported beside its uncorrected counterpart, and the two differ by a factor of about two on the independent Pereira comparison.

Sources: `AGREEMENT\_FRAMEWORK.md`, `COLOMBIA\_RECON\_CATALOGUE.md`,
`LESSONS\_LEARNED.md`, `PROCESSING\_LOG.md`, `catalogue/agreement`,
`catalogue/agreement/agreement\_log.txt`, `catalogue/agreement/cali\_AOI01\_footprints\_products.gpkg`,
`catalogue/agreement/pereira\_aoi02\_footprints\_products.gpkg`,
`catalogue/agreement/pereira\_grid250\_agreement.gpkg`, `catalogue/damage\_products`,
`catalogue/damage\_products/CEMS\_AOI01\_NorthernCali\_v2/EMSR916\_AOI01\_GRA\_PRODUCT\_source\_v1.dbf`,
`catalogue/damage\_products/CEMS\_AOI02\_Pereira\_v1/EMSR916\_AOI02\_GRA\_PRODUCT\_source\_v1.dbf`,
`catalogue/damage\_products/CEMS\_AOI04\_Quibdo\_v1/EMSR916\_AOI04\_GRA\_PRODUCT\_builtUpP\_v1.shp`, `catalogue/damage\_products/MS\_Cali\_overture\_footprints\_predictions.gpkg`,
`catalogue/damage\_products/MS\_Pereira\_overture\_reviewed.gpkg`,
`catalogue/damage\_products/Vantor\_Pereira\_google\_predictions.gpkg`,
`catalogue/damage\_products/Vantor\_Pereira\_overture\_predictions.gpkg`,
`catalogue/s1\_screen/out20/final/build\_final\_log.json`,
`catalogue/s1\_screen/out\_ablation/ATTRIBUTION.md`,
`catalogue/s1\_screen/out\_ablation/attr\_geometry\_only\_null.csv`,
`catalogue/s1\_screen/out\_ablation/attr\_geometry\_only\_null\_singlefootprint.csv`,
`catalogue/s1\_screen/out\_validarea/label\_geography.csv`,
`catalogue/s1\_screen/out\_validarea/product\_agreement\_corrected.csv`,
`catalogue/s1\_screen/out\_validarea/reproduction\_check.csv`,
`catalogue/s1\_screen/out\_validarea/technique\_ranking.csv`,
`catalogue/s1\_screen/out\_validarea/valid\_area\_summary.csv`,
`figures/fig5\_product\_agreement\_pereira.png`.

5. Synthetic aperture radar coherence screens: NISAR, Sentinel-1 and OPERA RTC

We screen three synthetic aperture radar (SAR) products: the coherence of the NASA-ISRO Synthetic Aperture Radar (NISAR) L-band interferograms, Sentinel-1 C-band 6-day coherence, and the backscatter change of the radiometric terrain corrected (RTC) Sentinel-1 product (RTC-S1) of the Observational Products for End-Users from Remote Sensing Analysis (OPERA) project. Coherence is the similarity between two radar images of the same ground. A building that collapsed or lost its roof between the two acquisitions loses coherence, and so do wet ground, vegetation, debris clearance and traffic. For every pair we averaged coherence inside each footprint of the Overture Maps Foundation open building-footprint dataset (Overture) and formed the coherence drop, pre-event minus co-event coherence, which is the damage proxy map (DPM) of section 3 and is carried in the product files as the column `dpm`.

Three provisional labels apply throughout: every NISAR product is a provisional version 1 release; every result below rests on one co-event pair per sensor; and all 20 m Sentinel-1 products carry a geolocation offset of about 40 m toward the sensor. The offset is two pixels of the 20 m grid and is smaller than the

140 m coherence window, so city and neighbourhood statistics are unaffected, while per-building work at 10 m requires the corrected digital elevation model (DEM). Section 5.6 states what the offset means for the intermediate case this section reports, per-building values read off a 20 m product.

What the screens can and cannot see follows from the geometry rather than from a damage taxonomy. A coherence value is the mean of a window far larger than a building: 19,600 square metres for Sentinel-1 and about 750 square metres for NISAR, against a median Pereira footprint of 61 square metres. Damage that leaves the scattering geometry of such a patch intact does not appear in either product, and 75 percent of Pereira footprints receive no 20 m pixel centre and therefore no value at all. Which damage states a block-scale decorrelation map cannot represent is a question for the mission leads, and section 8 records their answer in their own voice.

5.1 NISAR L-band coherence (provisional)

We used the provisional geocoded unwrapped interferogram (GUNW) products of tracks T011 descending, T047 ascending and T148 ascending, distributed by the Alaska Satellite Facility (ASF). Their coherence layer is unfiltered, at 5 range by 6 azimuth looks and 20 m posting. The 20 m figure is the posting, not the support: the ground cell of that multilook is 25 by 30 m, about 750 square metres. A NISAR value therefore summarises about 12 times the median Pereira footprint of 61 square metres.

Pre-event and post-event pairs span 12 days; every co-event pair spans 24 days because the satellite did not acquire anywhere over the area between the last pre-event pass on 25 July and the first post-event pass on 11 August, the gap reported in section 3.3. Inside the built-up mask the 12-day pre-event median runs from 0.55 in Quimbaya on T011D to 0.82 in Cali on T148A, the 24-day co-event median from 0.47 to 0.74 on the same two rows (Fig. 5 for the per-phase city levels; the co-event pairs themselves are mapped in section 3); the per-footprint medians of Table 5.2 are computed on a different support and differ by about 0.01. The longer interval is not what produces the drop: a 24-day pre-event baseline built from the geocoded single-look complex (GSLC) products of 26 June and 20 July with the same estimator changes the building median by 0.01 or less in every city, and the co-event pair sits 0.06 to 0.10 below both baselines.

No single NISAR frame covers all seven cities, so Table 5.1 states which track carries data where before Table 5.2 reports values.

Table 5.1. NISAR track availability across the seven cities, and the track used in Table 5.2. "No data" means the city and track combination exists in the product inventory but returns no valid coherence over any footprint. Source: catalogue/s1_screen/out_nisar/nisar_building_dpm_stats.csv.

City	T047A f004	T011D f086	T011D f087	T148A f003	Track used
Quibdó	no data	data	not in frame	not in frame	T011D f086
Armenia	data	data	not in frame	not in frame	T047A f004
Pereira and Dosquebradas	data	data	not in frame	not in frame	T047A f004
Manizales	data	data	not in frame	not in frame	T047A f004
Quimbaya	data	data	not in frame	not in frame	T047A f004
Sevilla	data	data	no data	no data	T047A f004
Cali	no data	no data	data, excluded as artefact	data	T148A f003

The selection rule follows from that matrix and from one documented exclusion: T047A f004 wherever it carries data, T011D f086 where T047A does not (Quibdó), and T148A f003 for Cali, where T047A f004 and T011D f086 are empty and T011D f087 is an artefact. We call this the preferred track.

Table 5.2. NISAR per-building coherence and drop on the preferred track per city, provisional products. Pre and co-event are median coherence inside footprints, dimensionless on 0 to 1. The DPM column is the median of the per-building drops, not the difference of the two medians, so it does not reproduce by

subtracting the two preceding columns. The share column is the share of buildings whose drop exceeds 0.2, a threshold at which half the Copernicus Emergency Management Service (CEMS) Destroyed buildings sit at a 24-day L-band baseline and which is in the range reported for C-band damage proxy maps in Turkey. The italicised last row is an excluded artefact, not a result. Source: catalogue/s1_screen/out_nisar/nisar_building_dpm_stats.csv.

City	Track	Buildings	Pre	Co-event	DPM	Share above 0.2
Quibdó	T011D f086	7,199	0.70	0.61	0.060	0.154
Armenia	T047A f004	24,450	0.75	0.69	0.052	0.079
Pereira and Dosquebradas	T047A f004	22,696	0.72	0.62	0.072	0.153
Manizales	T047A f004	16,569	0.67	0.63	0.031	0.046
Quimbaya	T047A f004	2,196	0.67	0.56	0.089	0.168
Sevilla	T047A f004	1,888	0.75	0.66	0.071	0.108
Cali	T148A f003	48,079	0.81	0.73	0.072	0.095
<i>Cali, excluded artefact</i>	<i>T011D f087</i>	<i>48,079</i>	<i>0.78</i>	<i>0.49</i>	<i>0.255</i>	<i>0.640</i>

We exclude a frame when two conditions hold together: the whole frame decorrelates over the city in the co-event pair, and it does not recover in the post-event pair. Cali on T011D f087 meets both, dropping from 0.79 pre-event to 0.53 co-event and staying at 0.50 afterwards, whereas T148A over the same city and nearly the same dates gives 0.82, 0.74 and 0.77 across the three phases. Fig. 7 puts the two Cali tracks side by side. These four exclusion figures are built-up mask medians, which is why they sit 0.01 to 0.04 above the per-footprint medians the same two frames carry in Table 5.2, 0.78 and 0.49 on T011D f087 and 0.81 and 0.73 on T148A, the largest gap being the T011D f087 co-event value, 0.53 against 0.49.

Applied openly, the criterion also touches Quibdó. On T011D f086 the Quibdó medians run 0.70 pre-event, 0.62 co-event and 0.45 post-event, the lowest post-event value in the file, so Quibdó fails the recovery condition. It does not fail the first condition: its co-event decorrelation is 0.078 against 0.264 for the excluded Cali frame. Persistent rainfall and dense vegetation over the Chocó are the expected cause of a low 12-day post-event pair there, against a frame-wide problem in the Cali case, and T011D f086 is the only NISAR track that carries data over Quibdó (Table 5.1), so we keep the Quibdó row and flag it rather than excluding it.

The ranking depends on which track is read. Table 5.2 mixes three tracks. On the one track that covers six of the seven cities, T011D f086, the drop order is Sevilla 0.076, Pereira and Dosquebradas 0.071, Armenia 0.062, Quibdó 0.060, Quimbaya 0.058, Manizales 0.040. On T047A f004 the same city, Quimbaya, moves from second lowest to highest, 0.058 to 0.089, and Sevilla from highest to mid, 0.076 to 0.071. Track choice therefore moves the L-band ranking by as much as the sensor comparison in section 5.3 does. Fig. 6 shows why the ranking is fragile: in every city the drop is spread across the whole urban area rather than concentrated in identifiable blocks, so a city median is an average over a diffuse field.

The product whose skill survives our verification tests is neither of these two single-track rankings but the mean of the two tracks, which scores a receiver operating characteristic area under the curve (AUC) of 0.805 against reviewer-rejected negatives and 0.832 against unflagged negatives. Both of those values are computed over all Pereira footprints and therefore carry the valid-area defect section 5.5 corrects; the two tracks scored separately in both views are in Table 5.5. Section 8 names the two-track mean as the product to calibrate.

Take-away: the NISAR city drops are small, 0.03 to 0.09 (Table 5.2), they rest on one co-event pair of a provisional product, and the city order is not a property of L-band alone because it changes between two L-band tracks.

5.2 Sentinel-1 C-band 6-day coherence

We processed 13 bursts of track 142 across three Interferometric Wide (IW) subswaths at seven acquisitions spaced 6 days apart from 21 July to 26 August, giving three pre-event pairs, one co-event pair (8 to 14 August) and two post-event pairs, with a 7 by 7 window geocoded to 20 m. The window is a square 140 m by 140 m patch of ground, about 19,600 square metres, so a Sentinel-1 value summarises about 321 times the median Pereira footprint. We kept 7 by 7 because it is the smallest window that delivers a usable look count at this posting: the observed coherence distribution implies 25 to 27 effective looks against the nominal 49, because bicubic geocoding correlates neighbouring pixels and the 22 m azimuth resolution is coarser than the 20 m posting. The consequence for the reader is that the coherence bias at low coherence is larger than 49 looks would predict, so the fixed 0.2 and 0.15 thresholds used below are conservative reporting conventions and not calibrated detection thresholds.

A geolocation regression test places 98.4 to 100 percent of valid pixels inside the ASF burst footprint, with 12 bursts passing and one warning. That test asserts an overlap above 80 percent, so it detects a mislocated swath and is blind to the two-pixel geoid offset described above; the two statements measure different things. The single warning is not a geolocation failure: burst 303901 in subswath IW2 was flagged because only 36 valid building pixels fall inside it, below the 200-pixel minimum of the urban check.

The 21 to 27 July pre-event pair is the lowest of the six pairs in 11 of the 13 bursts and within 0.001 of the lowest in the other two. We therefore report the drop against the two pre-event pairs nearest the event as the headline, the near-baseline drop (`dpm_late``), and the drop against all three as a lower bound, the full-baseline drop (`dpm``).

Table 5.3. Sentinel-1 6-day coherence and drop per city, with the NISAR drop over the buildings valid in both sensors. Pre (late) and co-event are median coherence inside footprints, dimensionless on 0 to 1; the DPM columns are medians of per-building drops and do not reproduce by subtracting the two preceding columns; the share column is the share of buildings whose near-baseline drop exceeds 0.2. The shared-building column is the number of Table 5.3 buildings that also carry a NISAR value, which is why the NISAR DPM column differs from the same city and track in Table 5.2. For Cali that count is 26,385 here and 26,381 in `fig11_values.csv`, column `n_both``, which is the denominator of the sensor ratio quoted in 5.2; the four building difference does not move any figure quoted from either file. In Cali both bursts show azimuth banding with a 2.1 to 2.2 km period in the 21 to 27 July pair and in both post-event pairs, faint in the co-event pair; the near-baseline drop avoids the two most affected pairs, and no Cali post-event or recovery layer is used anywhere in this report. Source: `catalogue/s1_screen/out20/final/s1_city_table.csv`.

City	Buildings	Pre (late)	Co-event	DPM late	Share above 0.2	Shared with NISAR	NISAR track	NISAR DPM
Quibdó	8,660	0.623	0.408	0.156	0.378	1,848	T011D f086	0.056
Armenia	24,406	0.679	0.525	0.121	0.257	7,947	T047A f004	0.057
Pereira and Dosquebradas	22,731	0.570	0.435	0.106	0.256	9,948	T047A f004	0.068
Manizales	16,708	0.688	0.574	0.096	0.184	6,049	T047A f004	0.034
Quimbaya	2,122	0.395	0.310	0.105	0.280	953	T047A f004	0.091
Sevilla	2,165	0.736	0.478	0.175	0.425	762	T047A f004	0.070
Cali	48,126	0.766	0.614	0.125	0.236	26,385	T148A f003	0.071

Coverage is uneven and is not a property of the city. Sentinel-1 bursts cover 100 percent of the Pereira and Quimbaya bounding boxes but only 15 to 57 percent of Quibdó across four bursts, 43 and 71 percent of Armenia, 43 and 74 percent of Manizales, 51 and 76 percent of Sevilla, and 46 and 65 percent of Cali.

Within the covered area, only footprints containing a 20 m pixel centre receive a value: 8,660 of 34,066 in Quibdó and 22,731 of 91,266 in Pereira and Dosquebradas. Those that do are the larger buildings, median area 183 square metres against 48 square metres for those that do not, so every per-building Sentinel-1 statement concerns the larger buildings. The two coverage maps ``out20/qa2/Quibdo_burst_coverage_map.png`` and ``out20/qa/pereira_burst_coverage_map.png`` show the footprints of record.

The city ranking has to be read against the pre-event level, because a coherence drop is bounded above by the coherence it starts from. Table 5.3 shows the constraint directly: pre-event medians run from 0.395 in Quimbaya to 0.766 in Cali while the near-baseline drop runs from 0.096 to 0.175, and across the seven cities the drop ranks with the starting level at a Spearman correlation of 0.43. Dividing the drop by the pre-event level reorders the cities: Quimbaya 0.267, Quibdó 0.250, Sevilla 0.238, Pereira and Dosquebradas 0.186, Armenia 0.178, Cali 0.163, Manizales 0.139, a Spearman correlation of 0.29 with the raw order. Fig. 8 shows the level the drop starts from, city by city, and Fig. 9 shows the drop itself at building and 250 m cell level. Sevilla falls from first to third and Quimbaya rises from sixth to first. Manizales is last on both. The raw order therefore does not survive normalisation, and section 8 inherits both orders rather than one.

The two sensors do not rank the cities the same way, and part of that is track choice rather than wavelength. On the buildings valid in both, the 6-day C-band drop is 1.3 times the 24-day L-band drop in Quimbaya and 3.0 times it in Manizales. Against the preferred-track NISAR order of Table 5.2 the Sentinel-1 order is uncorrelated across the seven cities, Spearman 0.00. Fig. 10 draws the two sensors on the shared buildings. Against the single-track T011D f086 order over the six cities that track covers, the two agree at Spearman 0.77, with the same extremes, Sevilla highest and Manizales lowest, and disagreement in the middle where Quibdó moves from second to fourth and Pereira from fourth to second.

At 250 m cell level the two sensors agree weakly and flag very different amounts of area. In cells holding at least five buildings in both sensors, the Spearman correlation of cell means runs from 0.003 in Armenia (795 cells) to 0.237 in Quimbaya (69 cells). On that same both-sensor set the share of cells above a 0.15 mean drop is 0.19 to 0.66 for Sentinel-1 against 0.00 to 0.13 for NISAR. Sevilla is the extreme case, 66.1 percent of its 56 shared cells above 0.15 in Sentinel-1 and none in NISAR. Fig. 11 draws that cell-level disagreement. Both shares in this paragraph are computed on the same cells; the published table ``s1_nisar_cells_250m.csv`` reports the Sentinel-1 share on the wider Sentinel-1-only set (0.19 to 0.58) and the NISAR share on the shared set, which is why the two do not match cell for cell.

Two limits bound any per-building reading. Where two adjacent Cali bursts both give a value, the two independent estimates of the same 7,555 buildings agree only at Spearman 0.138. Across cities the adjacent-burst repeatability runs from 0.14 to 0.37, with Quibdó at 0.296 and 0.372. We adopt the lowest of that range as the repeatability floor of the product, which is the conservative choice. The second limit is coverage, stated above: three quarters of Pereira footprints receive no value at all.

Take-away: rainfall moved 6-day C-band coherence by an amount comparable to the co-event drop itself, so a single co-event 6-day pair cannot separate rain from damage. Including the wet 21 July acquisition lowers the building pre-event median by 0.052 in Quimbaya to 0.109 in Sevilla and the full-baseline drop median by 0.043 to 0.083, against co-event near-baseline drops of 0.096 to 0.175 (Table 5.3); at burst level the wet pair sits 25 to 37 percent below the other pre-event pairs, which for burst 303904 in subswath IW1 is a median of 0.173 against 0.207 to 0.237 for the other five pairs. This is why we use two pre-event pairs and report the third as a lower bound, and why a coherence map shown to the team by a third party during the mission has to be read with its baseline dates in hand.

5.3 OPERA RTC backscatter change

As a third, weak proxy we differenced OPERA RTC-S1 backscatter at 30 m in vertical transmit vertical receive (VV) polarisation, using the closest pre-event and first post-event acquisition per burst: track 142 spanning 2 to 14 August and track 48 spanning 7 to 19 August. The change is the absolute log ratio in decibels divided by the standard deviation of the pre-event pairs, with that denominator floored at 0.5 dB. The floor matters because where it binds the quantity is not a z-score but twice the absolute log ratio in decibels, and it binds on 13 percent of built-up pixels across the seven cities, from 0.5 percent in Quibdó to 25 percent in Sevilla.

Inside footprints the median normalised change ('abs_z') runs from 0.84 in Manizales to 1.03 in Cali and the share of building pixels changing by more than 3 dB, a doubling or halving of backscatter power, from 0.071 in Armenia to 0.125 in Quibdó. Neither quantity orders the cities the way coherence does: across the seven cities the median normalised change correlates with the Sentinel-1 near-baseline drop at Spearman 0.39 and the share above 3 dB at Spearman 0.14, neither significant at seven points. Fig. 12 zooms on Pereira and shows the two layers failing to outline the same area.

Take-away: the RTC layer separates the seven cities no better than chance at this sample size, and it is a cross-check on a coherence flag rather than an independent screen.

5.4 Is the drop building-specific? The neighbourhood-baseline test in Pereira

Labelled Pereira buildings are not a random sample: they sit a median 1.4 km from the building-density peak of the city against 5.0 km for unflagged buildings, start from higher pre-event coherence, 0.63 against 0.51, and are larger, 334 m² against 183 m². All three confounders push the drop toward the claimed effect. The test that separates a building effect from a neighbourhood effect subtracts, for each labelled building, the median drop of the unflagged buildings within 200 m, so a neighbourhood-wide drop cancels and only the building-specific excess remains. We chose 200 m because it is the smallest round radius larger than the 140 m Sentinel-1 coherence window, so the baseline buildings are drawn from outside the window that produced the target building's own value while staying inside the same neighbourhood.

Labels come from the Microsoft (MS) AI for Good Pereira product and from CEMS. The MS reviewer classes are desk reviews of imagery, not field observations, so the class named `'ground_truth'` is a reviewer judgement like the others and we use the MS class names verbatim throughout rather than paraphrasing them.

Table 5.4. Pereira label test, raw drop and excess over the 200 m unflagged baseline, on the near-baseline Sentinel-1 drop and the NISAR T047A f004 drop. Provisional: one co-event pair per sensor, 41 to 46 labelled buildings in the headline class. MS reviewer classes are joined to footprints by nearest centroid within 3 m; CEMS points by point in polygon, then nearest footprint within 10 m. The two n columns differ because each sensor's excess is computed on its own valid population. Source: [catalogue/s1_screen/out20/final/fig12_pereira_label_test_stats.json](#).

Group	n (S1, NISAR)	S1 raw	S1 excess	NISAR raw	NISAR excess
Unflagged baseline	22,472, 22,435	0.105	reference	0.072	reference
MS rejected	65, 65	0.127	-0.007	0.090	+0.022
MS unsure	60, 52	0.161	-0.016	0.114	+0.013
MS confirmed	55, 62	0.218	+0.023	0.182	+0.048
MS ground truth	41, 46	0.366	+0.061	0.344	+0.135
CEMS Damaged	28, 28	0.320	-0.017	0.262	+0.085
CEMS Destroyed	31, 31	0.261	+0.004	0.316	+0.141
CEMS Possibly damaged	8, 8	0.284	-0.046	0.148	+0.001

The raw drop rises with label grade in both sensors and, in Sentinel-1, collapses once the neighbourhood is removed: the excess is +0.061 for 'ground_truth' and +0.023 for 'confirmed', while 'rejected', 'unsure' and all three CEMS classes lie between -0.046 and +0.004. That removes 76 to 79 percent of the raw 'ground_truth' and 'confirmed' gap. Fig. 13 shows the raw NISAR comparison before the neighbourhood is removed, Fig. 15 the same test on Sentinel-1 before and after, and Fig. 14 the two sensors on the same labelled buildings. NISAR on the same footprints keeps a building-specific excess of +0.135 for 'ground_truth', +0.141 for CEMS Destroyed, +0.085 for CEMS Damaged and +0.048 for 'confirmed'.

Significance is tested with a two-sided one-sample Wilcoxon signed-rank test of the per-building excess against zero. Fourteen tests are reported here, seven groups in two sensors, and at a Bonferroni threshold of 0.05 divided by 14, that is 0.0036, five survive: the NISAR 'ground_truth' result at $p = 1.8e-9$, NISAR CEMS Destroyed at $p = 1.5e-6$, NISAR CEMS Damaged at $p = 4.7e-6$, NISAR 'confirmed' at $p = 5.0e-4$ and the Sentinel-1 'confirmed' result at $p = 0.0023$. The Sentinel-1 'ground_truth' excess at $p = 0.014$ does not survive the correction, and neither does the NISAR 'unsure' excess at $p = 0.032$. We report the corrected reading, which is the conservative one for the Sentinel-1 claim and which removes the one result that would otherwise undercut NISAR.

Specificity has a floor, and the rejected class sets it. The NISAR excess for 'rejected' buildings is +0.022, which is 46 percent of the 'confirmed' excess of +0.048 and sits above the 'unsure' excess of +0.013, so the NISAR excess ladder is not monotone even though the raw drop is. On the two-track cross-check the rejected excess is higher still, +0.030 on T047A f004 against +0.048 for 'confirmed', and +0.040 on T011D f086 against +0.034 for 'confirmed', where the rejected point estimate exceeds the confirmed one. None of these rejected or unsure excesses is significant after correction, so the honest statement is that they are not distinguishable from zero at this sample size, not that they are zero. A rejected building that shows an L-band excess is exactly the case a site visit can settle, which is why section 8 asks for rejected buildings in the calibration sample.

The lower-bound layer gives a second reading, and it does not agree with the first for every class. On the full-baseline drop the Sentinel-1 excess is +0.068 for 'ground_truth', +0.020 for 'confirmed', -0.004 for both 'rejected' and 'unsure', +0.061 for CEMS Destroyed, +0.002 for CEMS Damaged and -0.094 for CEMS Possibly damaged. CEMS Destroyed therefore moves from null on the near-baseline layer, +0.004 on 31 buildings, to headline magnitude on the lower-bound layer, +0.061 on 14 buildings, and CEMS Possibly damaged from -0.046 to -0.094. The two layers differ in their pre-event baseline, one wet pair in or out, and in their matched CEMS population, so neither the sign nor the magnitude of a small CEMS class is stable in Sentinel-1. The independent NISAR track T011D f086 gives +0.138 on 46 'ground_truth' buildings against +0.135 on the T047A track, so the L-band result reproduces on a second track (Fig. 16). These are 41 Sentinel-1 and 46 NISAR 'ground_truth' buildings, not a single shared set.

Sample sizes for the same group differ between files and the difference is not an error. In 'label_stats_final.csv' the 'ground_truth' row carries $n = 41$ and a NISAR count of 27; in 'fig12_pereira_label_test_stats.json' the same group carries $n = 41$ for Sentinel-1 and 46 for NISAR. The 41 are the buildings with a valid Sentinel-1 value, the 46 those with a valid NISAR value, and the 27 the intersection of the two, which is the population the 'n_nisar' column reports. The same holds for 'confirmed', at 55, 62 and 34. The per-sensor counts are of record for this report, because each sensor's excess is computed on its own valid population. The CEMS counts differ for a different reason: Table 5.4 reports Damaged at 28 and Destroyed at 31 while the earlier cross-check table reports the same classes at 14 and 16, with correspondingly larger NISAR T047A excesses of +0.094 and +0.225. The two CEMS populations are matched differently and are not interchangeable; Table 5.4 and its documented 10 m matching rule are of record here.

Two further checks bear on whether the drop is real. No pre-event pair-to-pair analogue separates the labels, at an area under the curve of 0.42 to 0.50, while the co-event drop gives a permutation z of 9, so the separation is specific to the 8 to 14 August pair and is not a standing property of these buildings. And the same test run in two other cities gives one weak positive and one null: in Cali the 80 MS-flagged buildings carry a Sentinel-1 excess of +0.019 at p 0.045, and in Quibdó the 26 matched CEMS buildings carry -0.008 at p 0.627, with Cali's own eight matched CEMS buildings at -0.029 and p 0.844. Cali is a second, marginal Sentinel-1 building-specific excess and softens the reading that C-band carries no building-level signal at all; Quibdó is null and supports it. Neither city has a NISAR building-level test, because Cali's NISAR coverage comes from a different track and Quibdó has too few matched labels.

A fourth confounder sits underneath the three named at the head of this subsection and is treated separately in section 5.5. The labelled buildings are drawn from the optical producers' own valid areas, and those areas do not coincide: 43.2 percent of the positive labels but only 10.3 percent of the reviewer-rejected negatives lie inside the ground all three producers assessed. The 200 m neighbourhood baseline of this subsection removes a local level and does not remove that asymmetry, because the asymmetry is between the classes rather than within a neighbourhood. Section 5.5 restricts the comparison to the common valid area and reports what changes.

Take-away: on the 41 to 46 labelled buildings of Table 5.4, in one city and on one co-event pair per sensor, the Sentinel-1 drop is explained by the neighbourhood and the NISAR drop retains a building-specific residual on the larger buildings, with a specificity floor set by an unflagged-like excess on rejected buildings that the field team is asked to test; restricted to the common valid area the NISAR excess remains a separating feature at an AUC of 0.8458 on T011D while the Sentinel-1 excess falls below chance at 0.4600 (Table 5.5).

5.5 The technique ranking inside and outside the optical valid areas

One distinction governs everything in this subsection. The radar screens have no valid-area problem. Sentinel-1, NISAR, the OPERA RTC layer and the two products of the Balz coherence chain each return a value for every footprint they can resolve across the whole city, and where they return nothing the cause is geometric, a footprint holding no pixel centre or a burst that does not reach that part of the city, stated per city in Table 5.1 and Table 5.3. The labels these screens are scored against are optical, and every optical producer publishes a valid-area mask outside which it recorded nothing at all (section 4.1). A radar screen can therefore be validated against field observations anywhere the field team goes, while any comparison that involves an optical product, including the use of optical labels to score a radar screen, is valid only inside the common valid area of the optical producers concerned. Table 5.5 reports the Pereira ranking in both views for that reason: corrected, meaning the 12,023 footprints of the common valid area of section 4.1, and uncorrected, meaning all 91,266 footprints of the city.

The label set is the one section 4 describes. Positives are buildings the Microsoft reviewers marked 'ground_truth' or 'confirmed', or that carry a CEMS Destroyed or Damaged point within 15 m; negatives are the reviewer-rejected buildings that carry no such point. Over all footprints the comparison rests on 726 labelled buildings, 368 positives and 358 negatives; inside the common valid area 196 survive, 159 positives and 37 negatives.

Table 5.5. The twelve building-level techniques ranked against the same Pereira labels in both views, ordered by their corrected score. Every feature is oriented so that a larger value means more damage, which is the "orientation" column of the source file; the two coherence products of the Balz chain are inverted for scoring because low coherence means change. Values are given to four decimal places, as stored in the source file. The AUC is 0.5 at chance and 1.0 at perfect separation. The interval is a 95 percent stratified bootstrap interval on 2,000 replicates, reported for the corrected view because that is the view this report relies on. The n columns differ between techniques because each is scored on its own valid

population. Technique names are the "technique" column of the source file and retain its spelling. Source: catalogue/s1_screen/out_validarea/technique_ranking.csv. Providers: NISAR (NASA and ISRO) via ASF; Copernicus Sentinel-1 (ESA); OPERA RTC-S1 via ASF; Balz coherence products; labels from Microsoft AI for Good and the Copernicus Emergency Management Service; footprints from Overture Maps.

Technique	Sensor	Uncorrected AUC (rank)	n	Corrected AUC (rank)	n	Corrected 95 percent interval
NISAR T011D co-event coherence drop	NISAR	0.7542 (3)	725	0.8953 (1)	196	0.8405 to 0.9436
NISAR T047A co-event coherence drop	NISAR	0.7955 (2)	725	0.8819 (2)	196	0.8278 to 0.9291
NISAR T011D drop, neighbourhood excess	NISAR	0.6728 (5)	723	0.8458 (3)	196	0.7775 to 0.9036
Balz NISAR machine-learning coherence	NISAR	0.5690 (7)	704	0.8099 (4)	188	0.7169 to 0.8924
Balz Sentinel-1 machine-learning coherence	Sentinel-1	0.5441 (10)	726	0.7855 (5)	196	0.6993 to 0.8593
NISAR T047A drop, neighbourhood excess	NISAR	0.6917 (4)	723	0.7408 (6)	196	0.6602 to 0.8144
Sentinel-1 co-event coherence drop	Sentinel-1	0.8112 (1)	726	0.7204 (7)	196	0.6305 to 0.8018
Balz NISAR change fraction	NISAR	0.6052 (6)	704	0.6686 (8)	188	0.6220 to 0.7138
OPERA RTC backscatter absolute z	Sentinel-1 RTC	0.5636 (9)	726	0.6006 (9)	196	0.4882 to 0.7090
Balz Sentinel-1 change fraction	Sentinel-1	0.5441 (10)	726	0.5483 (10)	196	0.5072 to 0.5849
Sentinel-1 drop, neighbourhood excess	Sentinel-1	0.5665 (8)	723	0.4600 (11)	196	0.3660 to 0.5547
OPERA RTC backscatter log-ratio	Sentinel-1 RTC	0.4707 (12)	726	0.4524 (12)	196	0.3579 to 0.5446

The correction inverts the sensor ranking. The Sentinel-1 co-event coherence drop is first of the twelve over all footprints at 0.8112 and seventh at 0.7204 inside the common valid area, while the NISAR T011D co-event coherence drop moves from third at 0.7542 to first at 0.8953 and NISAR T047A holds second place while gaining 0.0864. The two machine-learning coherence products move furthest and move together, the NISAR one from 0.5690 to 0.8099 and the Sentinel-1 one from 0.5441 to 0.7855, which is the cleanest statement of the mechanism: a method that carries no location information gains most when location stops contaminating the comparison. The OPERA RTC log-ratio sits below chance in both views, 0.4707 and 0.4524, and is carried in the table as the negative control.

The bootstrap intervals say which parts of that reordering are separable. Inside the common valid area the interval for NISAR T011D, 0.8405 to 0.9436, does not overlap the interval for Sentinel-1, 0.6305 to 0.8018, so the corrected first place is defensible. Over all footprints the same two leaders, Sentinel-1 at 0.7803 to 0.8413 and NISAR T047A at 0.7650 to 0.8268, overlap heavily, so the uncorrected first place was never separable from second either.

Two readings of the table need guarding. The Sentinel-1 neighbourhood excess of section 5.4 scores 0.4600 inside the common valid area, below chance, under the same orientation that gives it 0.5665 over all footprints, so it ranks eleventh of twelve, above the negative control and below everything else. And precision inside the common valid area looks spectacular for every technique, 0.8470 to 0.9821 at the Youden threshold, only because the base rate of positives there is 0.8112, that is 159 of 196; the same column over all footprints runs 0.5383 to 0.9032 at a base rate of 0.5028 to 0.5090. No precision figure from the corrected view should be quoted without its base rate beside it.

The techniques agree with one another about as little in one view as in the other, and slightly less over all footprints. Table 5.6 gives the Spearman rank correlation between the six primary screens over the buildings each pair shares. The two NISAR tracks, which observe the same buildings with the same instrument from two look directions, agree at 0.3014 inside the common valid area and 0.1825 over all footprints. Sentinel-1 against the two NISAR tracks reaches 0.3407 and 0.2262 inside, against 0.1625 and 0.1084 over all footprints. Excluding the three pairs that compare a drop with its own neighbourhood excess, every other pair of the twelve lies between -0.0623 and 0.1825 over all footprints and between -0.0529 and 0.3407 inside the common valid area, and 54 of the 66 pairs rise inside it, 58 of 66 in magnitude. Apparent agreement between techniques is therefore deflated and not inflated by the uncorrected view, and the near-zero conclusion survives either way.

Table 5.6. Spearman rank correlation between the six primary screens over the Pereira buildings each pair shares, corrected first and uncorrected in parentheses. Every value is damage-oriented, that is computed after inverting the two coherence products so that a larger value means more damage in all six rows and columns; the raw-feature reading of the same pairs, where the Balz coherence keeps its own sign, is given in the paragraph below and in the "spearman_raw_feature" column of the source file. A value near zero means the two screens rank the same buildings almost independently of each other. Source: catalogue/s1_screen/out_validarea/technique_agreement.csv. Providers as Table 5.5.

	Sentinel-1 drop	NISAR T047A drop	NISAR T011D drop	Balz NISAR coherence	Balz Sentinel-1 coherence
NISAR T047A drop	0.3407 (0.1625)				
NISAR T011D drop	0.2262 (0.1084)	0.3014 (0.1825)			
Balz NISAR coherence	-0.0229 (-0.0526)	0.0798 (0.0207)	0.2125 (0.1240)		
Balz Sentinel-1 coherence	0.1915 (-0.0623)	0.2012 (0.1036)	0.1262 (0.0862)	0.0966 (0.1217)	
OPERA RTC absolute z	0.1094 (0.0492)	0.1345 (0.0568)	0.0687 (0.0367)	0.0159 (-0.0047)	0.0205 (0.0318)

Read on the raw features rather than on the damage-oriented ones, the Balz coherence correlations carry the sign the physics requires. Inside the common valid area the Balz NISAR coherence correlates with the NISAR T047A drop at -0.0798 and with the NISAR T011D drop at -0.2125, and the Balz Sentinel-1 coherence with the Sentinel-1 drop at -0.1915, because a larger drop and a lower coherence are the same statement about the same ground. That sign is a consistency check on the chain and not a second measurement.

Near-zero agreement between techniques is one fact with two consequences, and this report needs both. It is bad news for per-building confidence: two look directions of the same instrument over the same buildings ranking them at 0.3014 confirms that a per-building coherence value is dominated by estimation noise, which is what the 750 square metre and 19,600 square metre supports of sections 5.1 and 5.2 already imply. It is good news for design: near-independent screens are what makes a fused score worth more than its best component, and what makes a disagreement stratum informative, because a building two independent screens disagree about is a building whose true state neither of them determines and which a site visit therefore resolves. Section 8 carries that stratum.

Three limits bind this subsection. The corrected column of Table 5.5 rests on 196 labelled buildings of which only 37 are negatives, which is why the intervals are wide and why the reordering below the top two positions is not itself a ranking. The common valid area is the city centre and reaches 4.19 km from it, so nothing in Table 5.5 speaks to the periphery, which is where the radar screens are unvalidated and where section 8 sends field teams. And the labels are Microsoft reviewer verdicts and CEMS grades, both photointerpretation rather than field observation, so Table 5.5 measures agreement between a radar screen and an optical assessment and not accuracy against truth.

Take-away: restricted to the ground all three optical producers assessed, the NISAR co-event coherence drop is the best separating screen at an AUC of 0.8953 and the Sentinel-1 drop falls from first to seventh at 0.7204, the two machine-learning coherence products gain most, and the screens rank buildings almost independently of one another in both views (Table 5.6), which is bad for any single per-building call and good for fusion and for disagreement-based field sampling.

5.6 What the geolocation offset means for the per-building results

The 40 m offset is a property of our own Sentinel-1 geocoding chain, which uses Copernicus GLO-30 heights, and the source scopes it for city and neighbourhood statistics and for 10 m per-building work but not for the intermediate case reported here, per-building values read off a 20 m product. Two 140 m windows offset 40 m east and 5 m south still share about 69 percent of their area, (140 minus 40) times (140 minus 5) divided by 140 squared, so the expected effect on Table 5.3, Table 5.4 and the +0.061 Sentinel-1 excess is a dilution of the building-to-pixel assignment rather than a loss of it. That excess has not been recomputed with the footprints shifted 40 m east and 5 m south, so we label the Sentinel-1 building-level result provisional on that test rather than concluding against it.

The NISAR GUNW products arrive geocoded from ASF and do not pass through our DEM step, so they do not carry this offset, but we have not run an independent footprint-shift test on them. Until that test is run, the possibility remains that the geolocation quality is asymmetric between the two sensors in the same direction as the section's result, and the asymmetry belongs in the field calibration task rather than in a footnote.

Take-away: the Sentinel-1 per-building numbers in this section carry a two-pixel offset whose effect is dilution, the NISAR numbers are believed free of it, and neither statement has been tested by shifting the footprints.

Figures

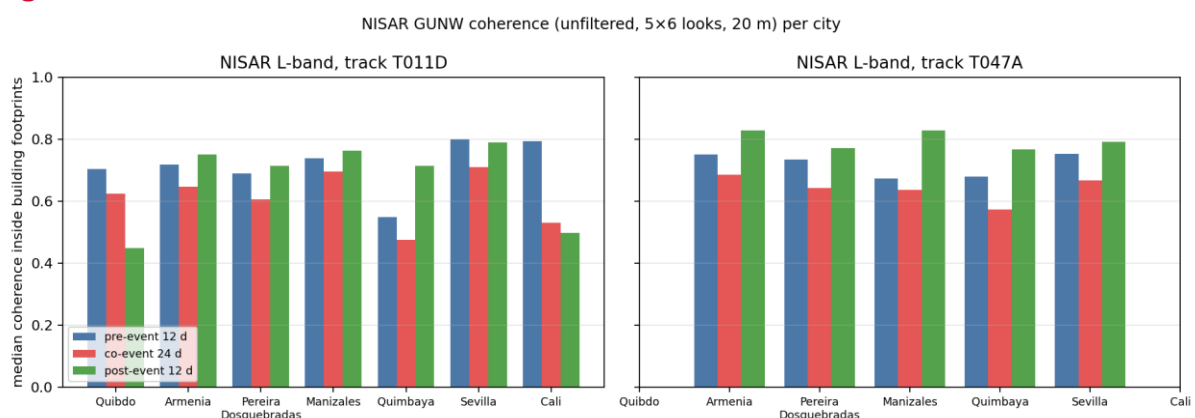


Fig. 5. NISAR footprint coherence per phase and city, showing a co-event drop of a few hundredths in every city and Quimbaya lowest throughout. Source: catalogue/s1_screen/out_nisar/gunw_city_stats.csv. Providers: NISAR (NASA and ISRO) via ASF; Overture Maps.

NISAR L-band GUNW coherence (unfiltered, 5x6 looks, 20 m), track 011 descending; Cali from frame 087. Hollow circles in Pereira = CEMS AOI02 damaged-building points

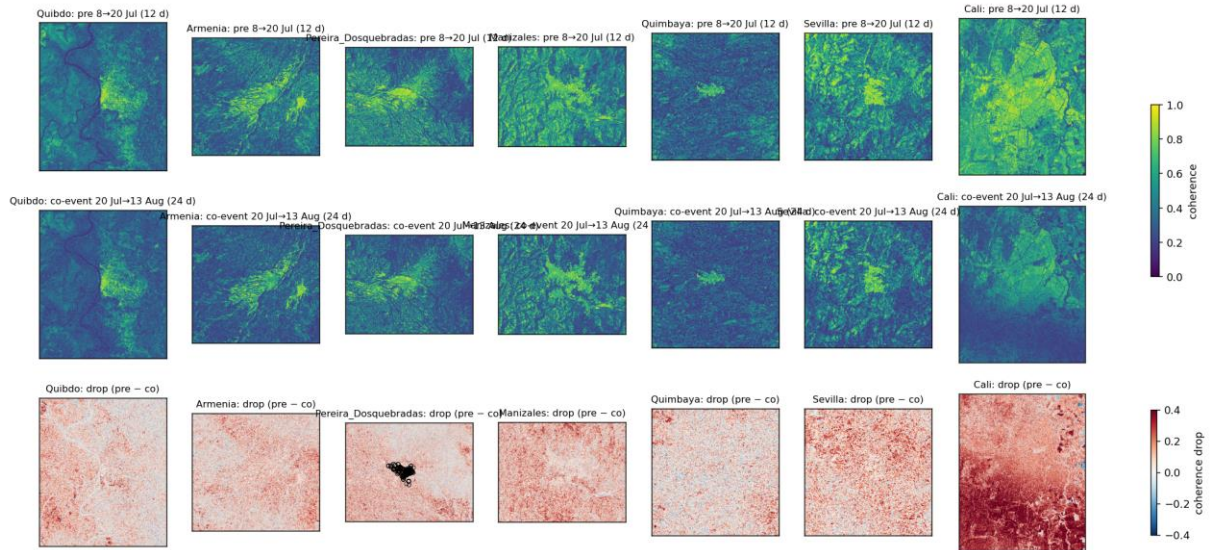


Fig. 6. NISAR coherence and drop maps for the seven cities, showing that the drop is spread across each urban area rather than concentrated in identifiable blocks. Source: the per-phase coherence rasters listed in catalogue/s1_screen/out_nisar/gunw_city_stats.csv, column "file". Providers: NISAR via ASF; Overture Maps.

Cali: the frame-wide coherence drop on track 011 (20 Jul→13 Aug, top) is not reproduced by track 148 (18 Jul→11 Aug, bottom)

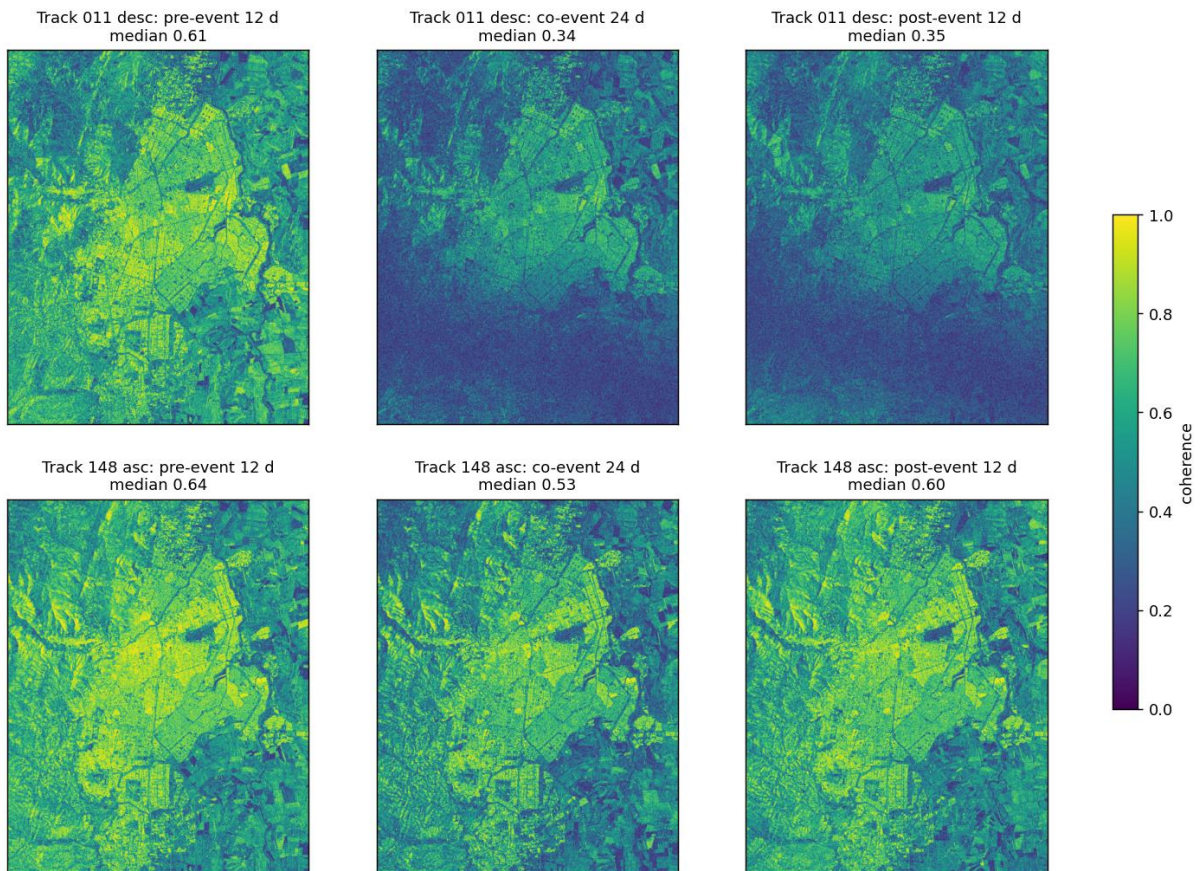


Fig. 7. Cali on T011D and T148A, showing that the frame-wide T011D decorrelation is an artefact and not a Cali result. Source: catalogue/s1_screen/out_nisar/gunw_city_stats.csv, Cali rows for T011D f087 and T148A f003, and the rasters they name. Provider: NISAR via ASF.

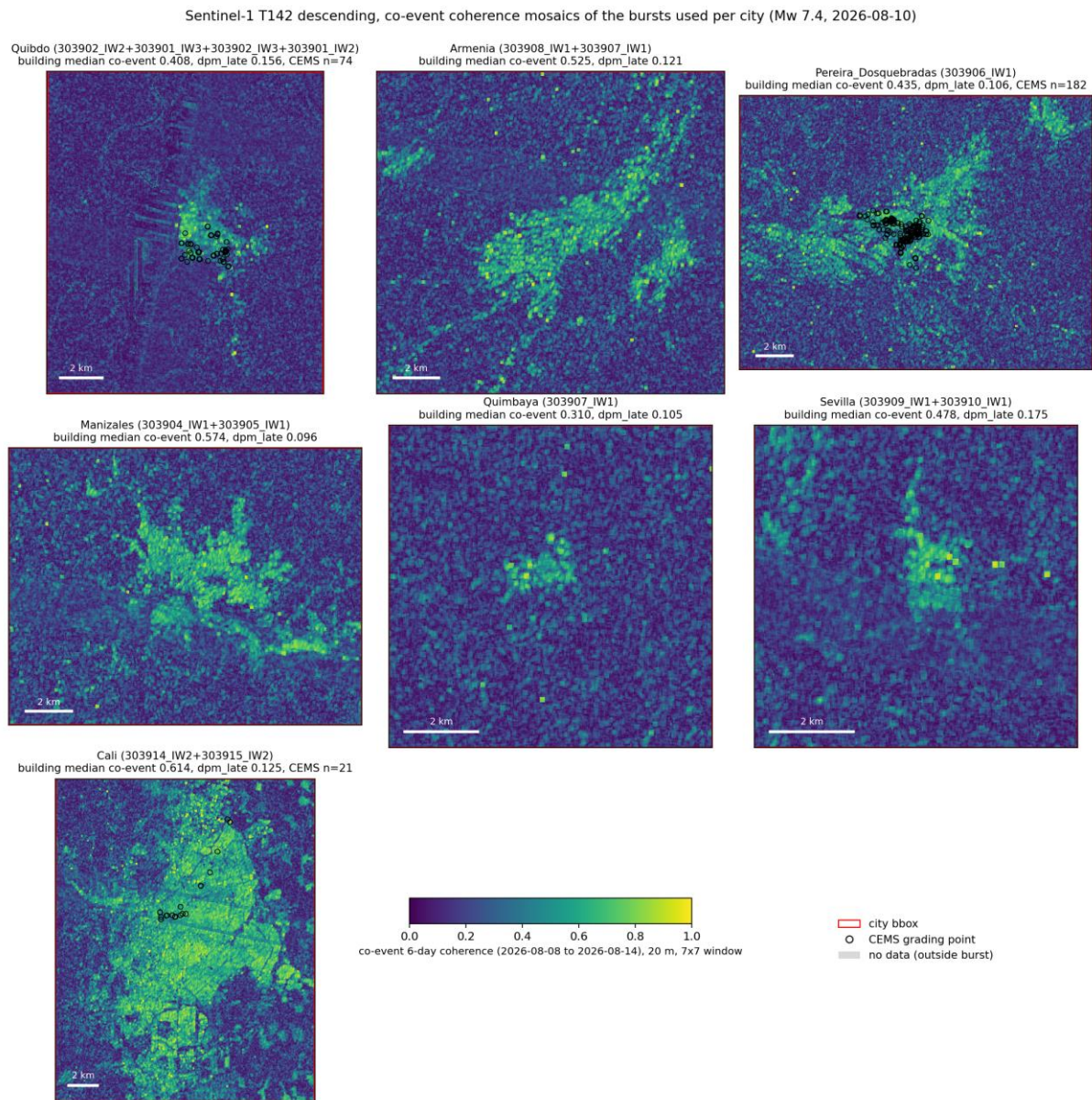


Fig. 8. Sentinel-1 co-event coherence for 8 to 14 August, showing that absolute coherence differs between cities by more than the co-event drop does, which is why the drop and not the level is the screen. Source: catalogue/s1_screen/out20/final/fig9_log.json and the per-burst co-event coherence rasters it lists. Provider: Copernicus Sentinel-1 (ESA), processed by us.

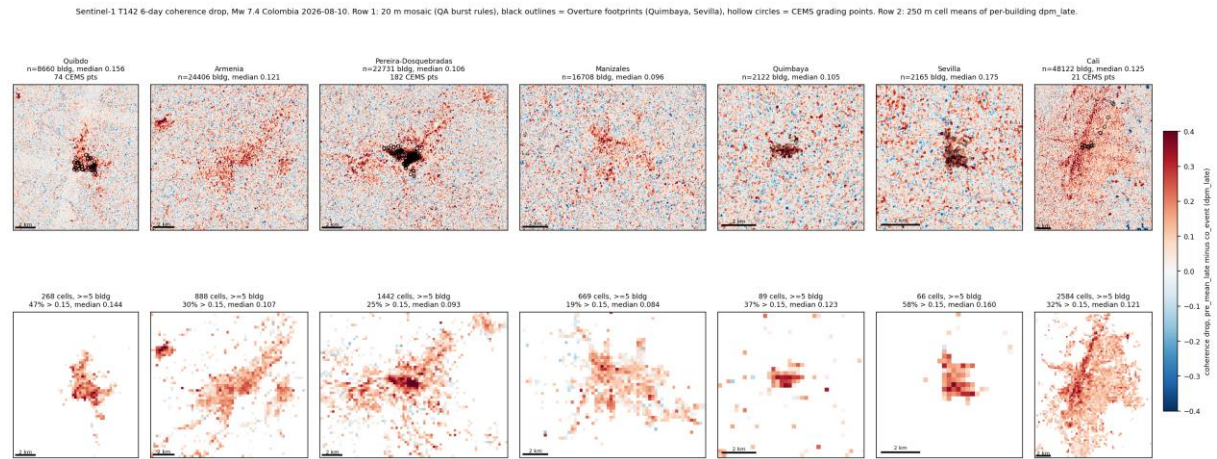


Fig. 9. the Sentinel-1 near-baseline drop at building and 250 m cell level, the layer that delimits the affected core. Source: catalogue/s1_screen/out20/final/fig10_stats.json and the seven '<City>_cells250.csv' files. Providers: Copernicus Sentinel-1 (ESA); Overture Maps.

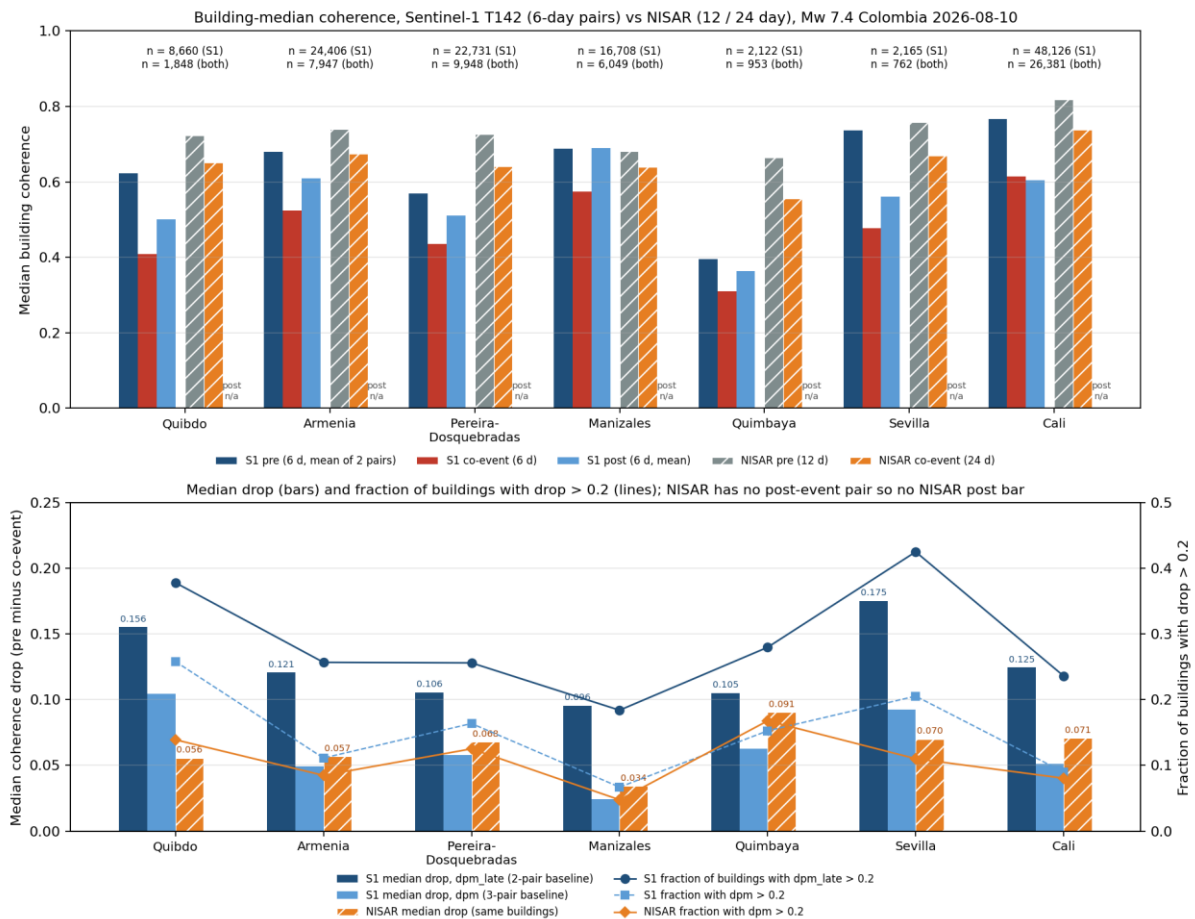


Fig. 10. the two sensors on shared buildings, showing the 1.3 to 3.0 drop ratio and a different city ranking. Source: catalogue/s1_screen/out20/final/fig11_values.csv. Providers: Copernicus Sentinel-1 (ESA); NISAR via ASF.

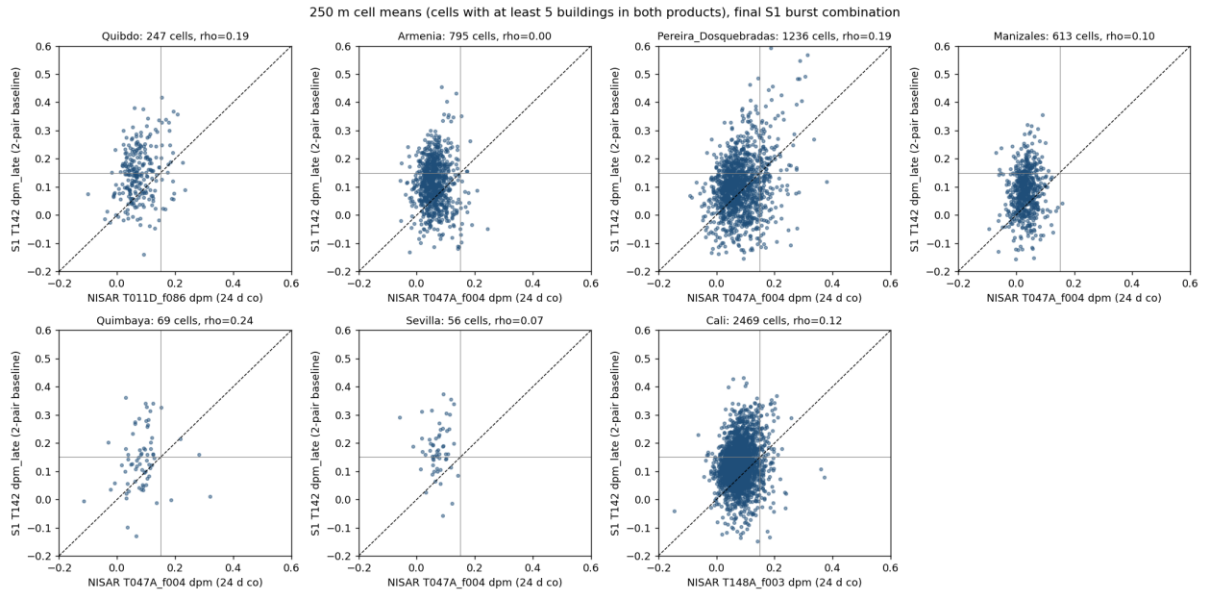


Fig. 11. sensor agreement at 250 m, weak everywhere and near zero in Armenia. Source: catalogue/s1_screen/out20/final/s1_nisar_cells_250m.csv. Providers as Fig. 10.

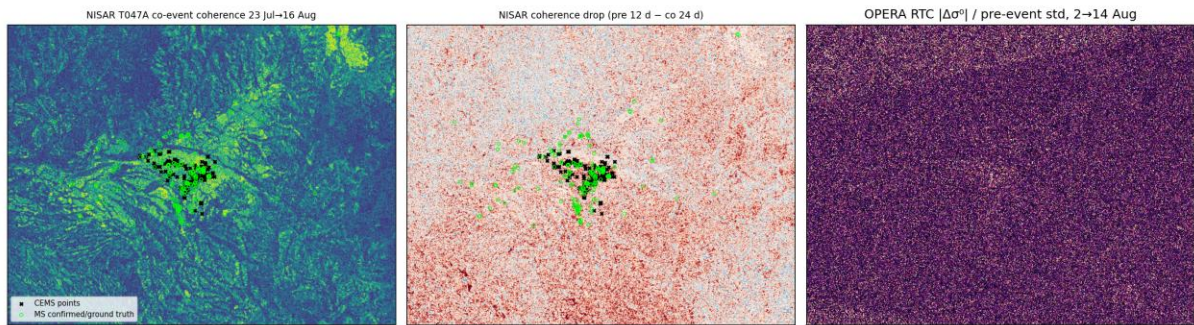


Fig. 12. Pereira zoom combining the coherence drop and the RTC backscatter change, showing that the two layers do not outline the same area. Source: catalogue/s1_screen/out_rtc/Pereira_Dosquebradas_logratio_dB.tif and Pereira_Dosquebradas_abs_z.tif with the drop from catalogue/s1_screen/out20/final/Pereira_Dosquebradas_buildings_s1_final.gpkg. Providers: Copernicus Sentinel-1 (ESA); OPERA RTC-S1 via ASF.

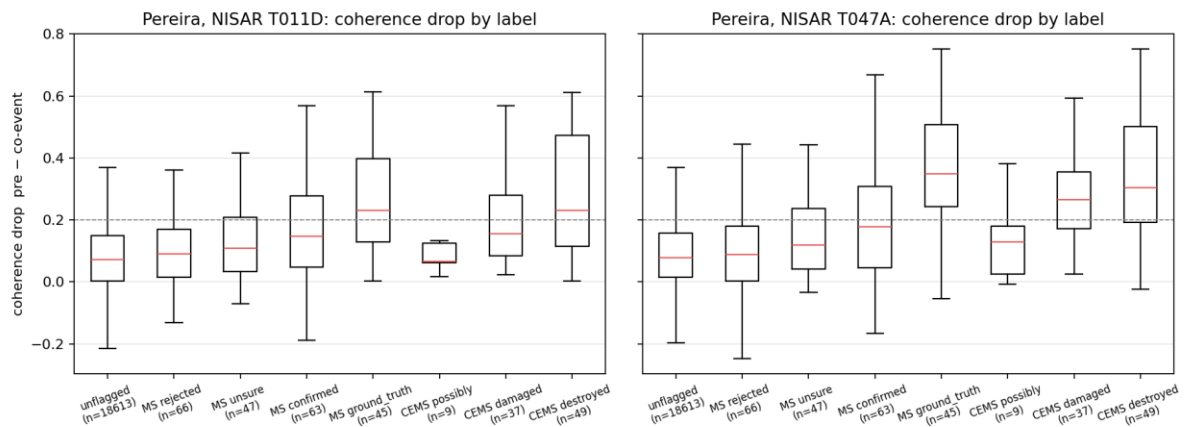


Fig. 13. NISAR drop by MS reviewer class and CEMS grade, the raw comparison before the neighbourhood baseline is removed. Source: catalogue/s1_screen/out_nisar/Pereira_Dosquebradas_T047A_f004_buildings_dpm.gpkg. Providers: NISAR via ASF; Microsoft AI for Good; Copernicus Emergency Management Service.

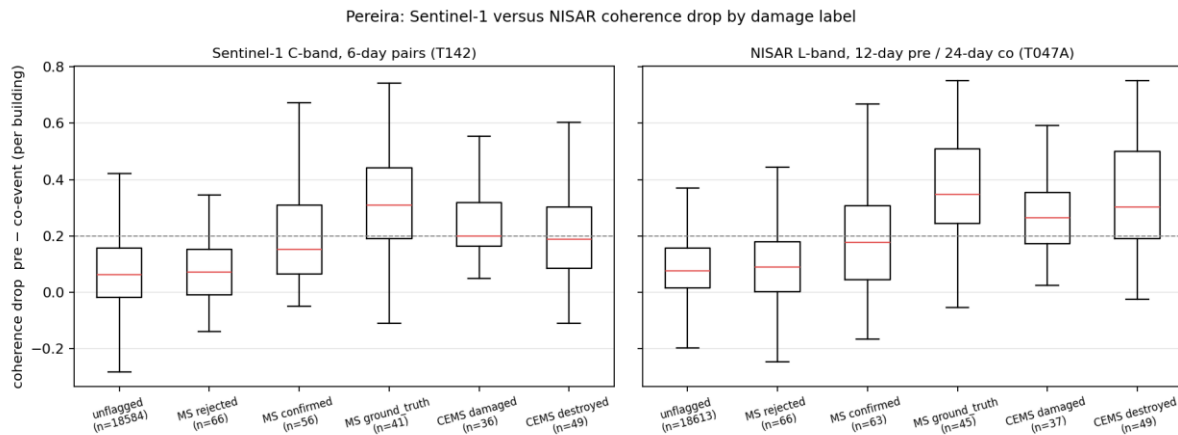


Fig. 14. the same Pereira labels in both sensors, agreeing by group and weakly building by building. Source: catalogue/s1_screen/out20/final/Pereira_Dosquebradas_buildings_s1_final.gpkg. Providers as Fig. 13, plus Copernicus Sentinel-1 (ESA).

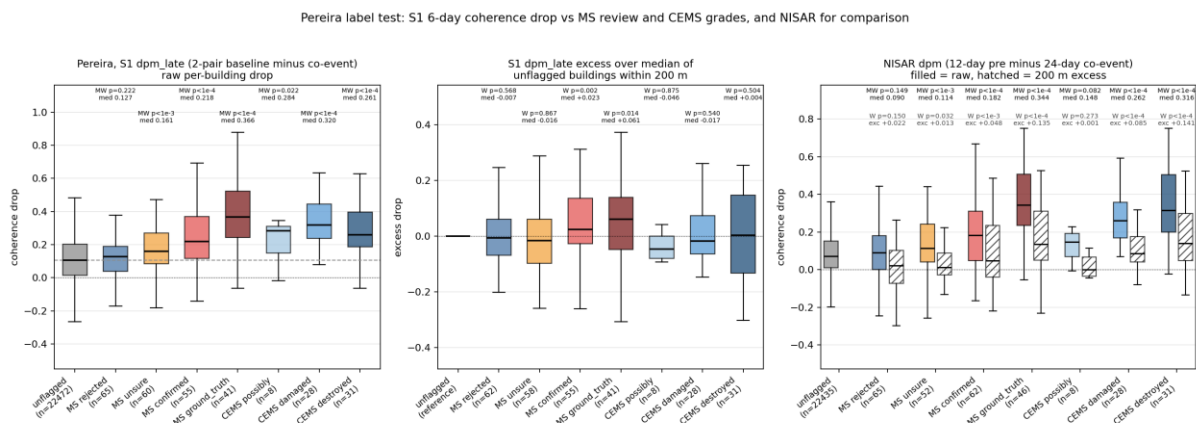


Fig. 15. the label test on the Sentinel-1 near-baseline drop, showing the raw drop against the excess over the 200 m baseline and the collapse of the gap. Source: catalogue/s1_screen/out20/final/fig12_pereira_label_test_stats.json. Providers as Fig. 14; footprints from Overture Maps.

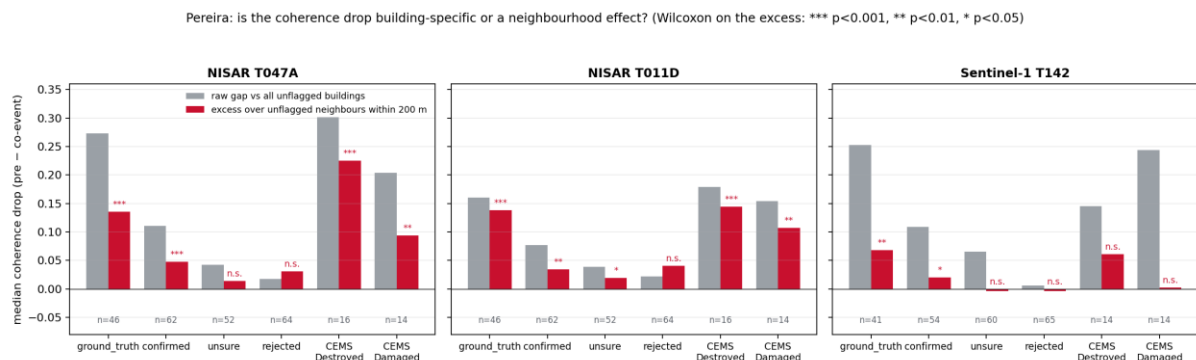


Fig. 16. the neighbourhood-baseline test on both NISAR tracks and Sentinel-1, showing that the L-band excess survives on two independent tracks while the C-band excess does not. Source: catalogue/s1_screen/out20/qa/local_baseline_excess_pereira.csv. Providers as Fig. 15, plus NISAR via ASF.

Section take-away: the Sentinel-1 drop is a block-level prioritisation layer that delimits the affected core in all seven cities (Table 5.3, Fig. 9) and the NISAR drop is the layer that retains a building-specific residual and is therefore the one to calibrate against field observations (Table 5.4, Fig. 16), and restricting the label comparison to the ground all three optical producers assessed strengthens that conclusion rather than qualifying it, moving NISAR to first at an AUC of 0.8953 and Sentinel-1 from first to seventh at 0.7204 (Table 5.5). The sensor contrast follows from estimation support first, 19,600 square metres against 750 (sections 5.2 and 5.1), and from wavelength second, and the two are confounded in this dataset because we have not run Sentinel-1 with a smaller window or NISAR degraded to the 140 m window of section 5.2. Everything here is provisional: one co-event pair per sensor, provisional NISAR products, a building-level claim resting on the 41 to 46 labelled buildings of Table 5.4 in one city, and a corrected ranking resting on 196 labelled buildings of which 37 are negatives, in the central 4.19 km of that same city. The radar screens themselves cover all seven cities in full, so field validation of a radar screen is possible anywhere the itinerary goes, and it is only the comparisons against optical labels that are confined to the common valid area. The instructions that follow from these results, with the city, the product file and the action for each, are in section 8.

Sources: `PROCESSING\_LOG.md`, `SAR\_SCREEN\_RESULTS.md`,
`catalogue/imagery\_acquisitions\_timeline.csv`, `catalogue/s1\_screen/out20/final`,
`catalogue/s1\_screen/out20/final/Pereira\_Dosquebradas\_buildings\_s1\_final.gpkg`,
`catalogue/s1\_screen/out20/final/SECTION\_7.md`,
`catalogue/s1\_screen/out20/final/Sevilla\_cells250.csv`, `catalogue/s1\_screen/out20/final/build\_final.py`,
`catalogue/s1\_screen/out20/final/fig10\_stats.json`, `catalogue/s1\_screen/out20/final/fig11\_values.csv`,
`catalogue/s1\_screen/out20/final/fig12\_pereira\_label\_test\_stats.json`,
`catalogue/s1\_screen/out20/final/fig9\_log.json`, `catalogue/s1\_screen/out20/final/label\_stats\_final.csv`,
`catalogue/s1\_screen/out20/final/s1\_city\_table.csv`,
`catalogue/s1\_screen/out20/final/s1\_nisar\_cells\_250m.csv`,
`catalogue/s1\_screen/out20/final/test\_burst\_geolocation\_table\_rerun.csv`,
`catalogue/s1\_screen/out20/qa/local\_baseline\_excess\_pereira.csv`,
`catalogue/s1\_screen/out20/qa2/Cali\_qa2\_results.json`,
`catalogue/s1\_screen/out\_experiments/pereira\_features\_README.md`,
`catalogue/s1\_screen/out\_nisar/Pereira\_Dosquebradas\_T047A\_f004\_buildings\_dpm.gpkg`,
`catalogue/s1\_screen/out\_nisar/gunw\_city\_stats.csv`,
`catalogue/s1\_screen/out\_nisar/nisar\_building\_dpm\_stats.csv`, `catalogue/s1\_screen/out\_rtc`,
`catalogue/s1\_screen/out\_rtc/Pereira\_Dosquebradas\_logratio\_dB.tif`,
`catalogue/s1\_screen/out\_rtc/rtc\_city\_stats.csv`, `catalogue/s1\_screen/out\_validarea/label\_geography.csv`,
`catalogue/s1\_screen/out\_validarea/technique\_agreement.csv`,
`catalogue/s1\_screen/out\_validarea/technique\_ranking.csv`,
`catalogue/s1\_screen/out\_validarea/valid\_area\_summary.csv`, `catalogue/s1\_screen/rtc\_screen.py`,
`figures/fig10\_s1\_drop\_maps.png`, `figures/fig11\_s1\_vs\_nisar\_bars.png`,
`figures/fig12\_s1\_pereira\_label\_test.png`, `figures/fig12b\_s1\_vs\_nisar\_cells250.png`,
`figures/fig18\_local\_baseline\_excess\_pereira.png`, `figures/fig1\_nisar\_city\_coherence.png`,
`figures/fig2\_pereira\_dpm\_by\_label.png`, `figures/fig3\_nisar\_maps\_seven\_cities.png`,
`figures/fig4\_pereira\_zoom.png`, `figures/fig6\_cali\_two\_tracks.png`,
`figures/fig7\_pereira\_s1\_vs\_nisar.png`, `figures/fig9\_s1\_coevent\_maps.png`.

6. The Liu and Balz 2026 test and the variable-combination experiments

Three sample sets govern what follows and no result below crosses between them. The neighborhood baseline test of section 5 rests on 41 to 46 buildings of the top Microsoft (MS) reviewer class 'ground_truth', depending on the sensor. The single-pair separation test of Table 6.3 rests on 218 damaged footprints, defined in its own source file as the MS 'ground_truth' class united with the Copernicus Emergency Management Service (CEMS) grades Destroyed and Damaged, against 360 'rejected' and 90,378 unflagged footprints. The classifier experiments of section 6.2 rest on 368 positives against either 358 reviewer-rejected negatives or 3,680 randomly drawn unflagged negatives, out of 91,266 open building outlines in Pereira and Dosquebradas. The 368 positives combine the MS 'ground_truth' class, a desk review of imagery and not a field observation, the 156 buildings of the MS 'confirmed' class, and CEMS points graded Destroyed or Damaged within 15 m of a footprint. The 218 set excludes the 'confirmed' class that the 368 set includes, so Tables 6.3 and 6.4 are not comparable row for row. No area under the receiver operating characteristic curve (ROC AUC) below rests on 46 buildings, and no excess value in section 5 rests on 368. We name each reviewer class by its own label, MS 'ground_truth' and MS 'confirmed', because the source files use the words reviewer-confirmed for the second of them. Nothing here is field-verified.

6.1 The deviation-guided single-pair rule

We implemented the deviation-guided dual-model coherence change detection of Liu and Balz (Liu and Balz 2026; see the reference list in section 9) as 'catalogue/s1_screen/balz_ccd' and ran it over Pereira and Dosquebradas on one Sentinel-1 6-day co-event pair of burst 303906 in vertical transmit, vertical receive (VV) polarization, 8 to 14 August, with the 2 to 8 August pair as control, and on one NISAR geocoded single look complex (GSLC) pair of track T011D frame 086, 20 July to 13 August. The NISAR products in this section are provisional and every NISAR number below carries that status. Two gradient-boosted regressors trained on ten million simulated three by three windows estimate the coherence of a window and the signed deviation of the analytical weighted estimator; a pixel is declared changed where the machine-learning (ML) coherence falls below 0.4 and the deviation exceeds 0.15, the rule of the paper. On an independent synthetic hold-out the ML coherence reaches a root mean square error of 0.153 against 0.176 for the nine-look sample coherence, and the deviation 0.168 against 0.373 for a zero baseline. These are single-day runs whose 10 m products carry the same 40 m toward-sensor offset as the 20 m products of section 5.

Table 6.1 gives the share of Pereira footprints with at least one changed pixel under the rule, by label class, on the Sentinel-1 pairs, with the count behind every share. The message of the table is the contrast between the co-event and control columns in the graded classes and the near absence of that contrast in the background.

Label class	n	Co-event pair, any pixel flagged	Control pair, any pixel flagged
CEMS Destroyed	54	22.2 % (12)	0.0 % (0)
CEMS Damaged	62	12.9 % (8)	1.6 % (1)
CEMS Possibly damaged	18	0.0 % (0)	0.0 % (0)
MS 'ground_truth'	136	13.2 % (18)	0.0 % (0)
MS 'confirmed'	155	7.7 % (12)	3.2 % (5)
MS 'unsure'	154	4.5 % (7)	3.2 % (5)
MS 'rejected'	361	2.8 % (10)	2.5 % (9)
Unflagged background	90,378	3.9 % (3,484)	2.7 % (2,484)

Table 6.1. Share of Pereira footprints with at least one changed pixel under the deviation-guided rule, by label class, on the Sentinel-1 co-event and control pairs. Source:

catalogue/s1_screen/out_balz/s1/Pereira_label_table.csv, columns "n" and "share_flagged_any", rows for pairs "co_0808_0814" and "pre_0802_0808"; the counts in brackets are the product of the two columns.

The CEMS and MS classes overlap in this table: 34 of the 136 MS 'ground_truth' footprints also carry a

CEMS grade. The counts are footprints, not points: 160 of 182 CEMS points fall within 15 m of a footprint, on 134 footprints, so they differ from the counts in sections 4 and 5.

Three footprint counts follow the same class through this section and they differ for one reason. The Liu and Balz label table joins the MS review layer to the footprints by representative point, matching 806 of 813 reviewed buildings, and keeps the CEMS grades as a separate attribute, so MS `ground\_truth` counts 136 and MS `rejected` counts 361. The rule comparison of Table 6.2 lets the CEMS grade take precedence over the MS class, which moves those 34 footprints out of `ground\_truth` and leaves 102, and one footprint out of `rejected` and leaves 360. The experiment table of section 6.2 joins the same review layer by centroid within 1.5 m, retains 359 rejected buildings of the 367 in the Microsoft product, and treats the one that carries a CEMS Destroyed or Damaged point as a positive, which leaves 358 negatives. The same two joins of the CEMS layer differ by one footprint, 134 against 135. The 15 m radius is the design constant behind those counts and it discards 47 of the 182 CEMS points, the ones that fall on no footprint at all.

Fig. 17 maps the two masks the rule produces over Pereira on both sensors. On the NISAR pair the rule flags 50.0 percent of the CEMS Destroyed footprints, 27 of 54, from a single 24-day pair and on provisional products. That number rests on 54 footprints and no confidence interval is available from the run, so it should be read as one pair over 54 buildings and not as a detection rate to plan around.

A per-footprint flag in Table 6.1 does not reliably belong to the footprint it is attached to. The 40 m toward-sensor offset is four pixels at the 10 m posting of these products, and the median Pereira footprint is 61 square metres, with 93 percent of footprints smaller than a single 20 m pixel and only 25 percent containing a pixel center. A team that walks to a flagged address and finds it intact should widen the search to the surrounding block rather than discard the product.

The rule suppresses false alarms in ground that decorrelates without damage, and we state that suppression against two baselines, which differ from each other by about five to six times rather than by an order of magnitude, 49.9 against 9.4 percent of unflagged buildings on Sentinel-1 and 53.3 against 9.1 percent on NISAR. Against the fixed threshold of the paper, sample coherence below 0.4 flags 49.9 percent of unflagged buildings on Sentinel-1 and 53.3 percent on NISAR, against 3.9 and 4.3 percent for the rule. Against a threshold lowered to 0.105 to match the any-pixel recall of the rule on the MS `ground\_truth` class, 9.8 percent of 102 footprints on Sentinel-1 and 14.7 percent of 102 on NISAR, the plain threshold flags 9.4 and 9.1 percent of unflagged buildings. The corresponding row of the comparison table reads 16.3 percent for NISAR because it divides by the 92 footprints of that class carrying at least one valid NISAR pixel rather than by all 102. We quote the summary file, which uses the same denominator for both sensors.

The matching fixes recall on one class only. Table 6.2 and Fig. 18 show what happens on the others.

Sensor and class	n	Rule, any pixel flagged	Sample coherence below 0.105, any pixel flagged
Sentinel-1, CEMS Destroyed	54	22.2 % (12)	33.3 % (18)
Sentinel-1, MS `ground_truth`	102	9.8 % (10)	9.8 % (10)
Sentinel-1, MS `confirmed`	146	7.5 % (11)	13.0 % (19)
Sentinel-1, MS `rejected`	360	2.5 % (9)	10.6 % (38)
Sentinel-1, unflagged background	90,378	3.9 %	9.4 %
NISAR, CEMS Destroyed	54	50.0 % (27)	44.4 % (24)
NISAR, MS `ground_truth`	92	16.3 % (15)	16.3 % (15)
NISAR, MS `confirmed`	137	15.3 % (21)	26.3 % (36)
NISAR, MS `rejected`	347	2.6 % (9)	8.9 % (31)
NISAR, unflagged	88,023	4.4 %	9.3 %

background			
------------	--	--	--

Table 6.2. The rule against a plain coherence threshold matched to it on one class, by sensor and label class. Source: catalogue/s1_screen/out_balz/fig15_rule_comparison.csv, columns "n" and "share_any", rules "Balz rule (coh_ml<0.4 & dev>0.15)" and "sample coh < 0.105 (equal any-pixel recall on ground_truth)". NISAR provisional. Every row of this table divides by the footprints of its class carrying at least one valid coherence pixel in that product, which is why the NISAR background share reads 4.4 percent here and 4.3 percent in the summary file cited in the preceding paragraph. The message is that recall is equal by construction only on the MS 'ground_truth' row: on Sentinel-1 the matched plain threshold detects more of every other graded class, and the rule's advantage lies in the background and 'rejected' rows. What the deviation term buys is background suppression, not detection.

What the rule gains in background suppression it loses in information, because it uses one pair and ignores the pre-event one. Table 6.3 gives the building-level separation of the 218 damaged footprints from intact ones for every single-pair quantity and for the pre-to-co coherence drop.

Quantity	ROC AUC against 'rejected'	ROC AUC against unflagged background
Co-event ML coherence, 3 by 3 window	0.544	0.562
Co-event sample coherence, 3 by 3 window	0.544	0.562
Co-event 7 by 7 sample coherence	0.572	0.589
Co-event deviation	0.600	0.560
Co-event change fraction under the rule	0.556	0.549
Pre-to-co drop, ML coherence	0.826	0.826
Pre-to-co drop, 7 by 7 sample coherence	0.863	0.881

Table 6.3. Building-level separation of damaged footprints by each single-pair quantity, against two negative sets. Source: catalogue/s1_screen/out_balz/s1/Pereira_auc_and_rules.json, key "auc", all seven entries; damaged n 218, 'rejected' n 360, background n 90,378 (keys "n_damaged", "n_rejected", "n_background"). The message is in the first two rows: at the estimator's own 3 by 3 window the learned coherence and the plain sample coherence are identical to three decimals, so on real Pereira footprints the learned estimator buys nothing over the sample estimate at equal window, although on the synthetic hold-out it lowers the root mean square error from 0.176 to 0.153. The 7 by 7 sample coherence scores higher than either because it is estimated on a larger window, not because the plain estimator beats the learned one.

Take-away: the single-pair rule is a conservative background filter with low recall, precise only relative to the fixed 0.4 threshold of the paper and not in the sense a surveyor needs, since on the Sentinel-1 co-event pair it flags 3,540 of the 91,266 footprints, of which 41, or 1.2 percent, carry a CEMS grade or an MS 'ground_truth' or 'confirmed' label; the pre-to-co drop it does not use separates damaged from intact footprints far better, and the sampling frames of section 8 are where a walkable shortlist is drawn.

6.2 Variable-combination experiments

Fig. 19 to Fig. 27 carry one panel per experiment and Fig. 28 ranks the ten headline sets on three metrics at once. The scoring grid and the audits that qualify it are two separate things and we describe them in turn. We scored feature sets on the same 368 positives, under both negative definitions, in a raw and an excess form, with logistic regression, a shallow gradient-boosted classifier and single-feature Otsu, CDF-0.84 and Youden thresholds, under five-fold cross-validation blocked on 500 m cells with seed 20260905; the ten headline sets of Table 6.4 are drawn from 1,296 scored configurations spanning nine experiments, 95 distinct feature sets, 26 model variants and three negative definitions, the third being an unflagged sample restricted to the blocks that contain a positive. Two audit lenses then qualify the ranking rather than extend it: a set of pre-event-only null feature sets, and a leakage audit that varies the fold scheme and the negative sampling.

The raw form uses the feature value; the excess form subtracts, per building, the median of the unflagged buildings within 200 m, so a neighborhood-wide signal cancels and only the building-specific part remains. Three design constants carry that protocol and we state what each buys. The 200 m radius is the one used by the neighborhood baseline test of section 5, so the two results are comparable, and at that radius 76,542 of the 91,266 footprints have at least five unflagged neighbors, so the local median is estimable across most of the city. The 500 m block is the cell of the identifier used for spatially blocked cross-validation, giving 1,172 blocks of which 941 contain a labelled building and 71 contain a positive, and the conclusion below is stable when it is varied. The unflagged negative set is ten times the positive count, drawn uniformly from a pool of 74,111 unflagged footprints with the same seed; ROC AUC does not depend on that ratio, so it affects only the precision-recall and Matthews correlation coefficient (MCC) columns, which we do not rank on.

Table 6.4 gives the ROC AUC of each headline set. Each cell is the best-scoring model for that set, and the model changes from cell to cell, so part of the spread between rows is a spread between models and the ordering is not a like-for-like ranking of feature sets. NISAR products are provisional.

Feature set	Raw, 'rejected'	Raw, unflagged	Excess, 'rejected'	Excess, unflagged	Model per cell, in column order
All features	0.898	0.916	0.881	0.895	logreg, xgboost, xgboost, xgboost
NISAR drop plus ShakeMap	0.880	0.889	0.853	0.876	xgb throughout
Sentinel-1 plus NISAR plus structure	0.872	0.894	0.798	0.818	logreg, xgb, logreg, logreg
All features without location	0.865	0.892	0.817	0.826	logreg, xgboost, logreg, logreg
Sentinel-1 plus NISAR fusion	0.843	0.860	0.705	0.764	logreg, xgb, logreg, logreg
ShakeMap only	0.841	0.844	0.841	0.844	xgb throughout
Sentinel-1 drop	0.810	0.801	0.556	0.603	logreg throughout
NISAR two-track drop	0.805	0.832	0.714	0.762	logreg throughout
Heights plus structure	0.726	0.709	0.726	0.709	logreg, xgb, logreg, xgb
Liu and Balz aggregates	0.669	0.657	0.654	0.673	xgb, xgb, lr, xgb

Table 6.4. ROC AUC of each headline feature set, in the raw and excess forms and against both negative sets. Source: catalogue/s1_screen/out_experiments/fig17_bar_values.csv, columns "roc_auc" and "model", by label, negatives and feature_form; the model labels are reproduced as that file writes them. The sets that carry location or ShakeMap keep their score in the excess form because those covariates have no excess form. The message of the table is that the four highest rows all carry location or ShakeMap, which the three tests below show is not damage information.

Three tests explain why that ranking is not a ranking of damage sensitivity.

The Sentinel-1 co-event drop loses most of its skill once the neighbourhood level is removed while the NISAR two-track drop does not (Fig. 19 and Fig. 20), and the Liu and Balz aggregates score at best 0.669 and score higher on the pre-event pair than on the co-event one (Fig. 21). ShakeMap is a location proxy inside one city (Fig. 22). Compared within a single negative set, the modified Mercalli intensity (MMI) falls from a ROC AUC of 0.811 over the whole labelled set to 0.591 within 2 km of the city center against 'rejected' negatives, and from 0.766 to 0.486 against unflagged negatives, and its rank correlation with distance to the center is negative, 0.580 and 0.532 in magnitude. The 2 km radius is the one that contains most of the labels, 241 of the 368 positives, so within it the geographic gradient is largely removed. A peak ground acceleration mask at 0.1 g is not a usable exposure filter either: it keeps 367 of the 368 positives and 341 of the 358 reviewer-rejected negatives, 95.3 percent of them, so it removes almost nothing from

either class. In the all-feature excess model the most important variable is distance to the city center, whose permutation removes 0.111 of ROC AUC, ahead of MMI at 0.032 and the NISAR T047A excess drop at 0.024.

Heights and structural covariates alone reach 0.726 and mostly re-introduce the dense-core confound when added to the drops (Fig. 23), distance to the city centre leads the all-feature excess model by permutation importance (Fig. 24), and a model built only from pre-event features nearly matches the co-event drops in the raw form (Fig. 26). Pre-event Sentinel-1 pairs reach 0.778 against 0.810 for the co-event Sentinel-1 drop and 0.807 for the co-event NISAR drop, so both sensors clear the null by about 0.03 and the raw null does not separate them; in the excess form the same three rows read 0.543, 0.556 and 0.711, so only the NISAR drop clears it. The Liu and Balz aggregates fail the same test by a wider margin: on the pre-event Sentinel-1 pair they score 0.694, above the 0.590 they score on the co-event pair, a null that beats the signal by 0.104.

The leakage audit separates the sensors under two of its three lenses (Fig. 27). Within 500 m blocks against 'rejected' negatives in the raw form, the Sentinel-1 drop scores 0.815 and the NISAR drop 0.805, which is no separation. On the same folds in the excess form, Sentinel-1 falls to 0.548 and NISAR holds 0.714. With negatives drawn inside the blocks that contain positives, so that a classifier cannot win by recognizing the dense core, Sentinel-1 falls to a range of 0.589 to 0.656 and NISAR to 0.736 to 0.782, each range taken over the two-feature and the single-feature version of that sensor's drop and over the two classifiers 'logreg' and 'xgb', four rows each. Blocked folds are why this is visible: the whole grid runs on 500 m blocked folds, and the random-fold figures quoted here are a diagnostic re-run inside the audit, under which 68 to 97 percent of test buildings sit within 100 m of a training building against 11 to 26 percent under 500 m blocks. The block size is not what carries the result: on 'rejected' negatives in the excess form the NISAR drop reads 0.714, 0.728, 0.715 and 0.709 at 500 m, 1,000 m and 2,000 m blocks and under unshuffled grouped folds at 500 m, while the Sentinel-1 drop reads 0.548, 0.488, 0.537 and 0.550.

Two Sentinel-1 excess numbers in this section look like the same configuration and are not. Table 6.4 gives 0.556 for a set holding the co-event drop alone, and the audit gives 0.548 for a set holding that drop together with its pre-event z-score; both use logistic regression, 500 m blocks and the same 368 positives against 355 negatives, and the audit's one-feature equivalent reads 0.560. The two-track NISAR pair is identical in both files at 0.714 because the two sets contain the same two features.

Fusion is worth little (Fig. 25). Sentinel-1 plus NISAR beats the best single sensor by 0.020 to 0.043 of ROC AUC in the raw form across both classifiers and both negative sets, and in the excess form the gain runs from a loss of 0.010 to a gain of 0.014.

Take-away: a feature set scoring 0.916 on raw values in one city is largely reporting where the labels are, since removing location costs it 0.024 and removing the neighborhood level costs the Sentinel-1 drop 0.254 of ROC AUC, and the only layer that keeps its skill once the neighborhood and the pre-event level are removed is the NISAR coherence drop, on provisional products and a single run.

6.3 What this section asks the field team to do

Calibrate the NISAR L-band coherence drop first, on
'catalogue/s1_screen/out_nisar/Pereira_Dosquebradas_T047A_f004_buildings_dpm.gpkg' and its
neighborhood-corrected column 'nisar_at_T047A_dpm_excess' in
'catalogue/s1_screen/out_experiments/pereira_features.gpkg', because it is the only layer whose skill
survives the null, within-block and excess tests. Treat the Sentinel-1 drop in
'catalogue/s1_screen/out20/final/Pereira_Dosquebradas_cells250.csv' as a block-level map for locating the
affected core and not as a per-building predictor. Do not use ShakeMap, distance to the city center or the

fixed Liu and Balz rule as damage predictors in Pereira. The per-city sampling frames that implement these three instructions, including the whole 250 m cells with unflagged controls, are in section 8.

Figures

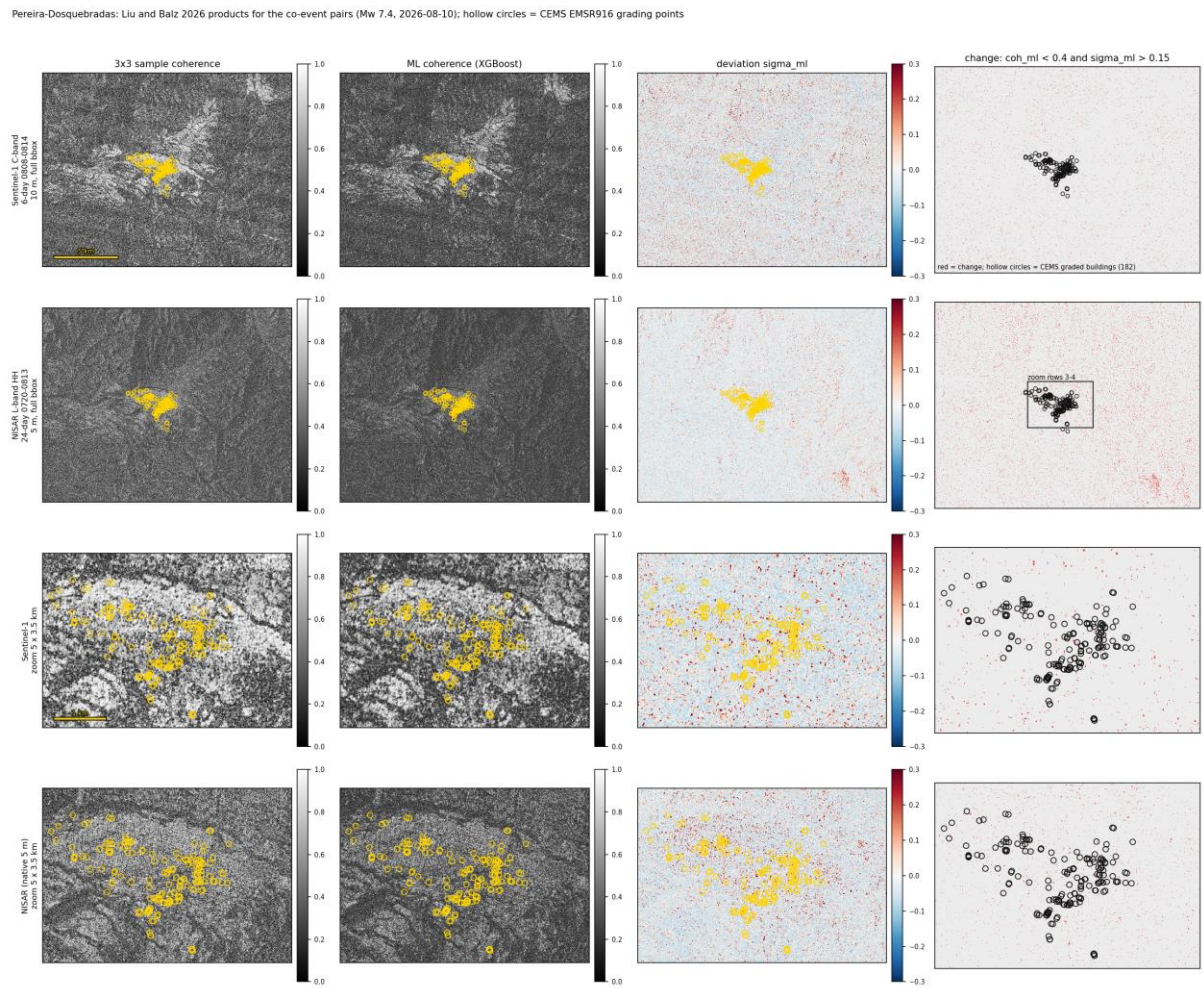


Fig. 17. ML coherence, deviation and change masks of the rule over Pereira on the Sentinel-1 and NISAR co-event pairs; the rule flags 3.9 and 4.3 percent of unflagged buildings where a fixed 0.4 threshold flags 49.9 and 53.3 percent. The masks carry the 40 m toward-sensor offset of section 5, so a flag locates a block and not a building. NISAR provisional. Providers: Copernicus Sentinel-1, NISAR via the Alaska Satellite Facility (ASF), Overture. Source: catalogue/s1_screen/out_balz/fig14_15_summary.json.

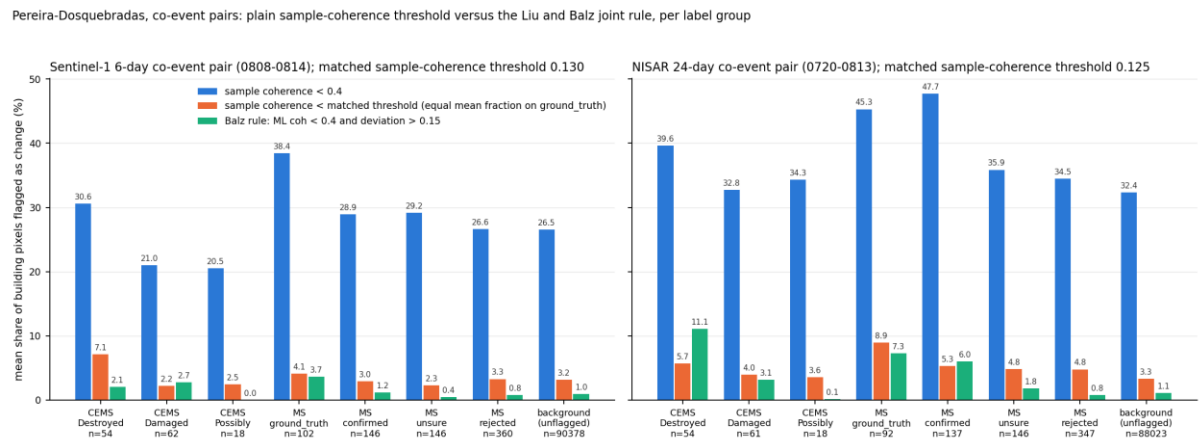


Fig. 18. the rule flags about 40 percent of the unflagged buildings that a threshold at equal recall flags, 3.9 against 9.4 percent on Sentinel-1, but misses most graded buildings and detects fewer of them than that threshold on every graded class except the one used to match it. The same 40 m offset applies. NISAR provisional. Providers: Sentinel-1, NISAR, Microsoft AI for Good, CEMS. Source: catalogue/s1_screen/out_balz/fig15_rule_comparison.csv.

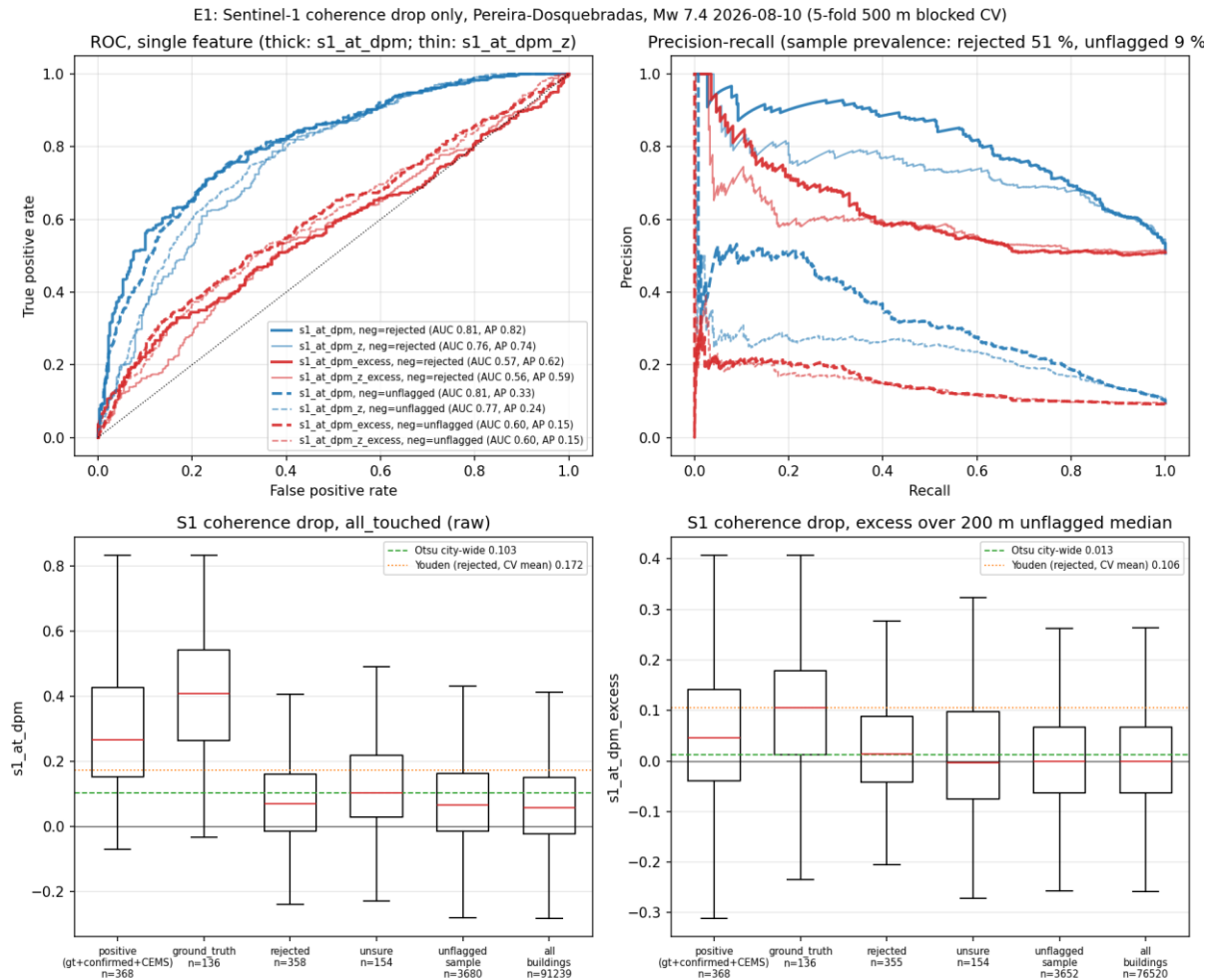


Fig. 19. the Sentinel-1 co-event drop reaches a ROC AUC of 0.810 on raw values and 0.556 once the 200 m neighborhood level is removed, so its raw skill is neighborhood skill. Providers: Copernicus Sentinel-1, Microsoft AI for Good, CEMS, Overture. Source: catalogue/s1_screen/out_experiments/E1_results.csv.

Fig. 16 E2: NISAR L-band coherence drop only, Pereira-Dosquebradas (Mw 7.4, 2026-08-10). 5-fold 500 m spatially blocked CV, pooled out-of-fold scores

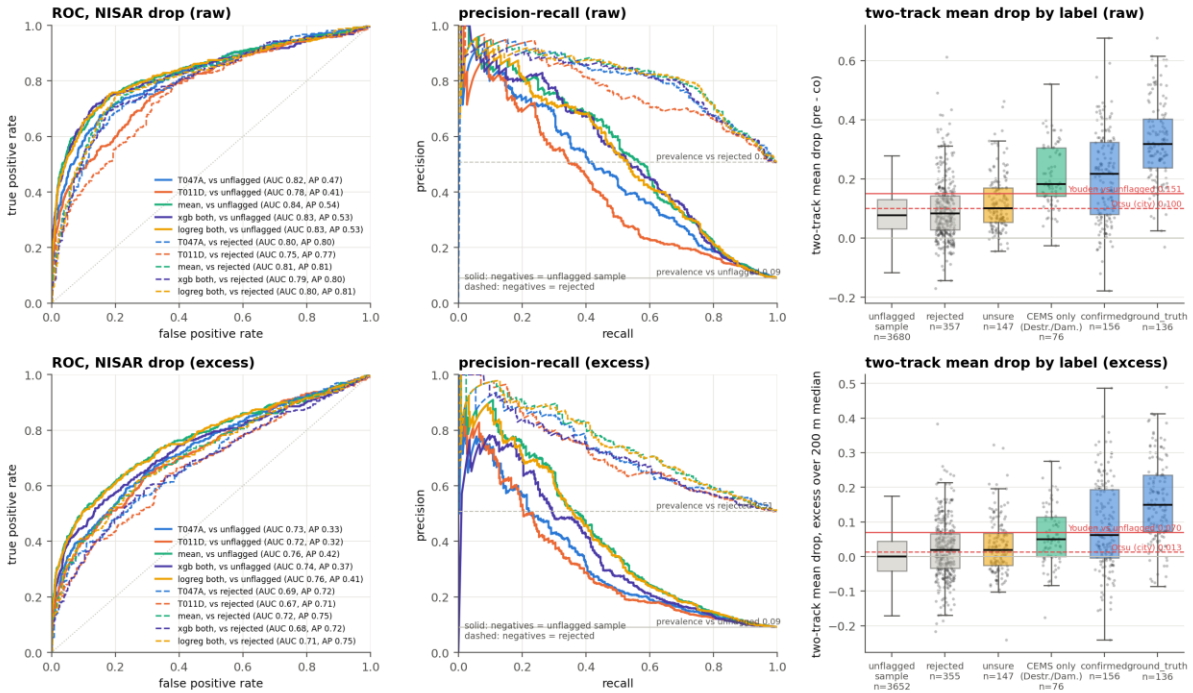


Fig. 20. the NISAR two-track drop reaches 0.805 raw and holds 0.714 in the excess form, the only single-sensor layer that does. NISAR provisional. Providers as above plus NISAR via ASF. Source: catalogue/s1_screen/out_experiments/E2_results.csv.

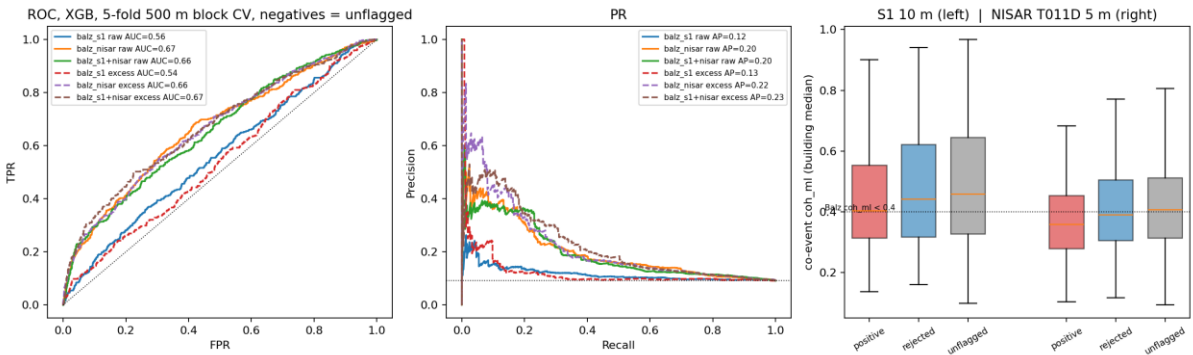


Fig. 21. the Liu and Balz aggregates score 0.669 at best and score higher on the pre-event pair, 0.694, than on the co-event pair, 0.590. NISAR provisional. Providers as above. Source: catalogue/s1_screen/out_experiments/E3_results.csv.

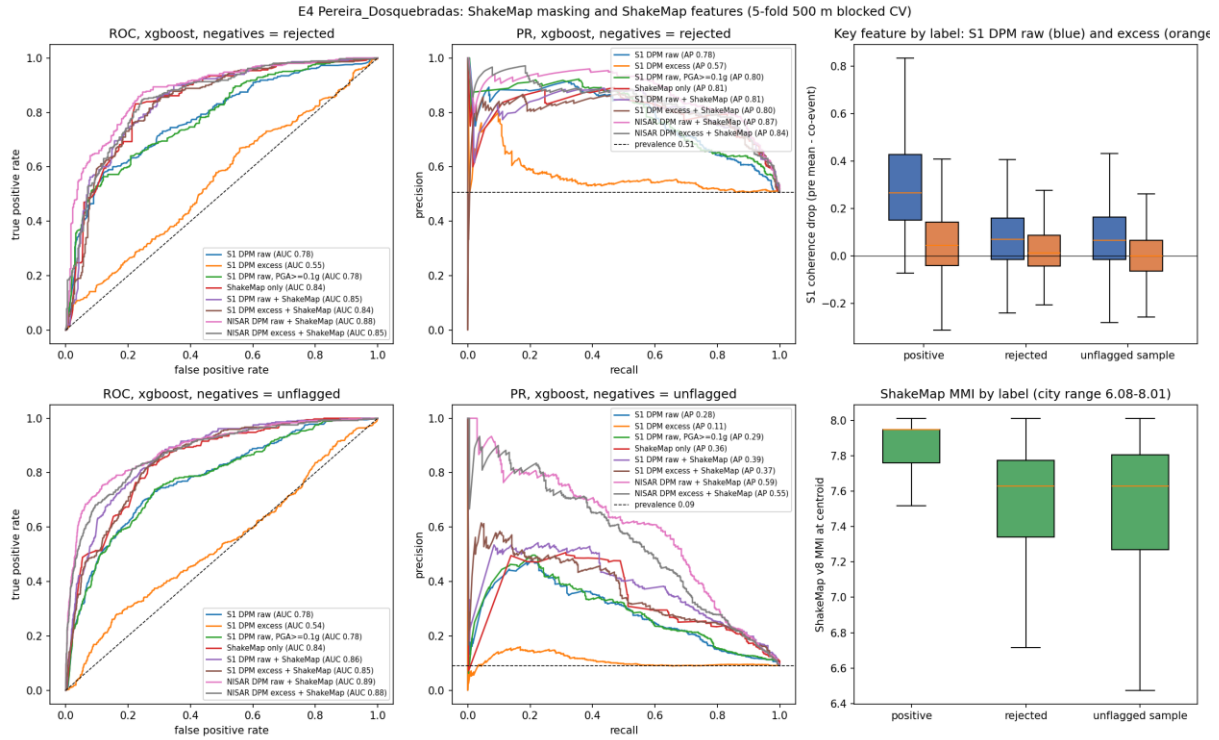


Fig. 22. ShakeMap alone scores 0.841 over the whole labelled set but 0.591 within 2 km of the city center, so it is a location proxy inside one city. Providers as above plus United States Geological Survey (USGS) ShakeMap. Source: catalogue/s1_screen/out_experiments/E4_results.csv.

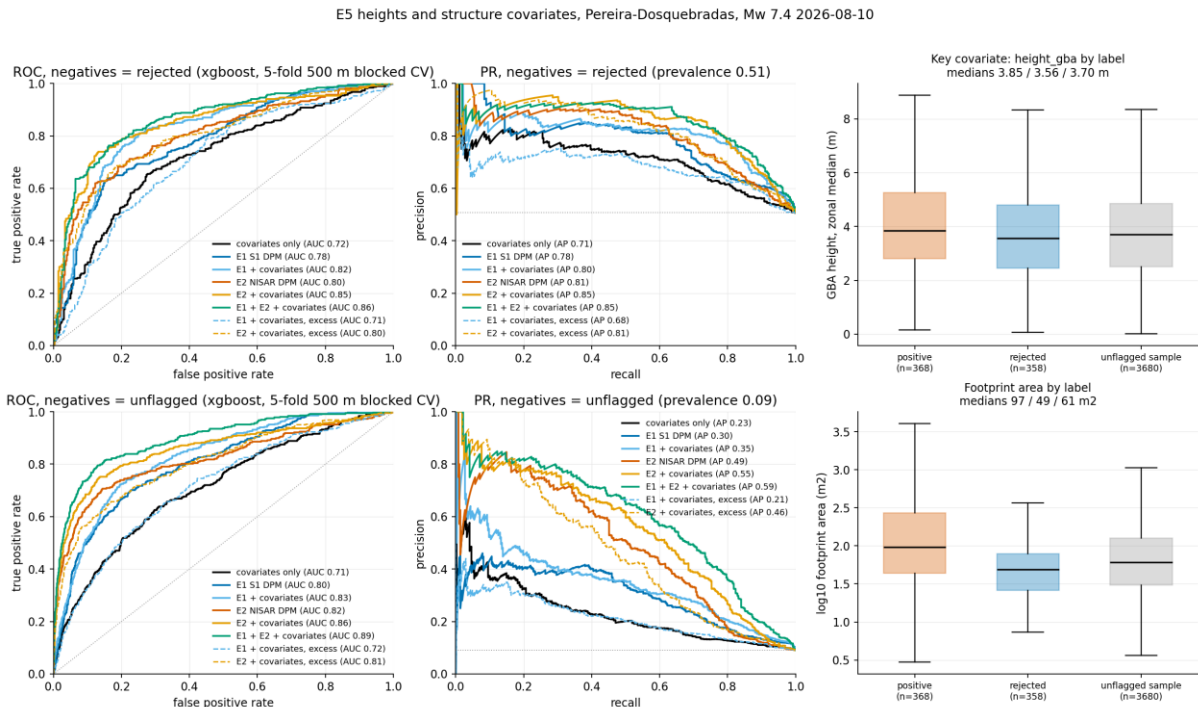


Fig. 23. heights and structural covariates alone score 0.726, and added to the SAR drops they mostly re-introduce the dense-core confound. Providers as above plus Technical University of Munich (TUM) GlobalBuildingAtlas. Source: catalogue/s1_screen/out_experiments/E5_results.csv.

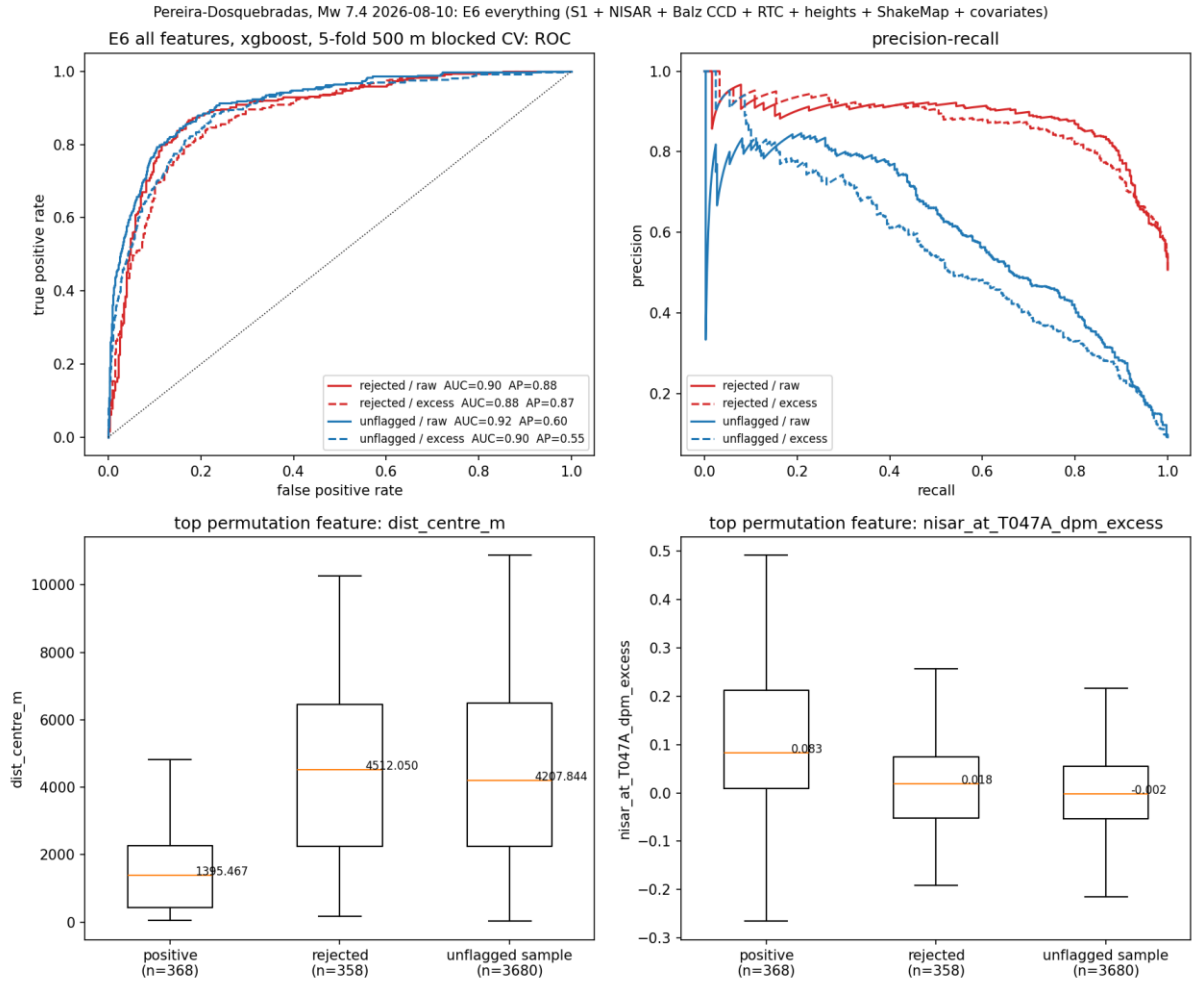


Fig. 24. in the all-feature excess model the leading variable by permutation importance is distance to the city center at 0.111 of ROC AUC, ahead of MMI at 0.032. NISAR provisional. Providers as above. Source: catalogue/s1_screen/out_experiments/E6_results.csv and E6_permutation_importance.csv.

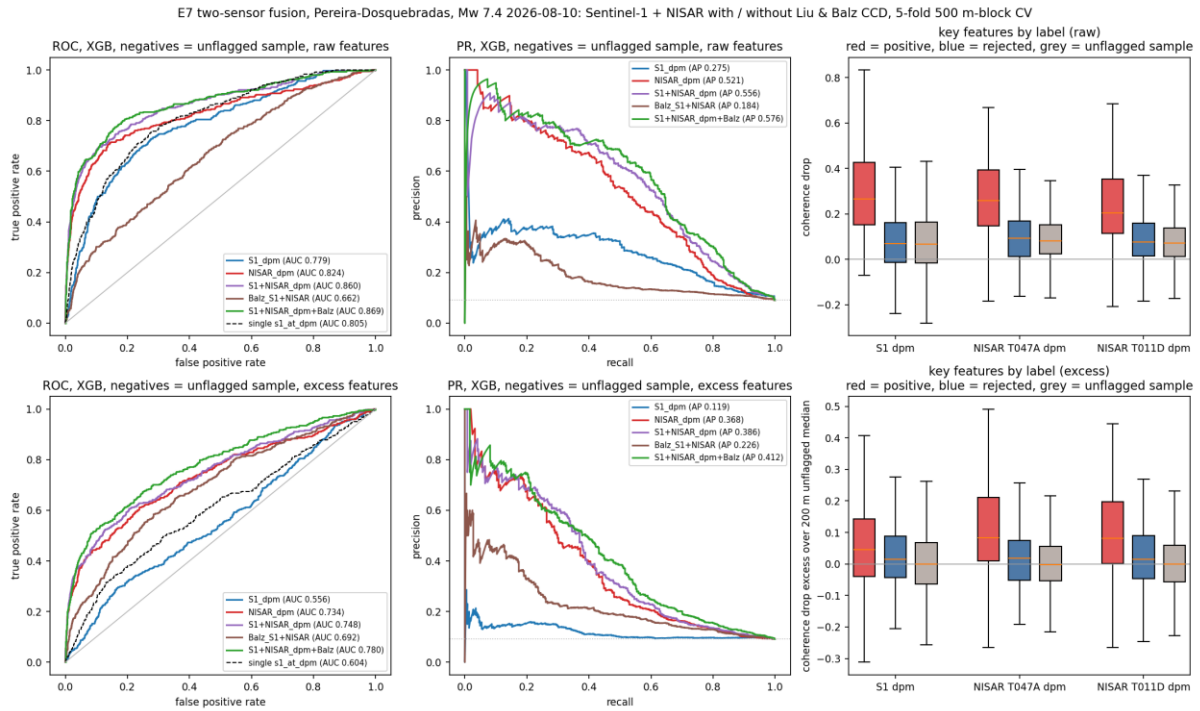


Fig. 25. fusing Sentinel-1 and NISAR adds 0.020 to 0.043 of ROC AUC over the best single sensor on raw values and nothing after the excess correction. NISAR provisional. Providers as above. Source: catalogue/s1_screen/out_experiments/E7_results.csv.

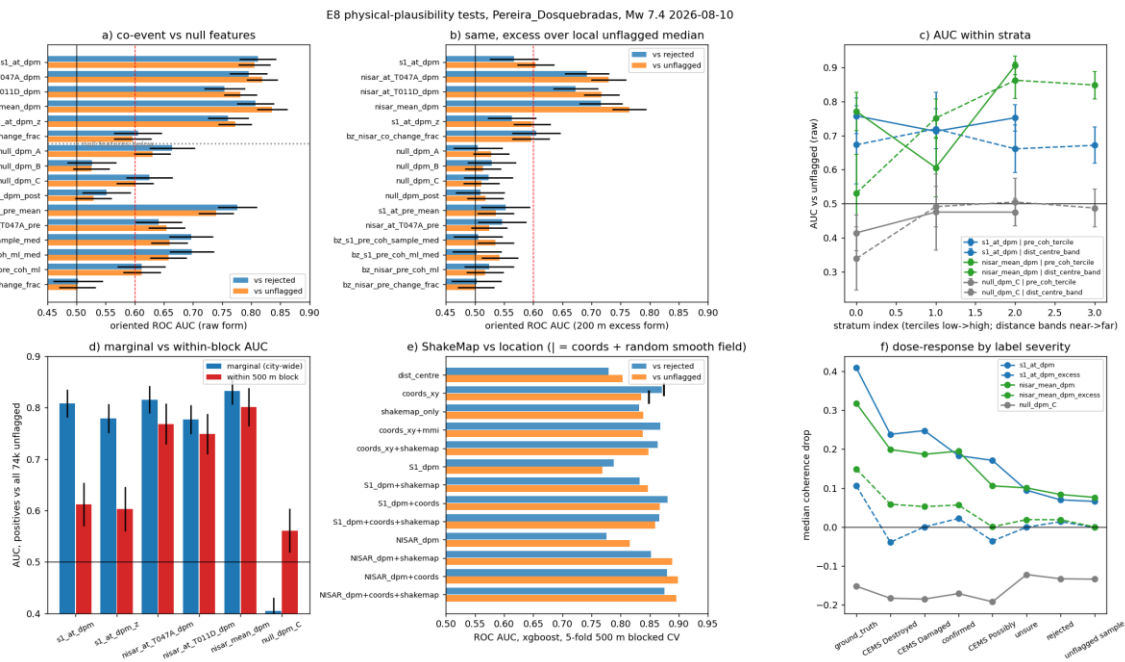


Fig. 26. pre-event-only null feature sets reach 0.778 against 0.810 for the co-event Sentinel-1 drop, so most of the raw Sentinel-1 skill is present before the earthquake. Providers as above. Source: catalogue/s1_screen/out_experiments/E8_results.csv.

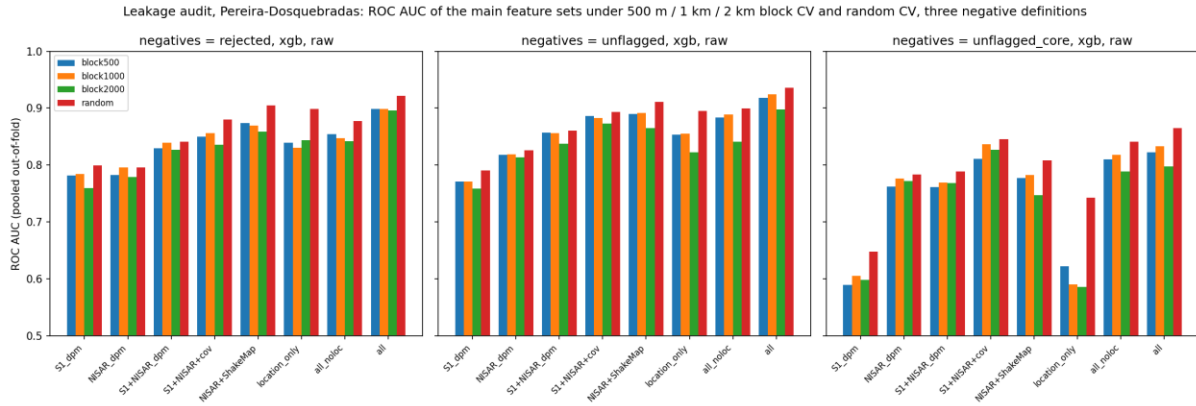


Fig. 27. with negatives inside positive-bearing blocks the Sentinel-1 drop falls to 0.59 to 0.66 while NISAR holds 0.74 to 0.78, so the Sentinel-1 skill on random negatives was spatial. NISAR provisional. Providers as above. Source: catalogue/s1_screen/out_experiments/leakage_audit/audit_results.csv.

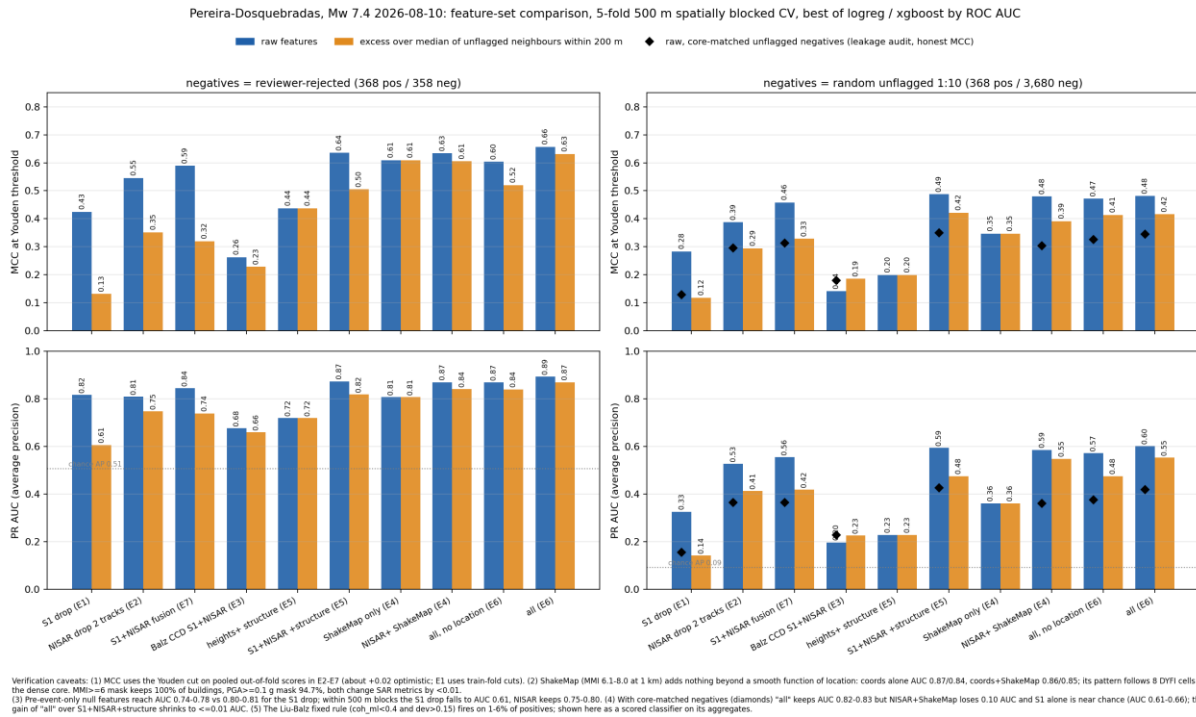


Fig. 28. ten feature sets ranked by MCC, precision-recall AUC and ROC AUC; only the NISAR drop keeps its ROC AUC in the excess form, 0.714 against 0.556 for Sentinel-1, while ShakeMap and all-feature sets keep theirs because they carry location. NISAR provisional. Source: catalogue/s1_screen/out_experiments/fig17_bar_values.csv.

Sources: 'PROCESSING_LOG.md', 'SAR_SCREEN_RESULTS.md', 'catalogue/s1_screen/balz_ccd', 'catalogue/s1_screen/balz_ccd/README.md', 'catalogue/s1_screen/out20/final/Pereira_Dosquebradas_cells250.csv', 'catalogue/s1_screen/out20/final/fig12_pereira_label_test_stats.json', 'catalogue/s1_screen/out_balz/fig14_15_balz.py', 'catalogue/s1_screen/out_balz/fig14_15_summary.json', 'catalogue/s1_screen/out_balz/fig15_rule_comparison.csv', 'catalogue/s1_screen/out_balz/s1/Pereira_auc_and_rules.json', 'catalogue/s1_screen/out_balz/s1/Pereira_buildings_balz.gpkg', 'catalogue/s1_screen/out_balz/s1/Pereira_label_table.csv',

```
`catalogue/s1_screen/out_experiments/E1_results.csv`,
`catalogue/s1_screen/out_experiments/E2_results.csv`,
`catalogue/s1_screen/out_experiments/E3_results.csv`,
`catalogue/s1_screen/out_experiments/E4_results.csv`,
`catalogue/s1_screen/out_experiments/E4_shakemap_confound.csv`,
`catalogue/s1_screen/out_experiments/E5_results.csv`,
`catalogue/s1_screen/out_experiments/E6_permutation_importance.csv`,
`catalogue/s1_screen/out_experiments/E6_results.csv`,
`catalogue/s1_screen/out_experiments/E7_fusion_vs_single.csv`,
`catalogue/s1_screen/out_experiments/E7_results.csv`,
`catalogue/s1_screen/out_experiments/E8_results.csv`,
`catalogue/s1_screen/out_experiments/fig17_bar_values.csv`,
`catalogue/s1_screen/out_experiments/leakage_audit/audit_leakage.py`,
`catalogue/s1_screen/out_experiments/leakage_audit/audit_results.csv`,
`catalogue/s1_screen/out_experiments/pereira_features.gpkg`,
`catalogue/s1_screen/out_experiments/pereira_features_README.md`,
`catalogue/s1_screen/out_experiments/pereira_features_summary.json`,
`catalogue/s1_screen/out_experiments/summary_table.csv`,
`catalogue/s1_screen/out_nisar/Pereira_Dosquebradas_T047A_f004_buildings_dpm.gpkg`,
`figures/fig14_balz_pereira_maps.png`, `figures/fig15_balz_label_test.png`, `figures/fig16_E1.png`,
`figures/fig16_E2.png`, `figures/fig16_E3.png`, `figures/fig16_E4.png`, `figures/fig16_E5.png`,
`figures/fig16_E6.png`, `figures/fig16_E7.png`, `figures/fig16_E8.png`, `figures/fig16_leakage_audit.png`,
`figures/fig17_feature_set_comparison.png`.
```

7. Building heights and how they enter prioritisation

Height is the one building attribute we can attach to a large share of the footprints in some cities and to almost none in others. Two sources of height are used: the GlobalBuildingAtlas (GBA), a global level of detail 1 (LoD1) building model that carries one height per polygon, and the floor counts already present on the open Overture building footprints, converted at 3.0 m per floor where no metric height exists. Buildings are binned into four classes at 6, 12 and 24 m.

Those four numbers are the pipeline's constants and not a design we optimised, so we state them plainly rather than justify them. At the pipeline's own 3.0 m per floor they correspond to two, four and eight storeys. The cut is right closed, which places a building of exactly 24.0 m in the 12 to 24 m class and a building of exactly 6.0 m in the below 6 m class. Subsection 7.4 shows what those two conventions do to the class the field team is being asked to visit.

7.1 Where a height attribute exists

Coverage splits the seven cities into two groups. GBA LoD1 heights cover 90.8 percent of the 91,266 Pereira and Dosquebradas footprints, 96.1 percent of the 158,489 Cali footprints and 73.8 percent of the 7,707 Quimbaya footprints, and three buildings in Armenia and none in Quibdó, Manizales or Sevilla. No GBA LoD1 height estimate exists for the polygons of those four cities, which matches the absence of the corresponding GBA height raster tiles: Quibdó, Armenia, Manizales and Sevilla each need one tile and have none, and Quimbaya has one of the two it needs. The LoD1 polygon files themselves were streamed for all seven cities, so this is a missing height estimate and not a failed match. Where GBA is missing, only Overture floor counts remain, and they are sparse: any height source at all reaches 50.4 percent in Quibdó, 4.1 percent in Armenia, 1.9 percent in Manizales and 0.2 percent in Sevilla. Every GBA polygon used here is derived from a Google Open Buildings footprint, so the height and the footprint outline are not independent of that product.

City	Overture footprints	GBA height	Any height	GBA tiles available
Pereira and Dosquebradas	91,266	90.8 %	92.5 %	2 of 2
Cali	158,489	96.1 %	96.3 %	2 of 2
Quimbaya	7,707	73.8 %	73.8 %	1 of 2
Quibdó	34,066	0 %	50.4 %	0 of 1
Armenia	106,040	< 0.1 % (3 buildings)	4.1 %	0 of 1
Manizales	67,014	0 %	1.9 %	0 of 1
Sevilla	6,726	0 %	0.2 %	0 of 1

Source: `catalogue/s1\_screen/out\_heights/height\_coverage.csv` and `gba\_height\_raster\_coverage.csv`. Per building files: `catalogue/s1\_screen/out\_heights/buildings\_heights\_<City>.gpkg`, one per city.

The Quimbaya gap is locatable on the ground, which matters because the field team will be standing in it. The city bounding box spans longitude 75.80 W to 75.72 W and latitude 4.58 N to 4.66 N. The one available tile covers latitude 4.6 N to 4.8 N and longitude 75.8 W to 75.6 W, so it contains the city in longitude and leaves the southern strip from 4.58 N to 4.60 N uncovered, a quarter of the latitude span. A quarter of the latitude span with no height attribute is what 73.8 percent coverage implies, and the two agree.

Take-away: a height attribute is usable on the tablet in Pereira and Dosquebradas, Cali and the northern three quarters of Quimbaya above latitude 4.60 N only, and the four remaining cities, together with the southern quarter of Quimbaya, need storeys recorded by hand in the field.

7.2 How tall the stock is

Where coverage is good the stock is low rise. The median height is 5.0 m in Pereira and Dosquebradas, with a 90th percentile of 7.7 m and 0.9 percent of buildings above 12 m, and 5.1 m in Cali, with a 90th percentile of 7.3 m and 0.7 percent above 12 m; Quimbaya is lower still at a 3.6 m median.

The figures for the four cities without GBA are not city statistics and must not be quoted as such. Manizales reports 40.8 percent of buildings above 12 m from 1,262 tagged buildings out of 67,014, Sevilla 21.4 percent from 14 out of 6,726, Armenia 17.3 percent from 4,337 out of 106,040 and Quibdó 0.16 percent from 17,174 out of 34,066. Each of the four is a statistic about whichever buildings happen to carry an Overture floor count, and in three of the four that is under 5 percent of the city.

The two height sources also disagree where both exist. On the 20,856 Pereira and Dosquebradas footprints carrying a GBA height and an Overture height estimate, the GBA height exceeds the Overture estimate by a median of 1.59 m with a median absolute deviation of 1.64 m and a correlation of 0.18; in Cali the offset is 0.56 m on 3,413 buildings at a correlation of 0.45, with the larger dispersion of the two, a median absolute deviation of 2.43 m. The Overture estimate is the metric height where one exists and 3.0 m times the floor count otherwise, and metric heights are rare, 70 of the Pereira and Dosquebradas footprints and 753 in Cali, so the comparison is in practice a comparison against floor counts.

The disagreement has a shape, and the shape is compression. Grouping the Pereira and Dosquebradas buildings by their Overture floor count, the median GBA height is 5.80 m at one floor, 6.02 m at two, 6.10 m at three, 6.48 m at four and 6.86 m at five, on 11,118, 7,441, 1,660, 318 and 182 buildings; at eight floors it is 9.11 m on 20 buildings. GBA moves 1.06 m across the first five storeys of the Overture attribute. Whatever the true heights are, one of the two sources is not resolving them.

Take-away: treat the height classes as relative bands and not as measurements, because the two sources that populate them differ by 1.6 m in the median, disagree at a correlation of 0.18, and move 1.06 m apart over five storeys of the same buildings.

7.3 The height stratified drop in Pereira and Dosquebradas

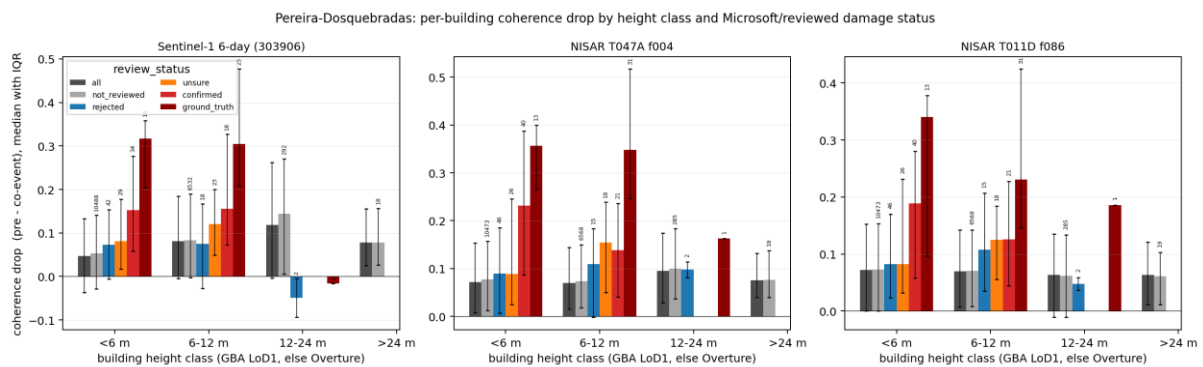


Fig. 29. median coherence drop with interquartile range by height class and by damage label in Pereira and Dosquebradas, for Sentinel-1 6 day and for tracks T047A and T011D of the joint National Aeronautics and Space Administration (NASA) and Indian Space Research Organisation (ISRO) mission, the NASA ISRO Synthetic Aperture Radar (NISAR), showing that the raw drop of the labelled classes stands above the population in the two classes below 12 m, that exactly one MS 'ground_truth' building falls between 12 and 24 m, and that no labelled or Copernicus Emergency Management Service (CEMS) graded building at all lies above 24 m. The three panels do not share a vertical axis, so the panels may be read within themselves but not against each other. The three CEMS grades are not drawn, although the text reports them, because the plotting script fixes the six Microsoft statuses. Data providers: European Space Agency (ESA) Copernicus Sentinel-1; NASA, Jet Propulsion Laboratory (JPL) and ISRO NISAR provisional products via the Alaska Satellite Facility (ASF); Technical University of Munich (TUM) GlobalBuildingAtlas; Overture Maps Foundation; Microsoft artificial intelligence (AI) for Good Lab; Copernicus Emergency Management Service.

What the subsection covers, and what it leaves out. The analysis runs on the 21,197 Pereira and Dosquebradas footprints that carry both a height and a Sentinel-1 value, which is 23.2 percent of the 91,266 footprints in the city; 84,395 carry a height and 22,731 carry a Sentinel-1 value. The labelled buildings are cut in the same proportion: 40 of the 134 MS 'ground_truth' buildings in the city, 52 of the 155 MS 'confirmed', 52 of 149 MS 'unsure' and 62 of 359 MS 'rejected'. Those four city totals are the height table's own join and they are not the product counts. The released Microsoft layer carries 136, 156, 154 and 367 in the same four classes, 813 reviewed footprints in all; matching them onto the Overture base used throughout section 7 keeps 795 of them, two of which match two Overture footprints each, which leaves the 797 labelled rows the four counts above are drawn from. The restriction is not spatial. Burst 303906 covers 100 percent of the Pereira and Dosquebradas bounding box, and the buildings that drop out are the small ones, because a footprint receives a value only when it contains the centre of a 20 m pixel: 93 percent of Pereira footprints are smaller than one pixel, the median footprint with a value is 183 square metres against 48 square metres without. Every statement below is therefore about the larger quarter of the building stock.

The geolocation offset, and what it does to a per building number. All 20 m Sentinel-1 screen products are displaced about 40 m toward the sensor, measured as a best footprint shift of 40 m east and 5 m south on this burst, because the orthometric heights of the Copernicus 30 m global digital elevation model (DEM) were used as ellipsoidal heights in our geocoding step. In the 21,197 building subset the median Overture footprint area is 188.5 square metres, a side of about 13.7 m. The shift is therefore about three building widths, and a Sentinel-1 value attached here to one footprint may have come from a neighbouring one. Section 5 states the same fact at the scale where it does not bite: two pixels lie inside the 140 m coherence window, so city and neighbourhood statistics are unaffected while per building work is not. What survives is every Sentinel-1 statement made at class level, where thousands of buildings average the

displacement away, and nothing that names an individual Sentinel-1 building. The NISAR drops are read from the provider's geocoded provisional products and do not pass through our geocoding step, so they do not carry this offset; no equivalent shift test has been run on them.

The population. Across the subset the population median drop rises with height, from 0.047 below 6 m to 0.081 between 6 and 12 m, 0.118 between 12 and 24 m and 0.078 above 24 m. NISAR track T047A is flat over the same classes at 0.072, 0.070, 0.095 and 0.076, and track T011D falls, 0.072, 0.070, 0.064 and 0.063. The three sensors are summarised on slightly different building sets, 21,197 for Sentinel-1 and 21,163 for each NISAR track, because a footprint may carry one sensor and not the other.

The labelled classes. The Microsoft AI for Good Lab Pereira product carries four human reviewer flags, which sections 4, 5 and 6 name by their own file labels, MS `ground\_truth`, MS `confirmed`, MS `unsure` and MS `rejected`, and which we name the same way here. All four are assigned from imagery at a desk and none is field verified, and the released product does not define what separates the first two. CEMS grades the same city independently into Destroyed, Damaged and Possibly damaged. Fig. 29 draws all three sensors by height class and label class at once. In the two classes below 12 m, where every labelled building except one lives, the raw Sentinel-1 6 day medians are:

Damage label	Below 6 m	6 to 12 m
population	0.047 (n 13,187)	0.081 (n 7,670)
MS `ground_truth`	0.317 (n 14)	0.305 (n 25)
MS `confirmed`	0.153 (n 34)	0.156 (n 18)
CEMS Damaged	0.189 (n 7)	0.256 (n 20)
CEMS Destroyed	0.171 (n 10)	0.223 (n 22)
CEMS Possibly damaged	0.177 (n 3)	0.193 (n 5)
MS `unsure`	0.081 (n 29)	0.121 (n 23)

Table 7.1. Median Sentinel-1 6 day coherence drop by damage label in the two height classes below 12 m, Pereira and Dosquebradas. Source: `catalogue/s1\_screen/out\_heights/pereira\_drop\_by\_height.csv`, sensor `S1\_6day`. The message of the table is that the separation depends on which label is used: the six labelled classes sit between 1.5 and 6.8 times the population median, and the reviewer unsure class sits at 1.7 and 1.5 times it.

What survives the neighbourhood baseline. Section 5 established that most of the raw Sentinel-1 gap in this city is a neighbourhood scale core drop and not a per building signal. On the same Sentinel-1 layer used here, taking the excess over the median of unflagged buildings within 200 m removes 73 percent of the MS `ground\_truth` gap and 82 percent of the MS `confirmed` gap, leaving +0.068 and +0.020; on NISAR track T047A the same correction leaves +0.135 for MS `ground\_truth` and +0.048 for MS `confirmed`, and track T011D leaves +0.138. Height does not change that verdict. Grouping the 368 positives of section 6, which are the MS `ground\_truth` and MS `confirmed` buildings together with the CEMS Destroyed and Damaged points, by height class, the Sentinel-1 excess is +0.019 on 58 buildings below 6 m and +0.026 on 73 buildings between 6 and 12 m, against a NISAR T047A excess of +0.092 on 60 and +0.087 on 82, and a T011D excess of +0.056 and +0.083 on the same counts. Over the whole population the same excesses are within 0.014 of zero in every class and on every sensor, as the construction requires.

One city. This is a Pereira and Dosquebradas result, and the field programme covers seven cities. We did not repeat it in Cali, which has the better height coverage of the two, 96.1 percent against 90.8 percent, and which has both Sentinel-1 per building products and labels: 256 Microsoft flagged footprints, which are the 256 of the released layer's 266 Cali flags that join the delivered building layer of section 5, of which 80 carry a Sentinel-1 value, and 8 of 21 CEMS graded points with a value. The field team should not assume the Pereira relation transfers. In Cali the same Microsoft flags separate only marginally on the raw drop, 0.135 against 0.125 for unflagged buildings, and a model trained on all Pereira labels and applied to Cali performs at chance, with an area under the curve of 0.458 to 0.527 across tiers and classifiers. Running the

same group-by on Cali is a change of input file and of label source, and it is the cheapest way to find out before the mission.

Take-away: once the 200 m neighbourhood is removed, the height stratified separation of damaged buildings from the population is carried by NISAR, at +0.09 in both classes below 12 m, and not by Sentinel-1, which retains +0.02 to +0.03 and whose raw class separation is a neighbourhood effect; and the whole relation has been tested in one city only.

7.4 The tall classes nobody has checked

Five labelled buildings lie between 12 and 24 m and none at all lies above 24 m, so the tall classes are the part of the stock where the radar layers are unchecked. Three of the five carry a damage grade, one from the Microsoft desk review and two from CEMS, and two are reviewer rejected; no building carries both a Microsoft flag and a CEMS grade. The five disagree with each other, and the direction of the disagreement depends on the sensor:

Group in the 12 to 24 m class	Sentinel-1 6 day	NISAR T047A	NISAR T011D
MS 'ground_truth' (n 1)	-0.017	0.163	0.186
MS 'rejected' (n 2)	-0.050	0.097	0.048
CEMS Damaged (n 2)	0.243	0.157	0.098
population (n 321, 317 on NISAR)	0.118	0.095	0.064

Table 7.2. Median coherence drop of every labelled group present in the 12 to 24 m class, Pereira and Dosquebradas. Source: `catalogue/sl\_screen/out\_heights/pereira\_drop\_by\_height.csv`, height_class `12-24 m`. No row of any labelled or CEMS graded status exists in the `>24 m` class for any of the three sensors.

The one desk graded damaged building in the class has a Sentinel-1 drop of -0.017, below the population and below zero, and sits above the population on both NISAR tracks; the two CEMS Damaged points sit at 0.243 on Sentinel-1, with one of the two above a 0.2 drop, and near the population on NISAR; the two reviewer rejected buildings are negative on Sentinel-1. Five points, pointing two ways on Sentinel-1 and consistently upward on NISAR, are the whole of the labelled evidence about tall buildings in this earthquake.

Against that, the 12 to 24 m class carries the highest share of buildings above a 0.2 Sentinel-1 drop of any class, 36.4 percent against 13.2 percent below 6 m. Four explanations compete with damage and none of them has been excluded. Layover, shadow and the mixture of roof and facade returns inside a 20 m cell all produce a drop over a tall building. So does footprint size: the median Overture footprint area rises from 157 square metres below 6 m to 263, 453 and 1,186 square metres across the four classes, so the tall classes are also the large footprint classes, which is the confound that subsection 7.5 has to remove from the structure covariates as well.

The class is also partly an artefact of the two constants stated at the head of this section. In the 21,197 building subset the 12 to 24 m class is 307 GBA heights and 14 heights derived from floor counts, and the above 24 m class is 13 GBA and 6 from floor counts. An eight floor Overture building with no GBA height is assigned exactly 24.0 m by the 3.0 m conversion and lands in the 12 to 24 m class on the closed boundary; 7 of the 27 floor derived buildings in that class citywide sit at exactly 24.0 m, and 3 of the 14 inside the Sentinel-1 subset do. The same eight floor buildings measured by GBA have a median height of 9.11 m, a factor of 2.6 below the converted value, and both numbers feed the one height class column.

Take-away: the elevated Sentinel-1 drop in the 12 to 24 m class is unverified, is contradicted by the one MS 'ground_truth' building in it, is supported by the two CEMS Damaged points, competes with layover, shadow and a median footprint area nearly three times that of the class below 6 m, and rests on a class 14

of whose 321 members in the Sentinel-1 subset take their height from a floor count, so the tall classes are a visit list and not a result.

7.5 Height as a stratification variable, not a damage predictor

Height on its own carries almost no information about damage in Pereira and Dosquebradas. The GBA height as a single feature reaches a receiver operating characteristic (ROC) area under the curve (AUC) of 0.537 against reviewer rejected negatives and 0.540 against unflagged negatives under logistic regression, where 0.5 is chance; the evaluation sets are 726 buildings with 368 positive and 4,048 buildings with 368 positive.

The set that reaches 0.726 and 0.709 in Fig. 28 is not height plus footprint area and neighbour density. It has five members, 'height_gba', 'height_overture', 'num_floors', 'area_m2' and 'n_neighbours_50m', and the two extra members matter. The two paired values also come from two different models: 0.726 is logistic regression against reviewer rejected negatives and 0.709 is shallow xgboost against unflagged negatives, because Fig. 28 reports the better of the two models per row.

The presence of the floor count is label correlated, which is the reason to look at the ablation rather than at the headline. Of the 368 positives, 48.9 percent carry an Overture floor count, against 23.2 percent of the 358 reviewer rejected negatives and 23.3 percent of the 74,111 unflagged buildings. A tree model reads the presence of an attribute as a feature, so a covariate set containing the floor count is partly reading which buildings a human chose to attribute.

Feature set	Rejected negatives, logreg	Unflagged negatives, logreg	Unflagged negatives, xgb
area and neighbour count only	0.728	0.689	0.662
those two plus the three height features	0.726	0.707	0.709
the same without the floor count	0.709	0.671	0.669
floor count only	0.579	0.641	0.634
GBA height only	0.537	0.540	0.489

Table 7.3. Area under the ROC curve for the structure covariate ablation, raw feature values, Pereira and Dosquebradas. Evaluation sets: 726 buildings with 368 positive against reviewer rejected negatives, 4,048 with 368 positive against unflagged negatives. Source:

'catalogue/sl_screen/out_experiments/E5_results.csv', 'feature_form' raw, columns 'roc_auc', 'n_eval' and 'n_pos_eval'.

The two columns say different things and the honest reading is the conjunction of both. Against MS 'rejected' negatives, removing all three height features changes the score by 0.002, which on 726 evaluations with no confidence interval and no repeated split is a tie and not a gain. Against unflagged negatives, adding them is worth 0.018 to 0.047, and the floor count supplies most of it, which is the same attribute whose presence is twice as common in the positive class.

The height and structure set is also not the best available, and the earlier claim that it sits below every combination containing a coherence drop is too strong. Three of the four coherence combinations of section 6 beat it in raw form on both negative sets: the Sentinel-1 drop alone at 0.810 and 0.801, the two NISAR tracks at 0.805 and 0.832, and their fusion at 0.843 and 0.860, all under logistic regression except the fusion against unflagged negatives, which is shallow xgboost. The fourth does not: the Liu and Balz coherence change detection (CCD) combination reaches 0.669 and 0.657, below the height and structure set on both. In neighbourhood excess form the ordering breaks further: against reviewer rejected negatives the Sentinel-1 drop alone falls to 0.556, the fusion to 0.705 and the two NISAR tracks to 0.714, all three under logistic regression and all three below 0.726. None of this makes the covariate set skilful. It makes

the comparison conditional on the negative set, the feature form and the model, which is why we report the ablation rather than a ranking.

Heights therefore enter prioritisation in two ways and no others. First, they stratify the field sample, so that the survey in Pereira and Dosquebradas, Cali and northern Quimbaya reports the coherence drop and the observed damage state by height class rather than as a city average. Second, they define a deliberate over-sample of the tall classes. In Pereira and Dosquebradas,

`catalogue/s1\_screen/out\_heights/buildings\_heights\_Pereira\_Dosquebradas.gpkg` filtered to `height\_class` in {12-24 m, >24 m} selects 774 buildings, of which 340 carry a Sentinel-1 value, and among those 340 the entire labelled evidence is five buildings, all in the 12 to 24 m class: one desk graded damaged, two graded Damaged by CEMS and two reviewer rejected (Table 7.2). A visit sample drawn from that selection, spread over the two classes and over both high and low drop values, decides whether the C band drop over tall buildings is geometry, footprint size or damage. That action belongs in the per city table of section 8, which at present schedules a height stratified survey in Quimbaya only.

Take-away: use the height classes to stratify the field sample where the attribute exists and to send a small number of visits to the 774 buildings of the two tall classes in Pereira and Dosquebradas, whose radar signals rest on five labelled buildings in the whole city and none at all above 24 m, and do not use height as a damage predictor, because on its own it scores 0.54 against a chance of 0.50 and inside the covariate set it is worth either nothing or a floor count that is itself label correlated.

Sources: `COLOMBIA\_RECON\_CATALOGUE.md`, `PROCESSING\_LOG.md`,
`SAR\_SCREEN\_RESULTS.md`, `catalogue/damage\_products/MS\_Pereira\_overture\_reviewed.gpkg`,
`catalogue/footprints/cities\_bbox.csv`, `catalogue/s1\_screen/out20/final/SECTION\_7.md`,
`catalogue/s1\_screen/out20/qa/local\_baseline\_excess\_pereira.csv`,
`catalogue/s1\_screen/out\_experiments/E5\_results.csv`,
`catalogue/s1\_screen/out\_experiments/fig17\_bar\_values.csv`,
`catalogue/s1\_screen/out\_experiments/pereira\_features.csv`,
`catalogue/s1\_screen/out\_experiments/pereira\_features\_README.md`,
`catalogue/s1\_screen/out\_heights/buildings\_heights\_`,
`catalogue/s1\_screen/out\_heights/buildings\_heights\_Pereira\_Dosquebradas.gpkg`,
`catalogue/s1\_screen/out\_heights/gba\_extract.log`,
`catalogue/s1\_screen/out\_heights/gba\_height\_raster\_coverage.csv`,
`catalogue/s1\_screen/out\_heights/height\_city\_stats.csv`,
`catalogue/s1\_screen/out\_heights/height\_coverage.csv`,
`catalogue/s1\_screen/out\_heights/join\_heights.py`,
`catalogue/s1\_screen/out\_heights/pereira\_drop\_by\_height.csv`,
`catalogue/s1\_screen/out\_heights/pereira\_height\_coherence.py`,
`catalogue/s1\_screen/out\_heights/pereira\_heights\_coherence.csv`, `figures/fig13\_drop\_by\_height.png`.

8. Recommendations for 15 to 26 September, and calibration sampling frames per city

8.1 The three spatial units, the quantity they carry, and how they nest

Three spatial units appear in the recommendations below, and all three carry the same quantity. Both radar layers, the Copernicus Sentinel-1 C-band radar and the L-band radar built by the National Aeronautics and Space Administration (NASA) and the Indian Space Research Organisation (ISRO), the NASA and ISRO Synthetic Aperture Radar (NISAR), report a coherence drop, the loss of similarity between two synthetic aperture radar (SAR) images of the same ground, computed per Overture Maps building footprint as the

pre-event coherence minus the co-event coherence. We report the layer named 'dpm_late', whose baseline is the 27 July to 8 August pair, as the headline drop and the layer named 'dpm' as a lower bound. Two thresholds recur below: a cell mean of 0.15 and a per-building value of 0.2, which are the reporting thresholds of the cell and city tables of section 5, and a cell enters any count in this section only when at least five of its buildings carry a value for the sensor being counted.

The 250 m cell is the unit of the Sentinel-1 prioritization layer and of the survey frame: the team surveys whole cells, flagged and unflagged buildings together, rather than visiting single flagged buildings. The cell size follows from what a coherence value is. Coherence is estimated in map geometry with a uniform 7 by 7 window, so each value summarizes a 140 m square of ground, in Pereira about 321 times the median building footprint, and only 25 percent of footprints contain a pixel centre and receive a value at all. The 300 m buffer around a strong-motion station is a stratum inside those cells wherever a station falls in a city, so that building typology and damage state can be paired with a recorded ground motion. That radius is the mission convention agreed for the survey design and not a quantity derived from our data. The 2.5 km survey buffer is the unit our tool design reserves for ranking areas across a city, on the principle that the prioritization unit, the tasking footprint and the validation unit must be the same polygon. No ranked 2.5 km buffer table exists in the project files, so nothing here ranks buffers: within a city the visit order comes from the 250 m cells and the label strata of section 8.2. Fig. 9 shows the cells for all seven cities and is the locator for every cell count below.

Cells set the visit order, station buffers add one stratum inside them, and the 2.5 km buffer is a placeholder for a ranking that does not yet exist.

8.2 The Pereira sampling frame: which strata, how large inside the survey cells, and how many carry a radar value

The sampling target is 100 to 200 buildings per stratum where the stratum is large enough, following the rule applied to the ten highest ranked areas per product in Turkey. That bullet also names the strata: the three-product intersection as the high confidence stratum, Microsoft only, Vantor only, Copernicus Emergency Management Service (CEMS) only, reviewer-rejected, and unflagged controls. Table 8.1 keeps every one of those strata and changes only how the Microsoft side is cut: the bare Microsoft flag is replaced by the four Microsoft reviewer classes, because a reviewer verdict carries more information than a flag, because the four classes are mutually exclusive where a set of product flags is not, and because every footprint Microsoft flags independently carries one of them. The three-product intersection and the pairwise agreement strata are recorded per building rather than visited separately, for the reason given below.

Three properties of Table 8.1 govern how it is used. First, the sizes in the fourth column are counted inside the 367 Sentinel-1 cells above a 0.15 mean drop, which is the frame the Pereira instruction of section 8.3 restricts the survey to, about 22.9 square kilometres holding 26,793 of the city's 91,266 footprints. The third column keeps the whole-city product counts so that the two can be compared. Second, the seven rows are disjoint and exhaust the frame: a footprint carrying a reviewer verdict is assigned to that verdict, a footprint with no verdict but a Vantor flag to row 5, a footprint with neither but a CEMS grade to row 6, and everything else is a control. Third, a coherence value exists on about one footprint in four, so the fifth and sixth columns, not the fourth, set what each stratum can calibrate.

Stratum	What it measures	In the released product	Inside the 367 cells	With a Sentinel-1 value	With a NISAR value	Also CEMS graded	Sampling instruction
1. Reviewer class 'ground_truth', a desk review	the per-building excess both sensors are	136	127	38	43	34	survey in full

of imagery	tested on						
2. Reviewer class 'confirmed'	ordinality of the reviewer ladder	156	97	37	38	9	survey in full
3. Reviewer class 'unsure'	the decision boundary	154	68	37	33	6	survey in full
4. Reviewer class 'rejected'	the false positive rate	367	93	21	16	1	survey in full
5. Flagged by Vantor, no Microsoft reviewer verdict	what one model sees alone	309 flagged	52	10	12	3	survey in full
6. CEMS graded, no reviewer verdict and no Vantor flag	recall, what the models missed	182 points released, 132 matched to a footprint	68	33	30	68	survey in full
7. Unflagged controls	precision, and the radar false alarm rate	91,266 footprints, 22,731 with a Sentinel-1 value	26,288	6,607	6,565	0	every building with both sensor values in a drawn set of cells, plus a top-up to 200

Table 8.1. Calibration strata for Pereira and Dosquebradas. The rows are disjoint and sum to the 26,793 footprints inside the survey cells, so a building belongs to exactly one row and the last column can be executed without double counting. Reviewer classes are desk verdicts by Microsoft reviewers on machine flags, not field observations, and the NISAR counts are provisional. The join of the reviewer classes onto the survey layer reproduces the product counts within three footprints per class (136, 159, 155 and 366 against 136, 156, 154 and 367). Data providers: Microsoft AI for Good, Vantor, Copernicus Emergency Management Service, Overture Maps, Copernicus Sentinel-1, NASA and ISRO NISAR via the Alaska Satellite Facility (ASF). Source: catalogue/agreement/agreement_log.txt, section "PEREIRA" (third column); catalogue/s1_screen/out20/final/Pereira_Dosquebradas_buildings_s1_final.gpkg joined to Pereira_Dosquebradas_cells250.csv and to catalogue/damage_products/Vantor_Pereira_overture_predictions.gpkg (columns four to seven, new computation for this report). Every stratum that decides whether the radar layers work holds fewer than 130 buildings inside the survey cells, which is why six of the seven rows are surveyed in full rather than sampled.

Three cross-cutting facts complete the frame. The high confidence stratum the mission framework asks for exists and is small: 19 footprints in Pereira are flagged by Vantor and independently by Microsoft and also carry a CEMS grade, 17 of them graded Destroyed and two Damaged, all 19 inside the 367 cells, 12 with a Sentinel-1 value and 13 with a NISAR value. Those 19 sit inside rows 1 to 3 of Table 8.1, 17 in 'ground_truth', one in 'confirmed' and one in 'unsure', so they are not a separate visit but a label to record. The pairwise independent agreement they come from is 45 footprints flagged by both products, 36 by Microsoft alone and 85 by Vantor alone, computed on the 29,191 footprints whose Microsoft flag was not imported from Vantor. Recording product membership per building is what makes these strata recoverable from one sample, which is also why section 8.4 asks for the stratum membership of every surveyed building.

The frame is smaller than the products suggest, in two ways the field team will meet on the first morning. Rows 1 to 6 inside the survey cells total 505 buildings, of which 176 carry a Sentinel-1 value and 172 a NISAR value, so the labelled Pereira frame plus 200 controls is about 705 buildings, a frame set by what lies inside the survey cells rather than by the whole-city product counts of the third column. And the radar

arm is a quarter of the optical arm: of the 26,793 footprints inside the cells, 6,783 carry a Sentinel-1 value, 6,737 a NISAR value and 3,069 both, because the missing three quarters are the footprints too small to contain a 20 m pixel centre. A surveyed building with no coherence value is still a full record: it grades the Microsoft, Vantor and CEMS products and it feeds the exposure model, and it is marked as carrying no radar value so that it never enters a precision or recall count for either sensor.

If the twelve days do not hold all seven cities at this frame, the order to protect is row 1, then row 6, then row 4, because those three fix, in turn, the per-building excess the layers are ranked on, the recall the desk products could not deliver, and the false positive rate. Rows 2, 3 and 5 follow, and the controls are drawn last but never dropped, since precision cannot be computed from flagged buildings alone. How many buildings a team can grade in a day is not a quantity in our files, so we ask the mission leads to set it and we will re-cut the frame to whatever rate they give.

The strata that decide whether the radar layers work hold 505 buildings between them inside the survey cells, 176 of them with a Sentinel-1 value, which is why six of the seven strata are surveyed in full and why the stratum membership of every surveyed building has to be written down.

8.3 What to do in each city

Table 8.2 is the printable version of this subsection, the Sentinel-1 250 m drop maps of section 5 locate the cells it names, and Fig. 30 is the map of the draw itself.

City	Action	File to load	Establishing section
Pereira and Dosquebradas	build the block-level visit list; survey whole 250 m cells by the strata of Table 8.1 with unflagged controls	catalogue/s1_screen/out20/final/Pereira_Dosquebradas_buildings_s1_final.gpkg (both sensors, 'review_status', 'cems'); catalogue/s1_screen/out20/final/Pereira_Dosquebradas_cells250.csv; catalogue/damage_products/Vantor_Pereira_overture_predictions.gpkg	4, 5, 6
Cali	field-check the CEMS graded points and the Microsoft flags, record the points with no footprint as an exposure finding, survey the buffer of station RAC2C	catalogue/s1_screen/out20/final/Cali_buildings_s1_final.gpkg ('ms_damaged', 'ms_damage_pct', 'cems')	4, 5
Sevilla	walk three high and three low 250 m cells	catalogue/s1_screen/out20/final/Sevilla_cells250.csv; catalogue/s1_screen/out20/final/Sevilla_buildings_s1_final.gpkg	5
Quibdó	record demolition dates and rainfall; use the CEMS AOI04 points as the only labels	catalogue/s1_screen/out20/final/Quibdo_cells250.csv; catalogue/damage_products/CEMS_AOI04_Quibdo_v1/	5

Armenia	georeferenced damage and no-damage observations, including the two-burst overlap; buffer of station ARMEC	catalogue/s1_screen/out20/final/Armenia_buildings_s1_final.gpkg; catalogue/s1_screen/out20/final/Armenia_cells250.csv	3, 5
Manizales	as Armenia; buffer of station MANIC	catalogue/s1_screen/out20/final/Manizales_buildings_s1_final.gpkg; catalogue/s1_screen/out20/final/Manizales_cells250.csv	3, 5
Quimbaya	observations by building height class, and storey counts against the modelled heights	catalogue/s1_screen/out20/final/Quimbaya_buildings_s1_final.gpkg; catalogue/s1_screen/out20/final/Quimbaya_cells250.csv	5, 7

Table 8.2. One action and the files that carry it, per city, for 15 to 26 September 2026. Each named file holds the attributes its action needs: the Pereira and Cali products carry the label columns, the cell tables carry the survey frame, and the Vantor product carries the flags that are absent from the Microsoft layer. Every NISAR layer inside these files is a provisional product. Data providers: NASA, ISRO and ASF (NISAR), Copernicus Sentinel-1, Copernicus Emergency Management Service, Microsoft AI for Good, Vantor, Overture Maps. Source: the section named in the last column. Three cities have a released damage product to calibrate and four are where the team's own observations become the only label.

Pereira and Dosquebradas. Build a block-level visit list, ordered inside each 250 m cell by the per-building NISAR drop, and not a list of single addresses. Two properties of the product fix that choice. The support of the estimate is the block and not the building: the median Overture footprint in Pereira is 61 square metres, 93 percent of footprints are smaller than one 20 m pixel, 25 percent contain a pixel centre, and the coherence window is a 140 m square. And the geocoded coherence is displaced about 40 m toward the sensor, a best footprint shift of plus 40 m east and minus 5 m north measured in both Pereira and Armenia, which is two pixels inside the 140 m window and therefore leaves city and neighbourhood statistics unaffected. So when a flagged footprint and its neighbour cannot be separated, which at a 61 square metre median footprint is the common case, the team surveys the whole block face containing the flagged footprint, records every building on it with its Overture identifier, and marks which footprint carried the flag.

NISAR is still the layer to order that list by. On the reviewer class 'ground_truth', the desk-review class of Table 8.1, the NISAR track T047A frame 004 per-building drop keeps a 200 m neighbourhood excess of +0.135, which is 1.89 times the citywide median drop of unflagged buildings, 0.072, on 46 buildings, and Sentinel-1 on the same class keeps an excess of +0.061 against its own citywide unflagged median of 0.105, a ratio of 0.58, on 41 buildings. The two values sit on different subsets of the 136 buildings, which is why they are stated as ratios to each sensor's own citywide unflagged median rather than compared directly; the denominator is the citywide median and not the 200 m local baseline, which for the same class is 0.166 on NISAR T047A and 0.204 on Sentinel-1. The largest NISAR excess in Pereira is not this class but CEMS Destroyed, at 1.96 times its own citywide unflagged median on 31 buildings. The 200 m radius is the neighbourhood baseline defined in section 5. Both results are provisional and rest on 41 to 46 labelled buildings.

Use the Sentinel-1 250 m cells to delimit the core, where 367 of the 1,442 cells with at least five buildings exceed a mean drop of 0.15. Sample the strata of Table 8.1 inside those cells and record the Overture footprint identifier, the number of storeys and the stratum membership on every surveyed building, so that precision and recall per building can be computed for NISAR, Sentinel-1, Microsoft and Vantor on one sample. Two stations are assigned to Pereira and they disagree. Every station-to-city distance in this section is the `dist\_to\_city\_km` column of the station file, which measures to the city point that file carries; Table 2.1 recomputes the same distances as great circles from the city bounding-box centroids and differs by up to 2.3 km, and neither convention changes which station is nearest. Station CBOCA at 6.6 km reads 1.8 percent g peak ground acceleration (PGA), 2.4 cm/s peak ground velocity (PGV) and instrumental Modified Mercalli Intensity (MMI) 3.9, while station FLND at 16.6 km reads 43.1 percent g, 23.4 cm/s and MMI 6.8, and CBOCA is the station the mission notes flag as "check instrument". Our building layer holds 25 footprints within 300 m of CBOCA, 10 of them with a Sentinel-1 value, and none at all within 300 m of FLND. Survey the CBOCA buffer in full: a graded damage state there is what arbitrates between an MMI 3.9 and an MMI 6.8 reading 10 km apart, and it is the survey the Global Earthquake Model (GEM) and Arup typology and European Macroseismic Scale 1998 (EMS-98) work was designed around.

Cali. Cali is the city where we do not yet have a usable sampling frame, and the recommendation is written that way: neither released optical product is a frame on its own, and the sources that would give one have to be obtained. Use the combined Sentinel-1 product for the co-event drop only and not for its post-event or recovery layers, which are contaminated by azimuth banding in both Cali bursts. A single Cali building's value is at the repeatability floor: in the 7,555-building overlap the two bursts agree on `dpm` at Spearman 0.138, so a Cali flag is evidence about a block and never about one address.

Three field tasks follow. Field-check the 21 CEMS graded points in the two Cali areas, 7 in AOI01 and 14 in AOI03, including the 8 graded Destroyed in AOI03. Only 10 of those 21 points carry an Overture identifier in the delivered Cali building layer the team will carry, and 8 of the 10 have a Sentinel-1 value. The other 11 carry no Overture identifier in that layer, so they cannot become calibration records; the agreement analysis of section 4 uses a looser join, a 15 m point-to-footprint match onto the Microsoft Cali layer, and matches 12 of the 21 points onto 11 distinct footprints, so the two counts are of two different joins and not of the same quantity; record them instead as an exposure finding, because the reason they have no footprint is that they are informal or newly built structures the exposure model does not hold. Second, field-check the Microsoft flags: the released Microsoft layer flags 266 Cali footprints and 256 of those join the delivered Cali building layer, 25 of them inside AOI01 and 2 inside AOI03, and Microsoft flags none of the CEMS points in either area. Of the 256, the 80 with a Sentinel-1 value carry a marginal signal, a raw drop of 0.135 against 0.125 for unflagged buildings (p 0.09) and a 200 m excess of plus 0.019 (p 0.045), so a confirmed or refuted flag in the field is what turns a marginal statistic into a usable one. Third, survey the 300 m buffer of station RAC2C, 8.8 km from the city, which recorded 11.6 percent g PGA, 25.4 cm/s PGV and instrumental MMI 6.9; our layer holds 80 footprints inside it, 33 with a Sentinel-1 value, so the buffer is surveyed in full.

None of that reaches the damage type Cali actually suffered: 56 collapses are reported there, dominated by mid-rise structural failures that nadir optical products under-see. Three sources would close that gap and none is a file in this data directory, so each is written as a request with a fallback. Pietro Milillo requests from the Charter value adds the Pléiades Neo affected-housing vectors for Cali, reported as 27 destroyed and 124 damaged; if the vectors do not arrive before 15 September the team works from the georeferenced PDF map and records positions by hand. The geolocated sites of the NZERP and NZSEE New Zealand reconnaissance preliminary report of 13 August sit in a members-only viewer with only the PDF saved to our library, so a member of the mission requests the site list directly, and the fallback is the report's per-city case descriptions. The crowd compilation at mapadelterremoto.com holds 3,857 points and is unofficial and not validation grade, so it is used to plan a route and never as a label.

Sevilla. Walk three cells with a high Sentinel-1 drop and three with a low one. Sentinel-1 exceeds a 0.15 mean drop in 38 of the 66 cells with at least five buildings, while NISAR exceeds it in none of the 59 cells that hold at least five buildings carrying a NISAR value. The town has 2,165 buildings with a Sentinel-1 value, of which 42.5 percent exceed a 0.2 drop, the highest share of the seven cities. No station buffer stratum exists here: the nearest strong-motion station, CTRUJ, is 43.1 km away. Six cells are enough to decide whether the C-band drop here is damage or the surface change of a small vegetated town.

Quibdó. Quibdó carries the second highest city drop in Sentinel-1, a building median of 0.156, and 127 of its 268 cells exceed 0.15. NISAR keeps decorrelating after the event where the other cities recover: inside built-up areas the median coherence falls from 0.703 in the 12-day pre-event pair to 0.625 in the 24-day co-event pair and 0.448 in the 12-day post-event pair on track T011D. The only labels are the 74 CEMS AOI04 graded points, 18 Destroyed, 8 Damaged and 48 Possibly damaged, of which 63 match a building footprint and 26 carry a Sentinel-1 value. No station buffer stratum exists here either: the nearest station, SOL, is 101.6 km away. Record demolition dates and rainfall on every surveyed site, because both explain a post-event coherence loss that is not damage.

Armenia and Manizales. Neither city has a released building-level damage product, and the two sensors disagree about where the damage is: the Spearman correlation of the 250 m cell means between Sentinel-1 and NISAR is 0.003 in Armenia and 0.102 in Manizales, so a field observation is the only thing that can break the tie. Sentinel-1 exceeds a 0.15 mean drop in 269 of 888 Armenia cells and 125 of 669 Manizales cells. Give priority inside those cells to the buildings that lie in the two-burst overlap, 4,976 in Armenia and 3,842 in Manizales, of which 695 and 753 respectively take their value from the secondary burst; the columns `'source_burst'` and `'dist_to_edge_m'` in the two GeoPackages identify them, and field observations there test whether the two-burst combination rule holds. Survey the 300 m buffers of stations ARMEC, 3.4 km from Armenia, which recorded 22.6 percent g PGA, 17.6 cm/s PGV and MMI 6.4, and MAN1C, 0.8 km from Manizales, which recorded 28.9 percent g, 17.3 cm/s and MMI 6.4. Those buffers hold 224 footprints in Armenia, 112 with a Sentinel-1 value, and 726 in Manizales, 190 with a Sentinel-1 value, so both are 100 to 200 building draws rather than full surveys.

Quimbaya. Quimbaya carries the highest NISAR per-building drop of the seven cities on track T047A frame 004, a median of 0.089 with 16.8 percent of 2,196 buildings above a 0.2 drop, but the ranking inverts on the other track: on T011D frame 086 the same buildings give 0.058, below Sevilla at 0.076 and Pereira and Dosquebradas at 0.071. The two facts that do not move are the reasons to go: nothing has been released to check Quimbaya against, and 33 of its 89 Sentinel-1 cells exceed 0.15. It is also the one city where a field storey count validates a modelled height product: a height attribute exists on 73.8 percent of its 7,707 footprints, all 5,687 of them modelled GlobalBuildingAtlas heights, against observed storey counts on 5 buildings, and no station is assigned to Quimbaya in our station table. Record observations by height class here and treat the class as a stratification variable, not as a damage predictor.

Every city has one file to load and one action; three have a released product to calibrate, and four are where the team's own observations become the only label.

8.4 What to record on every surveyed building

Six fields turn a survey into a calibration set: the Overture footprint identifier, the number of storeys, the EMS-98 damage grade, whether the damage is roof-visible or facade-only, the green, yellow or red tag of the national building inspection run by UNGRD, the Colombian agency responsible for post-earthquake building inspection, and whether the site has been cleared since the earthquake. Without the identifier a field observation cannot be joined to a footprint; without the clearing date a demolished building looks undamaged in post-event imagery. The damage grade and the inspection tag are not the same field and both are needed: the tag is a habitability judgment, it is assigned on a different logic from any product this report calibrates, and it will be absent on the unflagged controls. Only an EMS-98 grade lets us test

whether the Microsoft reviewer classes are ordinal, and only the roof-visible against facade-only distinction answers why nadir optical products under-see the mid-rise failures of Cali. Record with them the stratum membership of Table 8.1 and, in Pereira, whether the building carries a Microsoft flag, a Vantor flag, a CEMS grade or any combination, since the agreement strata are recovered from that field rather than from separate visits. Observations should return on an agreed nightly channel for all twelve days, with a log of every layer we send to the field, its version and date, so that the frames and any later ranking can be recomputed while the team is still in the country.

Survey whole 250 m cells by stratum with unflagged controls, record the footprint identifier, storeys, EMS-98 grade, inspection tag and stratum membership on every building, and send the observations back nightly so the frames and the ranking can be recomputed before the mission ends.

Figures

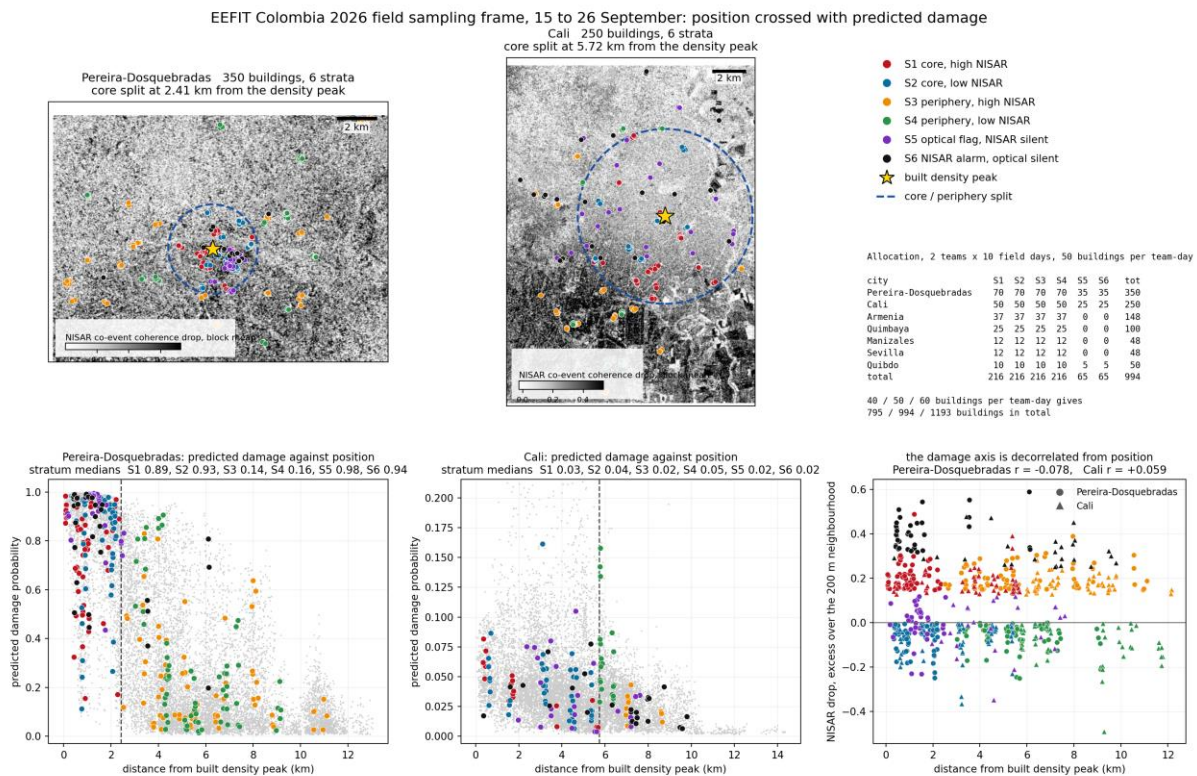


Fig. 30. the field sampling frame for 15 to 26 September. The two upper map panels place the drawn buildings on the NISAR co-event coherence drop in Pereira and Dosquebradas and in Cali, split into a core and a periphery at the distance from the built density peak, 2.41 km and 5.72 km, and coloured by the six position and prediction strata S1 to S6; the allocation panel gives the draw per city, 350 in Pereira and Dosquebradas, 250 in Cali, 148 in Armenia, 100 in Quimbaya, 48 in Manizales, 48 in Sevilla and 50 in Quibdó, 994 buildings in all at 50 buildings per team-day over two teams and ten field days. The three lower panels show that the predicted damage falls with distance from the density peak in both cities while the NISAR excess over the 200 m neighbourhood does not, which is why the draw crosses position with prediction rather than following prediction alone. This frame is complementary to the label strata of Table 8.1: Table 8.1 stratifies Pereira by what the released products say about each building, and this figure stratifies both product-bearing cities by where the building sits and what the radar predicts there. Data providers: NASA and ISRO NISAR via ASF (provisional), Copernicus Sentinel-1, Microsoft AI for Good, Overture Maps. Source: catalogue/s1_screen/out_ablation/sampling_allocation.csv, sampling_density_peaks.json, sampling_frame_Pereira.gpkg and sampling_frame_Cali.gpkg, produced by catalogue/s1_screen/out_ablation/fig22_sampling_design.py.

Sources: `AGREEMENT\_FRAMEWORK.md`, `ARCHITECTURE\_PLAN.md`,
`COLOMBIA\_RECON\_CATALOGUE.md`, `Emails/Re: Remote team support to Colombia2.pdf`,
`LESSONS\_LEARNED.md`, `SAR\_SCREEN\_RESULTS.md`, `catalogue/agreement/agreement\_log.txt`,
`catalogue/charter\_products`, `catalogue/damage\_products/CEMS\_AOI04\_Quibdo\_v1`,
`catalogue/damage\_products/CEMS\_AOI04\_Quibdo\_v1/EMSR916\_AOI04\_GRA\_PRODUCT\_builtUpP\_v1.shp`,
`catalogue/damage\_products/MS\_Pereira\_overture\_reviewed.gpkg`,
`catalogue/damage\_products/Vantor\_Pereira\_overture\_predictions.gpkg`,
`catalogue/s1\_screen/out20/final/Armenia\_buildings\_s1\_final.gpkg`,
`catalogue/s1\_screen/out20/final/Armenia\_cells250.csv`,
`catalogue/s1\_screen/out20/final/Cali\_buildings\_s1\_final.gpkg`,
`catalogue/s1\_screen/out20/final/Manizales\_buildings\_s1\_final.gpkg`,
`catalogue/s1\_screen/out20/final/Manizales\_cells250.csv`,
`catalogue/s1\_screen/out20/final/Pereira\_Dosquebradas\_buildings\_s1\_final.gpkg`,
`catalogue/s1\_screen/out20/final/Pereira\_Dosquebradas\_cells250.csv`,
`catalogue/s1\_screen/out20/final/Quibdo\_cells250.csv`,
`catalogue/s1\_screen/out20/final/Quimbaya\_buildings\_s1\_final.gpkg`,
`catalogue/s1\_screen/out20/final/Quimbaya\_cells250.csv`,
`catalogue/s1\_screen/out20/final/SECTION\_7.md`,
`catalogue/s1\_screen/out20/final/Sevilla\_buildings\_s1\_final.gpkg`,
`catalogue/s1\_screen/out20/final/Sevilla\_cells250.csv`, `catalogue/s1\_screen/out20/final/build\_final.py`,
`catalogue/s1\_screen/out20/final/build\_final\_log.json`,
`catalogue/s1\_screen/out20/final/fig12\_pereira\_label\_test\_stats.json`,
`catalogue/s1\_screen/out20/final/s1\_city\_table.csv`,
`catalogue/s1\_screen/out20/final/s1\_nisar\_cells\_250m.csv`,
`catalogue/s1\_screen/out20/qa/local\_baseline\_excess\_pereira.csv`,
`catalogue/s1\_screen/out\_ablation/fig22\_sampling\_design.py`,
`catalogue/s1\_screen/out\_ablation/sampling\_allocation.csv`,
`catalogue/s1\_screen/out\_heights/height\_coverage.csv`,
`catalogue/s1\_screen/out\_nisar/gunw\_city\_stats.csv`,
`catalogue/s1\_screen/out\_nisar/nisar\_building\_dpm\_stats.csv`,
`catalogue/strong\_motion\_stations\_shakemap\_v8.csv`, `figures/fig22\_sampling\_design.png`.

9. Data, code, reproducibility, licenses, and AI disclosure

What the field team carries can be traced to a named input and, for most of the chain, rerun from the script that made it; the parts that cannot be rerun from this directory are listed in 9.4, the single automated gate covers the Sentinel-1 bursts and nothing else, and a named set of inputs carries licenses that restrict what may be done with anything derived from them. All paths below are relative to the project data directory. The rule we applied throughout is that every number traces to a catalogued file and that no new number is computed in prose.

9.1 The chain, its code and its outputs

Table 9.1. The processing chain, one row per step, each naming its inputs, its settings and its outputs. The rows that cannot be rerun from this directory as they stand are listed in 9.4. Source:

PROCESSING_LOG.md sections A to G and the entry "2026-09-05, Liu and Balz 2026 package".

Step	Script and directory	Key settings	Main outputs
Harvest	`download_bursts.py`, `nisar/download_gunw.py`, `nisar/download_gslc.py`, `rtc/download_rtc.py`,	ComCat, ASF SearchAPI, Common Metadata Repository (CMR), Copernicus Data Space Ecosystem (CDSE) OData,	1,023 rows in `imagery_acquisitions_timeline.csv`; 82 Sentinel-1 single-look complex (SLC) scenes; 12 NISAR geocoded

	`footprints/` scripts; the remaining sources are inline code (9.4); `catalogue/`	Planetary Computer SpatioTemporal Asset Catalog (STAC), Charter geobrowser, CEMS activation API, Humanitarian Data Exchange (HDX), Overture, Google Open Buildings, GlobalBuildingAtlas	unwrapped interferogram (GUNW) files, provisional; 96 OPERA radiometric terrain corrected (RTC) GeoTIFFs over 19 bursts on 11 dates; 118 Sentinel-2 tiles and 46 Landsat scenes catalogued and not processed; 22 Vantor items; 36 OpenAerialMap images; 23 Charter products and 624 Charter acquisitions; 67 strong-motion stations, 66 of them Servicio Geológico Colombiano
Product agreement	no script in the directory (9.4); `catalogue/agreement/`	European Petroleum Survey Group (EPSG) code 32618; footprint centroid to footprint within 3 m on the same base, CEMS point to nearest footprint within 15 m; 250 m grid; provenance filter `damage from!= vantor`	`agreement_log.txt`, four GeoPackages
NISAR coherence, provisional	`gunw_screen.py`, `gunw_dpm.py`, `gslc_baseline.py`; `out_nisar/`	five range by six azimuth looks at 20 m, `wrappedInterferogramFiltering = no_filter`	coherence and `dpm` rasters, per building GeoPackages, `gunw_city_stats.csv`, `nisar_building_dpm_stats.csv`, all provisional
RTC change	`rtc_screen.py`; `out_rtc/`	10·log10(post/pre); track 142 pair 2 to 14 August, track 048 pair 7 to 19 August	city mosaics, `rtc_city_stats.csv`
Sentinel-1 coherence	`s1_screen.py`, `s1_derive.py`, `s1_aggregate.py`; `out20/`	range Doppler geocoding onto a Copernicus GLO-30 UTM 18N grid, 7 by 7 window, 20 m; co as 8 to 14 August; `dpm` against the mean of three pre-event pairs, `dpm_late` against the mean of the two pre-event pairs closest to the event, the second being the headline layer; post as the mean of two post pairs	`dpm`, `dpm_late`, `recovery` rasters, per building GeoPackages, `out20/final/` city tables
Liu and Balz test	`balz_ccd` package; `out_balz/`	Sentinel-1 burst 303906 at 10 m, NISAR geocoded single look complex (GSLC) track T011D frame 086 at 5 m, the latter provisional	two XGBoost regressors, coherence, deviation and change rasters
Experiments, heights, figures	the eight `run_E` and `E` drivers plus `build_pereira_features.py`, `build_summary_table.py`, `fig_E2.py` and `leakage_audit/audit_leakage.py` in `out_experiments/`; `gba_extract.py`, `join_heights.py` and `pereira_height_coherence.py` in `out_heights/`; `run_ablation.py` in `out_ablation/`; `run_zerohour.py`, `run_transfer.py`, `run_operational_value.py` and `refute_location_proxy.py` in `out_zerohour/`; `fig14_15_balz.py` in `out_balz/`; `build_final.py`, `fig10_s1_drop_maps.py`, `fig12_pereira_label_test.py`, `make_fig11.py` and `make_fig9.py` in `out20/final/`	500 m blocked folds, two negative definitions, raw and excess feature forms	`out_experiments/`, `out_heights/`, `out_zerohour/`, and 21 of the 32 files in `figures/`

The 3 m and 15 m matching radii and the 250 m grid are the same constants section 4 uses; the grid is set at 250 m because that is the scale at which the released damage products agree even where their building level labels do not, and it is the scale used for ranking. The 7 by 7 coherence window at 20 m posting was chosen, not derived, and its consequence is that every coherence value in this report has a ground support of 140 m by 140 m, which is the number 9.3 uses to argue that the geolocation offset moves no city or neighborhood statistic. The analysis code is 97 Python files under `catalogue/sl\_screen/`, with five shell drivers and two workflow definitions, re-counted at assembly at 19:39 on 5 September 2026.

Take-away: every row names its inputs, its settings and its outputs, and the NISAR, RTC, Sentinel-1 and Liu and Balz rows can be rerun as they stand from the scripts named in them, while the harvest row rests in part on inline code, the product agreement row has no script in the directory, and 17 of the 42 figures have no producing code, all of which 9.4 lists.

9.2 Tests and quality control

What is automatically tested in this chain, and what is not, does not follow the boundary a reader would expect. The change detection package carries 25 tests covering the simulator covariance by two factorizations, the preprocessing invariances, the ten window statistics against a slow reference loop, the phase gradient definition and the degenerate coherence cases. Its regressors were trained on 10,000,000 simulated windows at seed 42 and tested on 500,000 windows drawn with an independent seed, using XGBoost 3.2.0, giving a root mean square error of 0.153 for the machine learning coherence against 0.176 for the sample coherence, and 0.168 for the deviation against 0.373 for a zero baseline. Those are tests of the method against simulation, not of any Colombia product.

Only one gate tests a Colombia product, the regression test `out20/qa2/test\_burst\_geolocation.py`, and it tests the 20 m Sentinel-1 bursts alone. It requires that at least 80 percent of valid co-event pixels fall inside the ASF burst footprint and that median coherence inside building footprints exceeds the median outside by at least 0.1 in four of six pairs, and it treats the urban criterion as decisive only where at least 200 valid pixels fall inside footprints. The 80 percent overlap is set to separate a correctly geolocated burst from the subswath failure mode, which put IW2 and IW3 bursts about 0.65 to 0.75 degrees east and drove the overlap to near zero. The 0.1 median delta and the four of six pairs were chosen, not derived, and the twelve bursts on which the criterion is decisive clear the delta by roughly a factor of two, at 0.192 to 0.315, so the gate is not tuned to its own margin. On the corrected run all 13 bursts hold 0.984 to 1.0 of their valid pixels inside the footprint, and 12 of the 13 clear the urban criterion decisively while one, burst 303901 IW2, is left undecided because only 36 valid building pixels fall inside footprints there, that burst clipping only the fringe of its city.

Take-away: one automated gate stands behind the 20 m Sentinel-1 bursts, where 12 of the 13 cleared the urban criterion decisively and one was left undecided on 36 valid building pixels; the NISAR, OPERA RTC and 10 m Liu and Balz products carry no equivalent test, and the machine learning components of the last were validated against simulation before Colombia rather than on Colombia.

9.3 Properties of the delivered products

The products the field team carries have properties that outlive the processing that produced them, and these are the ones that change how a number should be read. All 20 m Sentinel-1 products carry a geolocation offset of about 40 m east and 5 m south. That is the toward-sensor direction on this descending track, and the cause is that Copernicus GLO-30 orthometric heights were used as ellipsoidal heights: an undulation of 26.6 to 27.8 m at 34 degrees incidence predicts 39 to 42 m, matching the measured shift, and the digital elevation model (DEM) step now adds the Earth Gravitational Model 2008 undulation for future per building work at 10 m. The 10 m Liu and Balz products carry the same offset, because they were geocoded from the same uncorrected DEM. City and neighborhood statistics are unaffected in both cases, because the aggregation window is 140 m.

Every Sentinel-1 table in this report comes from a corrected run. The geocoder selects the annotation block matching the subswath in the file name, and the six reprocessed IW2 and IW3 products, spanning bursts 303901, 303902, 303914 and 303915, geolocate with 99.7 to 99.95 percent of their valid pixels inside the ASF footprint, the whole set of 13 bursts sitting at 0.984 to 1.0 as reported in 9.2. Per city Sentinel-1 files are named ``<City>_T142_<burstid>_<swath>_<layer>.tif``, and any file carrying no swath token belongs to a superseded set in which the IW3 clip of a burst replaced the IW2 clip of the same burst identifier, which affected Quibdó and burst 303901. The Sentinel-1 pre-event mean is nodata aware, so the drop is empty rather than populated where the co-event pair has no data; in the superseded layer 6.5 percent of the pixels above a 0.3 drop in burst 303901 IW2 were artifacts of that kind. `dpm\_late` rather than `dpm` is the headline Sentinel-1 layer, because the 21 to 27 July pair sits 25 to 37 percent below the other pre-event pairs and inflates the pre-event spread.

Two further properties bound any per building reading. A fixed coherence threshold applied to these products should be read against 25 to 27 effective looks rather than the nominal 49 of a 7 by 7 window, because bicubic geocoding correlates neighboring output pixels and the 22 m azimuth resolution exceeds the 20 m posting, and 93 percent of Pereira footprints are smaller than one pixel with only 25 percent containing a pixel centre. And per building Sentinel-1 drops are noise dominated: across all overlapping burst pairs the adjacent-burst repeatability is Spearman 0.14 to 0.37, and on the 7,555 Cali buildings valued twice by two adjacent bursts it is 0.138.

Take-away: the geolocation offset moves no city or neighborhood number, the subswath, file naming and derived layer defects were all corrected before any table in this report was built, and the repeatability floor together with the effective looks result is why Sentinel-1 is reported at 250 m cell or neighborhood excess level and not per building.

9.4 What cannot be regenerated as the directory stands

Some of this chain cannot be rerun from the directory as it stands, and we set out where, so that nobody meets a gap in the field. The gaps that carry numbers this report prints come first, and the ones that carry none come after.

Gaps that carry numbers. First, the harvest of the USGS products, the NISAR and OPERA inventories, the Vantor STAC walk, the Charter geobrowser parse and the merge that produces the timeline was done with inline code that is not in the directory, so the 67 station file, the OpenAerialMap, Vantor and Charter catalogues and the 1,023 row timeline behind report sections 2 and 3 cannot be rebuilt from a script here. Second, the product agreement step has no producing code anywhere in the tree, so the section 4 tables built on `agreement\_log.txt` and the four GeoPackages cannot be regenerated. Third, the neighborhood baseline cross-check table `out20/qa/local\_baseline\_excess\_pereira.csv` is an output with no identified producer: the string appears in five code files, and in every one of them it sits inside prompt text or a source list that consumes the table as an established input rather than writing it. That table carries the Sentinel-1 `dpm` excess of +0.068 and the NISAR T011D f086 excess of +0.138 that section 5 uses to say the L-band result reproduces on a second track, so one of this report's two independent track claims rests on a file nothing here can rebuild. Fourth, 11 of the 32 files in `figures/` are produced by no code file of any kind, the log recording that the earliest of them were generated by inline matplotlib. Fifth, the NISAR, OPERA RTC and 10 m Liu and Balz chains carry no automated test in the directory, as 9.2 states.

Gaps that carry no number. The geocoding library named `Coherence\_Maps` lives outside the data directory and is versioned separately, with the patched annotation readers backed up as `\*.bak\_20260905`. And packaging the chain as one command line tool with a single event configuration is planned, not done. Separately from all seven, every NISAR product used in this report is provisional and is labelled as such wherever it appears.

Take-away: the external library and the unbuilt package are packaging work that changes no number, the other gaps sit under numbers this report prints, and the one to watch is the neighborhood baseline table, because a claim that no code here can rebuild is a claim the field team should treat as unrepeated until it is repeated.

9.5 Licenses and conditions of use

Table 9.2. The inputs whose license constrains what the mission may do with a layer derived from them. Three of them are non-commercial. This table is not the credit list: section 10 carries every institution whose data was used, including the ones whose licenses place no constraint. Source: report/REPORT_SPEC.md, Credits.

Provider	What we used	License or credit line
USGS	event, ShakeMap v8, PAGER, ground failure	public domain
Servicio Geológico Colombiano	strong motion stations	credit line, `sgc.png`
ESA and Copernicus	Sentinel-1 bursts, Sentinel-2	Copernicus Sentinel data 2026
Copernicus Emergency Management Service	EMSR916 grading points	CEMS products © European Union 2026
NASA, JPL and ISRO, via ASF	NISAR GUNW and GSLC, provisional	provisional products via ASF
OPERA and JPL, via ASF	RTC-S1 backscatter	credit line, `opera.svg`
Microsoft AI for Good Lab	Pereira and Cali damage layers	CC BY
Vantor and Maxar Open Data	Pereira damage layer	CC BY-NC 4.0, non-commercial
Planet Labs PBC	Crisis Response imagery	CC BY-NC 4.0, non-commercial
Airbus and CNES, via Charter and CEMS	Pléiades and Pléiades Neo	credit lines, `airbus.png`, `cnes.png`
International Charter	activation 1048, requester UNGRD	credit line and activation number
Overture Maps Foundation	building footprints	Open Database License (ODbL) and Community Data License Agreement (CDLA)
Google	Open Buildings v3	CC BY 4.0
TUM	GlobalBuildingAtlas LoD1 heights	CC BY-NC 4.0, non-commercial, cite Zhu et al. 2025, Earth System Science Data (ESSD), listed in 9.7
OpenStreetMap contributors and HOT	extracts	ODbL
OpenAerialMap	36 aerial images	CC BY 4.0, text credit, no logo in report/logos/
IGAC, UNGRD, UNITAR and UNOSAT, the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) and HDX, IOM DTM	orthophotos, situation reports, grades	credit lines, IOM as text only, no logo obtained

OpenAerialMap is not in the credits list of record and holds no logo file, so section 10 carries it as an added text credit, in the way IOM is handled, which is what CC BY 4.0 requires of us.

Take-away: three inputs the field pack depends on, the Vantor layer, the Planet imagery and the GlobalBuildingAtlas heights, carry non-commercial licenses, so layers derived from them may be shared inside the mission and published here but not used commercially.

9.6 AI disclosure

Parts of this report were drafted and checked with AI assistance (PeerPrism agent suite, Claude models) under the supervision of the remote-sensing group; every number was verified against the project files listed in section 9. In practice this meant that no number entered a section without a source tag naming the file and the location inside it, that each section was checked against those files by a reader who had not written it, and that the tables in 9.1 and 9.4 were built by walking the directory rather than by reading the processing log.

9.7 References

Five external works are named in the body of this report. They are listed here so that a reader can resolve each one, with the project file that carries the citation.

Liu, D. and Balz, T. (2026). Deviation-guided dual-model machine learning for SAR coherence change detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 238, 571 to 588. Implemented and tested in section 6.

Nowicki Jessee, M. A., Hamburger, H. W., Allstadt, K. E., Wald, D. J., Robeson, S. M., Tanyas, H., Hearne, M. and Thompson, E. M. (2018). A Global Empirical Model for Near Real-time Assessment of Seismically Induced Landslides. *Journal of Geophysical Research: Earth Surface*, 123, 1835 to 1859. The landslide model behind the USGS ground failure product used in section 2.4.

Voelker, B. and others (2024). *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. The EEFIT Turkey 2023 remote sensing workflow, section IV of which reports the ARIA against CEMS correlation of 0.31 quoted in section 4.1, and sections III-C and V of which supply the field-record and concurrency lessons of section 8.

Zhu, J., Baise, L. G. and Thompson, E. M. (2017). An Updated Geospatial Liquefaction Model for Global Application. *Bulletin of the Seismological Society of America*, 107, 1365 to 1385. The liquefaction model behind the USGS ground failure product used in section 2.4.

Zhu and others (2025). *Earth System Science Data (ESSD)*. The GlobalBuildingAtlas dataset paper that the TUM licence requires to be cited wherever the LoD1 heights of section 7 are used. The project files carry the citation in this short form only, as the licence line states it, and no volume, page range or digital object identifier (DOI) is recorded in the data directory.

Sources: `AGREEMENT\_FRAMEWORK.md`, `LESSONS\_LEARNED.md`,
`LITERATURE\_REVIEW.md`, `PROCESSING\_LOG.md`, `SAR\_SCREEN\_RESULTS.md`,
`catalogue/agreement`, `catalogue/oam.json`, `catalogue/s1\_screen`,
`catalogue/s1\_screen/balz\_ccd/README.md`, `catalogue/s1\_screen/balz\_ccd/models/meta.json`,
`catalogue/s1\_screen/out20/final/test\_burst\_geolocation\_table\_rerun.csv`,
`catalogue/s1\_screen/out20/qa/local\_baseline\_excess\_pereira.csv`,
`catalogue/s1\_screen/out20/qa2/Cali\_qa2\_results.json`,
`catalogue/s1\_screen/out20/qa2/test\_burst\_geolocation.py`,
`catalogue/s1\_screen/out\_experiments/pereira\_features\_README.md`,
`catalogue/s1\_screen/workflow\_s1\_synthesis.js`, `catalogue/s1\_screen/workflow\_variable\_combos.js`,
`catalogue/usgs\_products/usgs\_ground-failure\_info.json`, `report/REPORT\_SPEC.md`, `report/logos`,
`report/objectives/build\_objectives\_brief.py`, `report/objectives/build\_objectives\_excel.py`,
`report/workflows/eeffit-site-content.js`.

10. Credits

Table 10.1 reproduces the credits list of record, which is the credits paragraph of the report specification. For each institution the table gives the data or product credited, the credit line the specification prints, the licence the specification states, and the logo file. That paragraph names 27 institutions or provider groups and 31 logo file names. The logo directory holds 32 files. The one file present but not named in the credits paragraph is `usgs.png`, which carries the United States Geological Survey (USGS) mark for the entry the paragraph credits as public domain without naming a file. One credited body has no logo: the International Organization for Migration, Displacement Tracking Matrix (IOM DTM) is credited in text only.

Two providers whose data entered this report are not in the credits paragraph, and we have added them rather than leave them uncredited. OpenAerialMap appears as the last row of Table 10.1, marked as an addition, and the USGS row carries Landsat Collection 2 Level 2 alongside the event products. Both are sourced in the list below the table, and both are already carried in section 9.5.

Table 10.1 is the table of record for attribution, that is for the institution named, the credit line printed beside its logo, and the logo file itself. Section 9.5 is the table of record for the terms a reader must obey when reusing a product. The two tables are built from the same specification paragraph and do not disagree; where their wording differs, section 9.5 governs reuse and Table 10.1 governs attribution.

Table 10.1

Credits, credit lines, licences and logo files, one row for each entry in the credits list of record plus one addition. Take-away: three inputs the field pack depends on, the Vantor damage layer, the Planet imagery and the GlobalBuildingAtlas heights, carry non-commercial licences, so layers derived from them may be shared inside the mission and published here but not used commercially.

Institution or provider	Data or product credited	Credit line	Licence	Logo file
United States Geological Survey (USGS)	Event, ShakeMap, PAGER and ground failure products (section 2), and Landsat Collection 2 Level 2 scenes (section 3)	None stated in the credits list	Public domain	usgs.png (present in report/logos/, not named in the credits paragraph) 
Servicio Geológico Colombiano	Named in the credits list; the products used are named in the sections that use them	None stated in the credits list	Not stated in the credits list	sgc.png 
Copernicus, European Union and Copernicus Emergency Management Service (CEMS)	CEMS rapid mapping activation EMSR916 and its grading products (sections 3 and 4)	CEMS products © European Union 2026	Not stated in the credits list	copernicus.png, eu.png 
European Space Agency (ESA), Sentinel-1 and Sentinel-2	Copernicus Sentinel data	Copernicus Sentinel data 2026	Not stated in the credits list	esa.png 
NASA, Jet Propulsion Laboratory (JPL) and Indian Space Research Organisation (ISRO), NISAR	NISAR products, provisional	NISAR provisional products via ASF	Not stated in the credits list	nasa.png, jpl.png, isro.png 
OPERA and JPL	OPERA RTC-S1 backscatter products (section 5)	None stated in the credits list	Not stated in the credits list	opera.svg
Alaska Satellite Facility Distributed Active Archive Center (ASF DAAC)	Distribution of the Sentinel-1 and NISAR archives used	None stated in the credits list	Not stated in the credits list	asf.png 
Microsoft AI for	Microsoft damage	None stated in the	CC BY	microsoft.png

Good Lab	layers for Pereira and Cali and their reviewer labels (section 4)	credits list		
Vantor and Maxar Open Data	Vantor open imagery and its building level damage flags (sections 3 and 4)	None stated in the credits list	CC BY-NC 4.0	vantor.png 
Planet Labs PBC Crisis Response	Planet imagery released under crisis response	None stated in the credits list	CC BY-NC 4.0	planet.png 
Airbus and CNES	Pléiades and Pléiades Neo imagery, delivered through the Charter and CEMS	None stated in the credits list	Not stated in the credits list	airbus.png, cnes.png 
International Charter Space and Major Disasters	Charter activation 1048, requester the Unidad Nacional para la Gestión del Riesgo de Desastres (UNGRD)	Activation 1048, requester UNGRD	Not stated in the credits list	charter.png 
UNITAR and UNOSAT	Named in the credits list; the products used are named in the sections that use them	None stated in the credits list	Not stated in the credits list	unitar.png, unosat.png 
Humanitarian OpenStreetMap Team and OpenStreetMap contributors	Named in the credits list; the products used are named in the sections that use them	None stated in the credits list	ODbL	hotosm.png, osm.png 
OCHA and the Humanitarian Data Exchange (HDX)	Named in the credits list; the products used are named in the sections that use them	None stated in the credits list	Not stated in the credits list	ocha.png 
International Organization for Migration, Displacement Tracking Matrix (IOM DTM)	Named in the credits list; the products used are named in the sections that use them	Text credit; logo not obtained	Not stated in the credits list	none (text credit)
Instituto Geográfico Agustín Codazzi (IGAC)	Orthophotos published at emergencia.igac.gov.co	None stated in the credits list	Not stated in the credits list	igac.png 
Unidad Nacional para la Gestión del Riesgo de Desastres (UNGRD)	Named in the credits list; requester of Charter activation 1048	The mark reproduced is the Government of Colombia mark, credited with a note	Not stated in the credits list	ungrd.png 
Overture Maps Foundation	Overture building footprints for the seven cities (sections 4 and 7)	None stated in the credits list	ODbL and CDLA	overture.svg
Google	Google Open Buildings v3	None stated in the credits list	CC BY 4.0	google.png 
Technical University of Munich (TUM)	GlobalBuildingAtlas, Zhu et al. 2025, ESSD, listed in 9.7	None stated in the credits list	CC BY-NC 4.0	tum.png

				
German Aerospace Center (DLR)	World Settlement Footprint (WSF) datasets, credited if used; the WSF datasets were identified as fetchable and were not obtained for this analysis	None stated in the credits list	Not stated in the credits list	dlr.png 
Earthquake Engineering Research Institute (EERI), Learning from Earthquakes and Virtual Earthquake Reconnaissance Team (LFE and VERT)	Named in the credits list; the products used are named in the sections that use them	None stated in the credits list	Not stated in the credits list	eeri.png 
Institution of Structural Engineers (IStructE), Earthquake Engineering Field Investigation Team (EEFIT)	Mission sponsor and report series this document is written for	None stated in the credits list	Not stated in the credits list	istructe.png 
GEM Foundation	Named in the credits list; the products used are named in the sections that use them	None stated in the credits list	Not stated in the credits list	gem.png 
Arup	Named in the credits list; the products used are named in the sections that use them	None stated in the credits list	Not stated in the credits list	arup.png 
University of Houston	Author institution of the remote-sensing group	None stated in the credits list	Not stated in the credits list	uh.png 
OpenAerialMap (addition, not in the credits paragraph)	36 aerial images, including the IGAC emergency orthophotos (section 3)	Provided by OpenAerialMap	CC BY 4.0	none in report/logos/ (text credit)

Source for the institution names, the credit lines, the licence text and the logo file names in Table 10.1: `report/REPORT\_SPEC.md`, section "Credits (logo file in report/logos/, licence note)", the credits paragraph read entry by entry in the order printed there. Source for the presence of each logo file: `report/logos`, directory listing, 32 files, each file name in the table verified present.

Nineteen rows read "Not stated in the credits list" in the Licence column, because the credits paragraph names those institutions and their logos without stating a licence, and no licence has been inferred here.

Sources for the cells that are not in the credits paragraph

Some cells of Table 10.1 carry text that the credits paragraph does not contain. Each is established elsewhere in the project and is sourced here rather than under the source line above.

- USGS, product: catalogue/usgs_products, directory listing, files usgs_event.json, usgs_shakemap_info.json, usgs_losspager_alerts.json, usgs_losspager_exposures.json, usgs_losspager_losses.json and usgs_ground-failure_info.json, the USGS products used in section 2; catalogue/imagery_acquisitions_timeline.csv, column "program", value "USGS Landsat C2 L2 (Planetary Computer)", the Landsat scenes used in section 3

- Copernicus, European Union and CEMS, product: catalogue/cems_api.json, results[0].activationTime and results[0].aois, the EMSR916 rapid mapping activation and its seven areas of interest used in sections 3 and 4
- NASA, JPL and ISRO, product: catalogue/s1_screen/out_nisar, directory listing, the NISAR coherence and damage proxy products used in section 5, all provisional
- OPERA and JPL, product: SAR_SCREEN_RESULTS.md, section 2 "OPERA RTC-S1 backscatter change (30 m, VV, gamma0)" and section 4 "OPERA status (asked)", the OPERA products available and used
- ASF DAAC, product: catalogue/asf_s1.csv, column "URL", all values on datapool.asf.alaska.edu, the Sentinel-1 archive distribution; catalogue/s1_screen/out_nisar, directory listing, the NISAR products obtained through the same archive
- Microsoft AI for Good Lab, product: AGREEMENT_FRAMEWORK.md, section "Pereira, Microsoft vs Vantor", paragraph "Reviewer labels (MS)", the Pereira damage layer and its reviewer labels; section "Cali, Microsoft vs CEMS", the Cali damage layer
- Vantor and Maxar Open Data, product: catalogue/vantor_items.json, list length 22, the open imagery items; AGREEMENT_FRAMEWORK.md, section "Pereira, Microsoft vs Vantor", the Vantor building level damage flags
- International Charter, product: the expansion of UNGRD is editorial and is flagged below; the activation number and the requester are in the credits paragraph.
- Overture Maps Foundation, product: catalogue/footprints, directory listing, files overture_buildings_<City>.gpkg for the seven cities; catalogue/s1_screen/out_heights/height_coverage.csv, column "n_overture", the footprint counts these products supply to sections 4 and 7
- DLR, product: catalogue/s1_screen/out_zerohour/zero_hour_inventory.md, table rows "DLR World Settlement Footprint 3D (WSF3D, building volume/height)" and "DLR WSF Evolution (built-up extent time series 1985-2015)", availability column value "No", both listed as fetchable and neither obtained
- IStructE and EEFIT, product: report/REPORT_SPEC.md, line 1 title "Report specification, EEFIT Colombia 2026 remote-sensing support" and line 3 deliverable name report/EEFIT_Colombia_2026_RemoteSensing_Report.docx, the report series and mission record this document belongs to
- University of Houston, product: report/REPORT_SPEC.md, section "Site style (taken from the lab theme milillo-2026/style.css)", the University of Houston lab site this report is styled after, the remote-sensing group's institution
- OpenAerialMap, product, credit line and licence: catalogue/oam.json, meta.found 36, meta.provided_by "OpenAerialMap" and meta.license "CC-BY 4.0", the aerial images and their licence, printed in the table in the spaced form the report uses

One class of cell is deliberately absent from this list. The Planet row rewords the specification's own entry "Planet Labs PBC Crisis Response" and adds nothing to it, so the blanket source line above covers it.

Use of the logos

The logos are reproduced for attribution only, and two heights govern them, one for each deliverable. In the report site the specification keeps each logo at a height of at most 48 px in a greyscale on hover strip. In the Word document the report format specification sets each logo at 1.2 cm high beside its licence line. The two constants govern different outputs and are not conversions of one another: they coincide only at one particular screen resolution, and the Word document is built at print resolution, so the same logo file is printed at a different physical size in the two deliverables. The strip is assembled at build time from the files in `report/logos/`.

Reproduction of a logo in this credits section is an attribution, not a claim of endorsement of this report by the institution concerned, and it does not extend any licence beyond the one printed in Table 10.1. Where the Licence cell of a row states a licence, a reader reusing that product reproduces that licence line. Where the Licence cell reads "Not stated in the credits list", the credits paragraph states no licence and none has been inferred here: take the terms of reuse from section 9.5, and where section 9.5 states none either, ask the provider before reusing the product outside the mission.

Sources: `AGREEMENT\_FRAMEWORK.md`, `SAR\_SCREEN\_RESULTS.md`, `catalogue/asf\_s1.csv`, `catalogue/cems\_api.json`, `catalogue/footprints`, `catalogue/imagery\_acquisitions\_timeline.csv`, `catalogue/oam.json`, `catalogue/s1\_screen/out\_heights/height\_coverage.csv`, `catalogue/s1\_screen/out\_nisar`, `catalogue/s1\_screen/out\_zerohour/zero\_hour\_inventory.md`, `catalogue/usgs\_products`, `catalogue/vantor\_items.json`, `report/EEFIT\_Colombia\_2026\_RemoteSensing\_Report.docx`, `report/REPORT\_SPEC.md`, `report/logos`.

11. Zero hour damage estimate without post event remote sensing

A zero hour product is one that can be issued minutes to hours after an earthquake, from data that exist before any satellite has flown over the epicentre. We built one for Pereira and Dosquebradas and tested it against controls that a rapid response team can check for itself. It ranks 250 m neighbourhoods usefully and it does not identify damaged buildings: within one city its building level score is dominated by the position of the building, its margin over a model given nothing but map coordinates is small and changes sign depending on which negatives and which spatial blocking are used, and it does not carry to a second city. The tables below give the evidence and the section closes with what this does and does not change for the 15 to 26 September field programme.

Two classes of result in this section are provisional and are labelled where they appear. All NISAR (NASA ISRO Synthetic Aperture Radar) layers are provisional version 1 products. The transfer and null model results come from single day runs whose two verification passes did not complete, so they carry the confidence of a single run and are labelled "single run" throughout. Every label used as truth here is a desk review of optical imagery, not a field observation.

Acronyms used in this section at first occurrence: AUC (area under the curve); ROC AUC (area under the receiver operating characteristic curve, 0.5 is chance); PR AUC (area under the precision recall curve, whose chance value is the base rate and therefore differs between the two negative sets used here); MCC (Matthews correlation coefficient); CV (cross validation); XGBoost (a gradient boosted decision tree classifier, referred to as xgb in the result tables); MMI (Modified Mercalli Intensity); PGA (peak ground acceleration); PGV (peak ground velocity); PSA (pseudo spectral acceleration); Vs30 (time averaged shear wave velocity in the top 30 m); GBA (GlobalBuildingAtlas); GHS (Global Human Settlement layer); LoD1 (level of detail 1); CEMS (Copernicus Emergency Management Service); PAGER (Prompt Assessment of Global Earthquakes for Response); DYFI (Did You Feel It); FFM (finite fault model);

UTM (Universal Transverse Mercator); DEM (digital elevation model); SAR (synthetic aperture radar); EEFIT (Earthquake Engineering Field Investigation Team).

11.1 What exists at time zero

Two families of layer are available with no post event acquisition. The static family is the building and site inventory: Overture and Google Open Buildings footprints for all seven cities, GBA LoD1 heights, Copernicus GLO-30 elevation and slope, GHS built height and population at 100 m, the GEM global exposure model for Colombia, and the 1999 Quindio typology baseline for Armenia and Quimbaya. The event triggered family becomes available within minutes to hours of the origin and needs no imagery: the USGS origin solution and moment tensor, ShakeMap with MMI, PGA, PGV and spectral accelerations together with their standard deviations, the Vs30 grid that ShakeMap itself used, PAGER exposure and losses, the ground failure liquefaction and landslide probability grids, the FFM, and DYFI intensities. The full inventory records the release latency, the on disk status and the download path for each layer, and it is explicit that several of these layers are fetchable rather than already held: the ground failure probability grids, the Vs30 raster and the GEM exposure model were not on local disk at the time of writing.

Height is the one static attribute whose coverage splits the seven cities, and the split depends on which column is read. GBA LoD1 heights cover 90.8 percent of footprints in Pereira and Dosquebradas, 96.1 percent in Cali and 73.8 percent in Quimbaya, and essentially none of the other four cities. Any height source at all, which adds the Overture floor counts converted at a fixed metres per floor, is a different picture in Quibdo: 50.4 percent of its 34,066 footprints carry a height, supplied by 17,174 floor counts. Armenia, Manizales and Sevilla remain near zero on both columns.

City	Overture footprints	GBA height (frac_gba_height)	Any height (frac_any_height)
Pereira and Dosquebradas	91,266	90.8 %	92.5 %
Cali	158,489	96.1 %	96.3 %
Quimbaya	7,707	73.8 %	73.8 %
Quibdó	34,066	0.0 %	50.4 %
Armenia	106,040	0.0 %	4.1 %
Manizales	67,014	0.0 %	1.9 %
Sevilla	6,726	0.0 %	0.2 %

Source: 'catalogue/sl_screen/out_heights/height_coverage.csv', columns 'n_overture', 'frac_gba_height' and 'frac_any_height', one row per city.

Take-away: a zero hour feature vector can be assembled today for all seven cities, but a metric height enters it only in Pereira and Dosquebradas, Cali and northern Quimbaya, and in Quibdo the usable height attribute is a floor count on half the stock rather than nothing at all.

11.2 Tiers, protocol and the design constants

Four tiers were scored on the Pereira and Dosquebradas labels under the protocol used for the remote sensing experiments of sections 5 to 7: 5 fold cross validation with folds blocked on 500 m cells, logistic regression and shallow XGBoost, and 368 positives scored against either 358 reviewer rejected buildings or a sample of 3,680 unflagged buildings.

Tier A is static data only, 40 features: height, height class, floor count, footprint area, compactness, elongation, neighbour counts at 50, 100 and 250 m, built density at 100 and 250 m, elevation, slope, GHS built height and population, and the Overture class and subtype one hot columns. Tier B adds the event layers, 64 features: the full ShakeMap set with its uncertainties, Vs30, five ground failure models and four source distance metrics. Tier C adds the pre event radar archive, 69 features. Tier D is the post event remote sensing reference carried over from sections 5 and 6.

Tier C is not a zero hour tier. Its five additional features are `s1\_pre\_mean\_late`, `s1\_at\_pre\_mean`, `nisar\_pre\_T047A`, `bz\_s1\_pre\_coh\_ml\_med` and `bz\_s1\_pre\_deviation\_med`, all of them Sentinel-1, NISAR or Liu and Balz derived layers. Tier C is therefore available with no post event image and is not available with no remote sensing at all, and the headline of this section stays on tier B.

Tier	Data needed	ROC AUC (rejected)	MCC (rejected)	ROC AUC (unflagged)
A, static only	nothing post event	0.8723	0.6282	0.8273
B, static plus event	ShakeMap at T plus 30 min	0.8891	0.6501	0.8642
C, plus pre event radar archive	pre event radar archive, not a zero hour tier	0.8921	0.6670	0.8638
D, NISAR coherence drop, single feature	post event imagery at T plus 1 to 3 days	0.8062	0.5417	0.8345
D, Sentinel-1 plus NISAR plus covariates, 12 features	post event imagery at T plus 1 to 3 days	0.8719	0.6356	0.8944
Null, x and y coordinates alone	nothing	0.8739	0.5922	0.8425

Source: `catalogue/s1\_screen/out\_zerohour/zero\_hour\_results.csv` (tier rows) and `null\_models.csv` (null row). All tier rows are the XGBoost classifier except the two D rows, which are the classifier reported in sections 5 and 6: logistic regression for the rejected column and, for D with 12 features, XGBoost in the unflagged column. The two 12 feature D rows carry `n\_gt` and `n\_blocks` as not a number, so no ground truth count is quoted beside them.

Both post event references must be read together. Against reviewer rejected negatives the single feature NISAR drop, at 0.8062, sits below every zero hour tier, which is the reading that would support issuing a zero hour product and not waiting. The 12 feature post event stack reaches 0.8944 against unflagged negatives, the highest ROC AUC anywhere in the file and above tier C at 0.8921, which is the reading that would support waiting. Any sentence comparing zero hour against "post event" therefore has to name which post event product it means.

Three design constants govern these numbers and all three were varied. The 500 m blocking scale is the reported one. At 2 km blocking, tier B XGBoost falls from 0.8891 to 0.8780 on rejected negatives and its PR AUC falls from 0.4763 to 0.3250 on unflagged negatives, a relative drop of 32 percent, with the number of blocks falling from 209 to 51 on the rejected set. A 1 km scheme at 124 groups gives tier B XGBoost 0.8701 on rejected negatives. Random folds give 0.9301, but with 94.0 percent of test positives sharing a 500 m block with a training positive, and the coordinate null itself inflates to 0.9117 under the same scheme. Quoting the random fold figure would overstate tier B by about 0.04 AUC and would overstate the coordinate null by about the same amount, which is the reason blocked folds are used and 500 m is the scale reported. The 250 m cell size of the operational product and the 5 percent top fraction are reported alongside their alternatives in 11.4 and 11.5, and the 1:1 class rebalancing is carried explicitly into the count correction in 11.6.

Take-away: tier B is the tier worth discussing, its headline of 0.8891 on rejected negatives is a 500 m blocked figure that falls to 0.8780 at 2 km blocking and inflates to 0.9301 under leaky random folds, and it must always be compared against a named post event product rather than against "post event imagery" in the abstract.

11.3 What the null models show

Four tests bear directly on whether the tier B score is anything more than a location prior, and they do not all point the same way.

First, the coordinate null. A model given nothing but the UTM x and y of each building reaches ROC AUC 0.8739 against rejected negatives and 0.8425 against unflagged negatives, while distance to the city centre alone reaches 0.8392 and 0.8220. The random feature control sits at 0.4739 and the neighbour count null at 0.5286, so the coordinate result is a real spatial effect and not an artifact of the evaluation.

Second, the margin of tier B over the coordinate null, with intervals. On rejected negatives at 500 m blocking the difference is 0.0151 for XGBoost, with a bootstrap interval of [-0.0154, 0.0522] and a fraction of bootstrap replicates at or below zero of 0.1605 over 2,000 resamples. At 1 km blocking the same comparison gives -0.0039 with interval [-0.0436, 0.0379] and a fraction of 0.5525, which is a tie, and the logistic regression row at 1 km leans to the null at -0.0268 with a fraction of 0.8610. On unflagged negatives the ordering reverses for XGBoost, which gains 0.0217 at 500 m and 0.0203 at 1 km with fractions of 0.0970 and 0.0680, while the logistic regression rows give 0.0077 and -0.0100. The sign of the margin depends on the negative set and on the blocking scale.

Third, signal that is not location. A feature set built from building attributes with no coordinates and no ShakeMap reaches 0.7695 on rejected negatives, and adding the 64 tier B features on top of the two coordinates gains 0.0183 over the coordinates alone. Removing the four explicit source distance features from tier B costs essentially nothing: 0.8910 against 0.8893 for logistic regression on rejected negatives, and 0.8584 against 0.8642 for XGBoost on unflagged negatives, at 60 features against 64.

Fourth, the test that removes location by construction rather than by adjustment. Comparing positives against negatives only inside the same 500 m block, the pooled discrimination of the density and environment features collapses: 'built_density_100m' falls from a pooled ROC AUC of 0.7558 to 0.4865 and 'ghs_built_h_m' from 0.8166 to 0.6498, on 35 mixed blocks and 631 within block pairs. Footprint area survives better, 0.6756 pooled against 0.5832 within blocks, and on unflagged negatives it does not fall at all, 0.6115 pooled against 0.6177 within 71 mixed blocks and 12,705 pairs.

Feature	Pooled ROC AUC (rejected)	Within 500 m block	Mixed blocks	Pairs
'built_density_100m'	0.7558	0.4865	35	631
'built_density_250m'	0.7933	0.4168	35	631
'ghs_built_h_m'	0.8166	0.6498	35	631
'mmi'	0.8113	0.5214	35	631
'area_m2'	0.6756	0.5832	35	631

Source: 'catalogue/sl_screen/out_zerohour/refute_locationproxy/refute_within_block_auc.csv', rows with negatives 'rejected', columns 'auc_pooled', 'auc_within_500m_block', 'n_blocks_mixed_500m', 'n_pairs_500m'. Message: the density, built height and shaking features lose most or all of their discrimination once positives and negatives are compared only within the same 500 m block, while footprint area loses least, so most of the pooled skill of these features is a statement about where a building is rather than about the building.

Fig. 34 shows the same tier under random and blocked cross validation, where the margin over the coordinate null narrows as the blocks grow. The strongest single feature in tier B illustrates the same point and cannot be presented as a shaking covariate. 'mmi_std', the ShakeMap MMI standard deviation, has the largest permutation importance in tier B logistic regression, a mean ROC AUC drop of 0.0841 with a standard deviation of 0.0060. Its raw pooled AUC against rejected negatives is 0.1311, that is, it predicts in the inverted direction, giving a sign agnostic AUC of 0.8689, and its Spearman correlation with distance from the city centre is 0.7748. The physical reading is that ShakeMap uncertainty is lowest near the instrumented core, so a low 'mmi_std' marks the core; the feature enters the model as a location proxy.

Distance from the city centre carries this argument, so the centre must be defined. It is the density peak of building centroids on a 500 m grid, at UTM 423750.0, 531750.0, or longitude -75.68763 and latitude 4.81045. Against that centre the four label families sit at very different distances, so the concentration

argument is not one statement but four. The 368 training positives have a median distance of 1,395.5 m with 44.3 percent within 1 km; the 136 ground truth buildings sit at 432.7 m with 75.0 percent within 1 km; the 76 CEMS destroyed or damaged only points sit at 816.9 m with 57.9 percent within 1 km; and the 156 confirmed buildings sit at 2,276.7 m with only 10.9 percent within 1 km, against 5,191.0 m and 6.3 percent for all 91,266 footprints.

Take-away: the tier B building score is dominated by position, its margin over a coordinate model is small and changes sign with the negative set and the blocking scale, a location free building feature set still reaches 0.7695 so location is not the only signal, and the core concentration that drives all of this is a property of the two most concentrated label families and not of the confirmed family.

11.4 Which negative set, and what a top 5 percent list buys a field team

Every headline in this section depends on which negatives it is scored against, and the two sets have different base rates, so their PR AUC values must never be compared as a ratio. Tier B XGBoost gives PR AUC 0.8846 on 726 evaluated footprints, 368 positives against 358 reviewer rejected negatives, and 0.4763 on 4,048 evaluated, 368 positives against 3,680 unflagged negatives. Those sit on base rates of 0.5069 and 0.0909. Expressed as lift over base rate the two values are 1.75 and 5.24, which is the opposite ordering to the raw PR AUC values. The unflagged set is the operational one, because a deployed model scores every footprint in the city and not a curated pair of reviewed classes.

Precision without recall is the number most likely to be misused by a survey planner, so the two travel together here. On the reviewer rejected universe, taking the top 5 percent of the ranked list means 37 buildings, of which 35 are positives: precision 0.9459, lift 1.866, recall 0.0951 and ground truth recall 0.1397. On the operational unflagged universe the same 5 percent is 203 buildings, of which 118 are positives: precision 0.5813, lift 6.394, recall 0.3207 and ground truth recall 0.4632.

Universe	Top fraction	k	Precision	Lift over base rate	Recall of 368 positives	Recall of 136 ground truth
labelled rejected	5 %	37	0.9459	1.866	0.0951	0.1397
labelled rejected	10 %	73	0.9589	1.892	0.1902	0.2279
labelled unflagged	5 %	203	0.5813	6.394	0.3207	0.4632
labelled unflagged	10 %	405	0.4741	5.215	0.5217	0.6691

Source: `catalogue/sl\_screen/out\_zerohour/operational\_value\_topk.csv`, rows `labelled\_rejected,B|xgb` and `labelled\_unflagged,B|xgb`, `top\_frac` 0.05 and 0.1, columns `k`, `precision`, `lift`, `recall`, `gt\_recall`. Message: on the curated rejected universe the top 5 percent looks almost perfect and holds under a tenth of the damaged buildings, and on the operational universe the same list is right slightly more than half the time and holds a third of them, so a survey planned from precision alone would be planned wrongly.

Take-away: report the negative set with every headline, compare the two sets by lift over base rate and never by a ratio of PR AUC values, and never state a precision without the recall beside it.

11.5 Ranking buildings and 250 m cells against the post event layers

The question a field team actually faces is whether a zero hour ranking is worth acting on before the post event imagery lands. Two tests answer it directly, one at building level and one at cell level, both on the 74,999 footprints of the Microsoft valid area in Pereira and Dosquebradas.

Fig. 33 plots precision and recall against the fraction of the list taken. At building level, taking the top 10 percent of the ranked list, tier B recovers 60.1 percent of the 368 strict positives, the coordinate null recovers 46.7 percent, the post event NISAR coherence drop recovers 61.4 percent and the post event

Sentinel-1 drop recovers 47.8 percent. The paired differences carry bootstrap intervals. Tier B beats the coordinate null by 0.1332 on the strict positives, with a bootstrap mean of 0.1161, interval [0.0124, 0.2280] and a fraction of replicates at or below zero of 0.016, and by 0.1218 on the 616 union positives, with bootstrap mean 0.0991, interval [0.0232, 0.1816] and fraction 0.007. Against the NISAR drop it does not win: the difference is -0.0136 on the strict positives with fraction 0.600, and 0.0519 on the union positives with fraction 0.249. Tier B also beats tier A by 0.0897 on the strict positives with fraction 0.017, so the ShakeMap and ground failure layers do add to the static inventory even where they do not add over the coordinates.

At cell level the ranking is much stronger for every scorer, and it is stronger for the trivial baseline too. Against the 34 CEMS destroyed cells among 3,809 cells of 250 m holding at least one building, ranking cells by building count alone gives a cell AUC of 0.9016, the post event Sentinel-1 damage proxy gives 0.9251 and the post event NISAR proxy gives 0.8182, while tier B summed over the cell gives 0.9817 and tier B mean over the cell on rejected negatives gives 0.9874. Requiring at least 10 buildings per cell reduces the field to 2,122 cells and separates the model from the baseline more clearly, because the building count baseline falls to 0.8221 while tier B summed stays at 0.9669.

Cell ranking (250 m, at least 1 building, 3,809 cells)	Cell AUC, 34 destroyed cells against the rest
Building count alone	0.9016
Post event NISAR coherence drop (provisional)	0.8182
Post event Sentinel-1 coherence drop	0.9251
Coordinate null, mean over cell	0.9602
Tier A, mean over cell	0.9564
Tier B, summed over cell	0.9817
Tier B, mean over cell (rejected negatives)	0.9874

Source: `catalogue/sl\_screen/out\_zerohour/operational\_cems\_destroyed\_cells.csv`, rows with `min\_bld` 1, column `cell\_auc\_destroyed\_vs\_rest`. Message: at 250 m cell scale every zero hour scorer, and the coordinate null with it, separates the destroyed cells better than either post event coherence layer does, and the trivial building count baseline already reaches 0.9016, so the cell product is worth issuing and the margin it holds over counting buildings is the part that has to be defended.

Take-away: at building level tier B beats a coordinate model on the operational universe but does not beat the post event NISAR drop, and at 250 m cell level every scorer including the coordinate null outperforms both post event layers while a building count baseline already reaches 0.9016, so the deliverable is a ranked cell list and the honest claim for it is speed and coverage rather than discrimination.

11.6 A citywide count, and why it must not be issued as one number

Summing the model probability over all 91,266 Pereira and Dosquebradas footprints gives a raw total of 11,505.9, which is an artifact of the 1:1 class rebalancing used in training and must never be quoted alone. Correcting for that rebalancing, under an assumed true positive odds of 368 over 74,111, gives 172.9.

The corrected value is not one number either. Under the same odds the six scorers give 172.9, 204.5, 238.2, 266.0, 510.9 and 1,095.1, a spread of 6.3 to 1 that is far larger than the correction applied to any one of them. And 172.9 sits below four counts already observed in the same file: 292 for the Microsoft ground truth plus confirmed classes, 441 for the Microsoft Pereira extended set, 446 for the Microsoft damaged flag matched locally, and 625 for the union of Microsoft, CEMS destroyed or damaged, and Vantor labels. The comparison is not like for like: the reference counts pool damage grades from several products, while the model estimate is of the 368 strict positive class only. That is exactly why the comparison has to be stated rather than left out, because a single citywide figure of 173 damaged buildings placed in front of structural engineers would sit below the damage already catalogued.

Take-away: no single citywide count is issued from this work; the corrected estimates span 172.9 to 1,095.1 across scorers, they are estimates of the strict positive class alone, and four already observed catalogues of pooled damage grades exceed the lowest of them.

11.7 Transfer to Cali, and back

Fig. 32 draws the transfer in both directions. Trained on all Pereira and Dosquebradas labels and applied to Cali against the Microsoft damaged flags, 256 positives among 92,597 evaluated footprints, the 256 being the flags of the released Cali layer's 266 that fall inside the evaluated valid area, the zero hour model performs at chance: ROC AUC runs 0.4577 to 0.5271 across the eight tier and classifier combinations. Lift over the base rate in the top 10 percent runs 0.8594 to 1.4843 across the same eight rows, the lowest of them being the event only XGBoost row, which is below the base rate. At 250 m cell level the Spearman correlation of the model against the label density runs 0.0874 to 0.2372 across those rows, against 0.2337 for a null that ranks cells by building count alone, so the cell agreement that does survive the transfer is explained by counting buildings.

Against the eight CEMS destroyed or damaged footprints in Cali the AUC values are higher, running 0.5434 to 0.8645 across the same eight combinations. With eight positives that is not evidence, and three checks say so. First, a permutation test on the same eight positives clears one sided p 0.05 for four of the eight scores, not one: 0.8645 at p 0.0000, 0.7210 at p 0.0115, 0.6899 at p 0.0325 and 0.6698 at p 0.0445. Second, the bootstrap interval on the best of them is [0.7211, 0.9877], on eight positives. Third, footprint area on its own, with no model at all, reaches 0.8001 on the same eight positives, which is 93 percent of the best model score. On the 256 positive Microsoft label set the same permutation test gives 0.4577 at two sided p 0.0195, significantly below chance, and 0.5015 at p 0.9285.

Trained inside Cali on blocked folds rather than transferred, the model reaches 0.6058 to 0.6699 across the eight rows, on 256 positives. The reverse direction, trained on Cali and applied to Pereira and Dosquebradas against all 74,837 negatives, gives 0.6065 for tier A and 0.6945 for tier B, and both of those are the better classifier of the two: the other classifier gives 0.4378 for tier A and 0.4869 for tier B, below chance in both cases. The tier B figure of 0.6945 comes from an extrapolated distance and intensity gradient rather than from building properties, since removing the distance features drops it to 0.4805. Trees do not extrapolate the shifted feature ranges at all, giving 0.2966.

Take-away: the model learns one city's spatial pattern of reported damage and carries none of it to the next city, the one apparently positive transfer rests on eight positives where a single static field reaches 93 percent of the score, and any number quoted from this subsection must carry the range across classifiers rather than the better half of it.

11.8 What this changes for 15 to 26 September

Section 11 changes nothing about the field programme of 15 to 26 September 2026. The zero hour product is a pre event and minutes to hours capability whose value is realised before a mission is mobilised, and the cities, the sampling frames and the priority lists for this mission are set by sections 5 to 8 from post event observations. Two concrete uses follow from it nonetheless, and both are preparatory rather than field actions.

1. Pereira and Dosquebradas: the 250 m cell ranking in ``catalogue/sl_screen/out_zerohour/operational_cells_250m.csv`` and the destroyed cell test in ``operational_cems_destroyed_cells.csv`` are the products to carry as a cross check on the survey frame, not as a target list. Action: before the survey blocks are fixed, confirm that the cells the team plans to visit include the top 10 percent of that ranking, and record any top ranked cell that the post event products do not also flag, since those cells are where a zero hour product and an imagery based product disagree.

2. All seven cities: the three zero hour layers that were fetchable but not held on disk, the ground failure probability grids, the Vs30 raster and the GEM exposure model for Colombia, are named with their download paths in `catalogue/s1\_screen/out\_zerohour/zero\_hour\_inventory.md`. Action: pre stage them for Colombia before the next event, because a capability that must download them after the origin time is not a minutes to hours capability.

Take-away: nothing in this section moves a team or changes a city allocation for this mission, and the one operational instruction it produces is to pre stage three named layers so that the capability tested here can actually run at T plus 30 minutes next time.

Key points

1. A zero hour model built only from data available before any post event image reaches ROC AUC 0.8891 on reviewer rejected negatives and 0.8642 on the operational unflagged negatives in Pereira and Dosquebradas (11.2), which is a real result and not a chance one, since a random feature control sits at 0.4739 (11.3).
2. A model given nothing but the map coordinates of each building reaches 0.8739 and 0.8425 on the same two sets, and the tier B margin over it is 0.0151 at 500 m blocking and -0.0039 at 1 km blocking on rejected negatives, so the building level score is dominated by position (11.3).
3. Location is not the only signal: a building attribute set with no coordinates and no ShakeMap reaches 0.7695, and adding the tier B features on top of the coordinates gains 0.0183 (11.3).
4. The two post event references disagree about whether waiting is worth it. The single feature NISAR drop sits below every zero hour tier at 0.8062, while the 12 feature post event stack reaches 0.8944, the highest value in the file (11.2). At building level the NISAR drop also recovers more strict positives in the top 10 percent than tier B does, 61.4 percent against 60.1 percent (11.5).
5. At 250 m cell scale every scorer separates the 34 CEMS destroyed cells well, tier B at 0.9817 and the coordinate null at 0.9602, but a building count baseline already reaches 0.9016, so the cell product must be defended against counting buildings and not against chance (11.5).
6. Precision and recall must travel together: the top 5 percent of the operational ranking is right 58.1 percent of the time and contains 32.1 percent of the positives (11.4).
7. No citywide count is issued. The prior corrected estimates span 172.9 to 1,095.1 across six scorers, they estimate the strict positive class alone, and four already observed catalogues of pooled damage grades run from 292 to 625 (11.6).
8. The model does not transfer. Pereira to Cali runs 0.4577 to 0.5271 against 256 Microsoft positives, and the one apparently positive result, 0.5434 to 0.8645 against eight CEMS positives, is matched to 93 percent by footprint area alone (11.7).
9. Blocking scale decides the headline. Tier B XGBoost gives 0.8891 at 500 m, 0.8701 at 1 km and 0.8780 at 2 km on rejected negatives, and 0.9301 under random folds where 94.0 percent of test positives share a block with a training positive (11.2).
10. The section changes nothing about the 15 to 26 September field programme. Its one operational instruction is to pre stage the ground failure grids, the Vs30 raster and the GEM exposure model for Colombia before the next event (11.8).

Figures

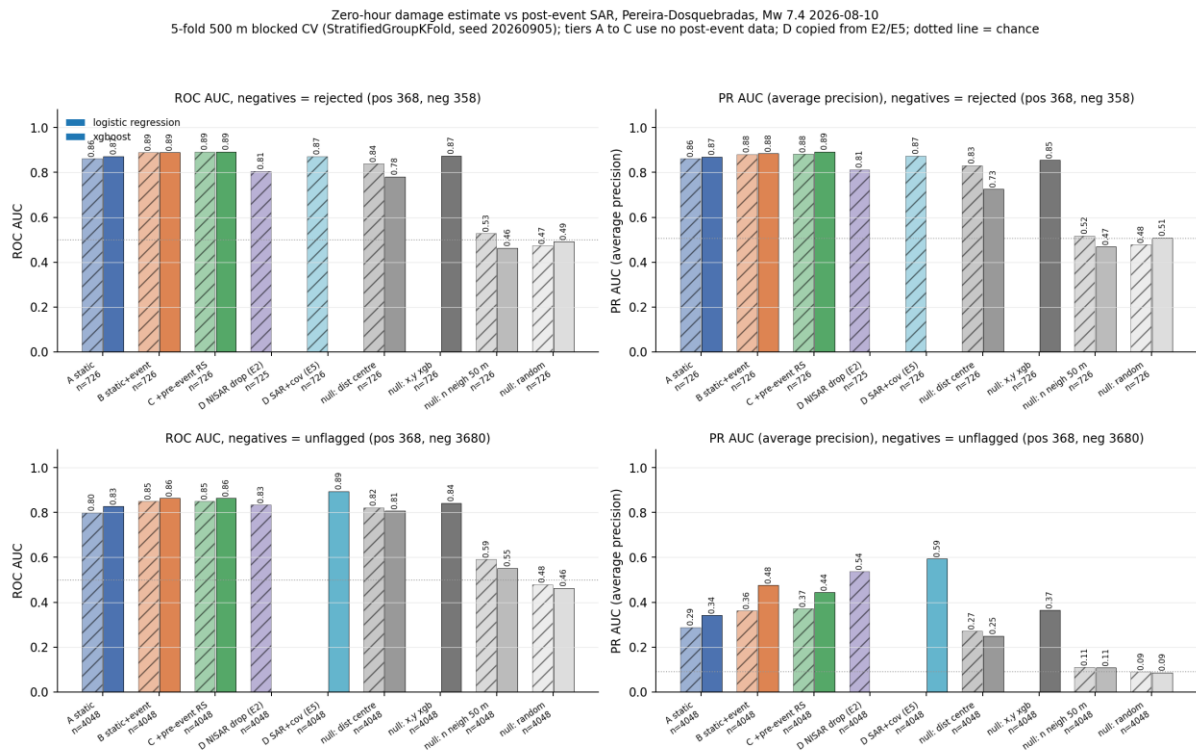


Fig. 31. zero hour tiers A, B and C against the post event remote sensing references and the coordinate null, ROC AUC on both negative sets, Pereira and Dosquebradas. Data providers: USGS (ShakeMap, PAGER, ground failure, finite fault); Overture Maps Foundation; Google Open Buildings; TUM GlobalBuildingAtlas; European Commission Joint Research Centre GHSL; ESA Copernicus (GLO-30 DEM, Sentinel-1); NASA, JPL and ISRO NISAR provisional products via the Alaska Satellite Facility; Microsoft AI for Good Lab; Copernicus Emergency Management Service. Message: the zero hour tiers sit between the two post event references and above the coordinate null by a margin smaller than the gap between the two post event references themselves. Source: ``catalogue/sl_screen/out_zerohour/zero_hour_results.csv`` and ``null_models.csv``.

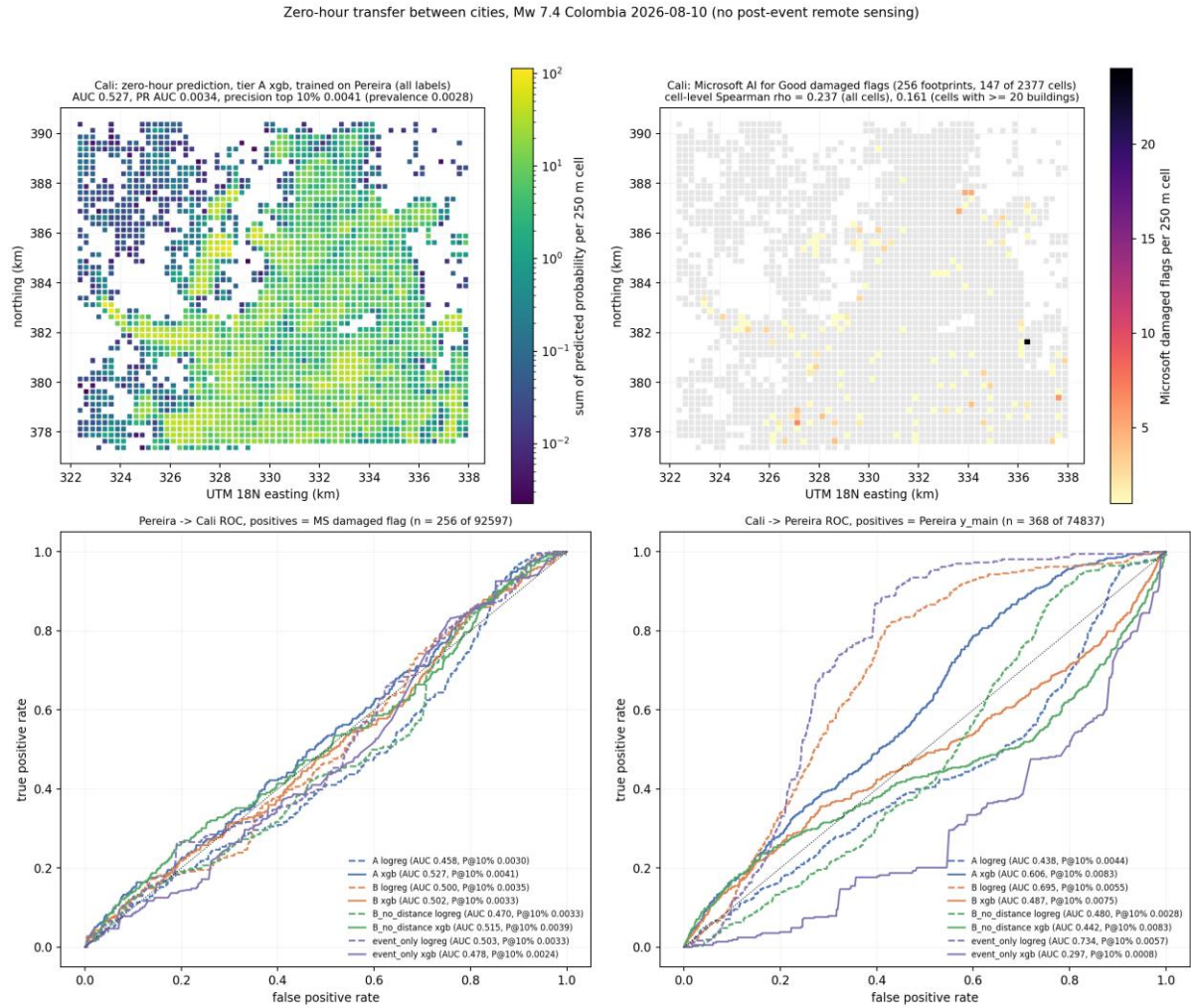


Fig. 32. transfer of the zero hour model between Pereira and Dosquebradas and Cali, ROC AUC and top 10 percent lift by tier, classifier and label family. Data providers: as Fig. 31. Message: every transfer direction and tier sits at or near chance against the 256 positive Microsoft label set, and the only values above chance rest on eight CEMS positives. Source: ``catalogue/s1_screen/out_zerohour/transfer_results.csv`` and ``refute_locationproxy/refute_transfer_permutation.csv``.

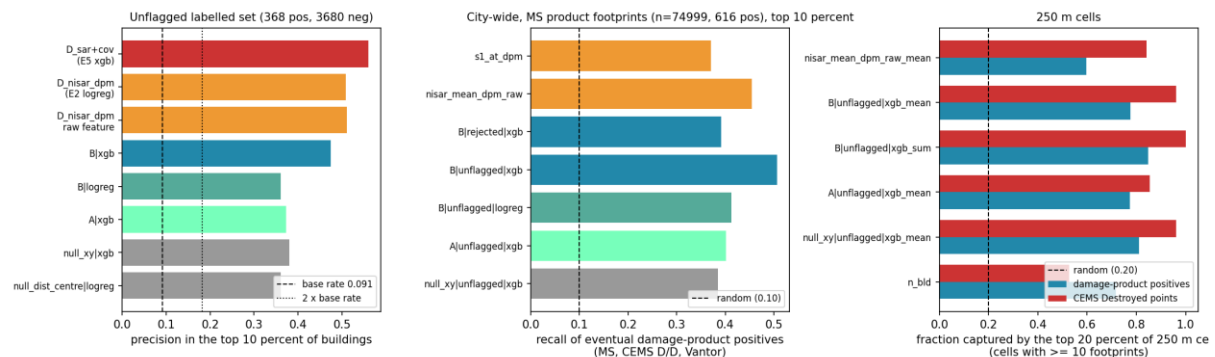


Fig. 33. operational value of the ranked list, precision and recall against top fraction for tier B on both universes, with the coordinate null and the post event layers. Data providers: as Fig. 31. Message: the top of the list is precise on the curated universe and thin on recall, and on the operational universe the same list

trades precision for the recall a survey actually needs. Source:
`catalogue/s1\_screen/out\_zerohour/operational\_value\_topk.csv`.

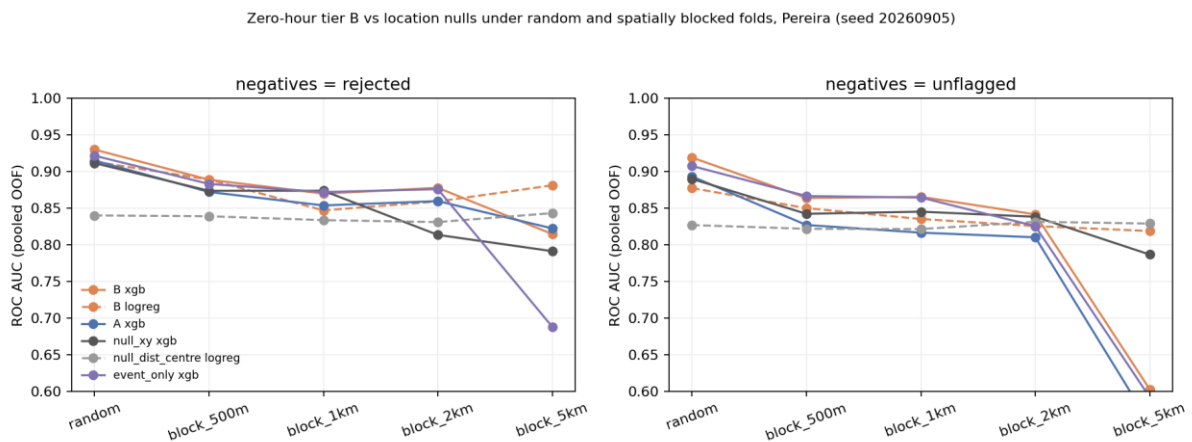


Fig. 34. ROC AUC of tier B and of the coordinate null under random, 500 m blocked, 1 km blocked and 2 km blocked cross validation, with the fraction of test positives sharing a block with a training positive. Data providers: as Fig. 31. Message: the apparent skill of tier B under random folds is a leakage artifact that inflates the coordinate null by the same amount, which is why the 500 m blocked figure is the one reported. Source: `catalogue/s1\_screen/out\_zerohour/refute\_locationproxy/refute\_fold\_schemes.csv` and `zero\_hour\_results.csv`.

Reproducibility constants for this section, the random seed, the fold construction, the bootstrap resample count and the five scripts that produced these tables, are stated once in section 9 and are not repeated here.

Sources: `catalogue/s1\_screen/out\_experiments/build\_pereira\_features.py`,
`catalogue/s1\_screen/out\_experiments/pereira\_features\_summary.json`,
`catalogue/s1\_screen/out\_heights/height\_coverage.csv`,
`catalogue/s1\_screen/out\_zerohour/feature\_importance.csv`,
`catalogue/s1\_screen/out\_zerohour/null\_models.csv`,
`catalogue/s1\_screen/out\_zerohour/operational\_aggregate\_counts.csv`,
`catalogue/s1\_screen/out\_zerohour/operational\_bootstrap\_top10\_recall.csv`,
`catalogue/s1\_screen/out\_zerohour/operational\_cells\_250m.csv`,
`catalogue/s1\_screen/out\_zerohour/operational\_cems\_destroyed\_cells.csv`,
`catalogue/s1\_screen/out\_zerohour/operational\_value\_topk.csv`,
`catalogue/s1\_screen/out\_zerohour/refute\_locationproxy/refute\_B\_vs\_null\_bootstrap.csv`,
`catalogue/s1\_screen/out\_zerohour/refute\_locationproxy/refute\_fold\_schemes.csv`,
`catalogue/s1\_screen/out\_zerohour/refute\_locationproxy/refute\_label\_core\_concentration.csv`,
`catalogue/s1\_screen/out\_zerohour/refute\_locationproxy/refute\_shakemap\_variability.csv`,
`catalogue/s1\_screen/out\_zerohour/refute\_locationproxy/refute\_size\_prior.csv`,
`catalogue/s1\_screen/out\_zerohour/refute\_locationproxy/refute\_transfer\_permutation.csv`,
`catalogue/s1\_screen/out\_zerohour/refute\_locationproxy/refute\_within\_block\_auc.csv`,
`catalogue/s1\_screen/out\_zerohour/refute\_locationproxy/refute\_xy\_plus\_features.csv`,
`catalogue/s1\_screen/out\_zerohour/transfer\_results.csv`,
`catalogue/s1\_screen/out\_zerohour/zero\_hour\_inventory.md`,
`catalogue/s1\_screen/out\_zerohour/zero\_hour\_results.csv`,
`catalogue/s1\_screen/out\_zerohour/zerohour\_lib.py`, `figures/fig19\_zero\_hour\_vs\_rs.png`,
`figures/fig19b\_zero\_hour\_transfer.png`, `figures/fig19c\_operational\_value.png`,
`figures/fig19c\_refute\_folds.png`.

12. How the survey targets were chosen

Sections 1 to 11 say what the products show and where they disagree. This section says how the 15 to 26 September field program was turned into a list of buildings, and what the resulting sample will and will not be allowed to claim in December. It matters because the answer is not a list of interesting places. It is a probability sample, drawn with a recorded inclusion probability on every unit, cut to an itinerary that was fixed before the sample existed, and it will fail to produce any defensible accuracy number at all if the field team treats it as a suggestion.

Everything in this section is provisional in the same sense as the rest of the report. All NISAR (NASA ISRO Synthetic Aperture Radar) quantities used in the stratification are provisional version 1 products and are labeled as such. The acceptance test rests on synthetic truth fields, not on field observations, and the interval rule derived from it is a calibration on those synthetic fields and is labeled as a stated modeling choice where it appears.

Acronyms used in this section at first occurrence: PPS (probability proportional to size); PSU (primary sampling unit); SRS (simple random sampling); Kish deff (the Kish weighting design effect, a measure of how much unequal weights inflate the variance); EMS (Copernicus Emergency Management Service); CEMS as used in earlier sections.

12.1 The frame: everything that could have been chosen

The sampling unit is a 250 m cell inside one of the seven city boxes, and inside a drawn cell, a building footprint. The frame retains 5,278 cells holding 331,359 buildings after five exclusion rules, each recorded with its reason so that no cell leaves the frame by judgment. A cell leaves the frame when it holds no building footprint, when its centroid does not sit on built up land, when it is cut by a burst edge, when it carries no valid post event radar layer, or when it fails the conditional Cali post event banding rule described in section 5. The frame was built before the itinerary was adjudicated and it is unchanged by anything in this section.

The population the design can make probability statements about is smaller than the frame, and the gap is the single most consequential number in this section. It is 2,746 cells and 178,273 buildings. Cali and Quibdo fall out of it because their scheduled days are fully committed to named non probability building lists and to strong motion station buffers, so the frame's mechanism selects no unit there and no unit there carries an inclusion probability. Cali takes 2,310 cells and 129,028 buildings with it, and with them 248 of the 389 optically comparable cells in the entire frame; Quibdo takes 222 cells and 24,058 buildings, and with them the Copernicus certainty stratum and the highest optically comparable fraction anywhere at 0.297.

The valid area rule of section 4 carries into the sampling unchanged: optical comparisons are made only inside the intersection of the producer masks, and outside a producer mask is no evidence and never undamaged.

Take-away: the frame stands at 5,278 cells, the design population is 2,746, and no number about Cali or Quibdo may be quoted from this sample however the field days there go.

12.2 The strata, and what each one is for

The frame defines 21 design strata and each answers a named question. Position blocks are cut at the within city median distance to the density core, and they are a blocking variable rather than a contrast: they exist so that the comparison below is made between buildings in comparable neighborhood contexts. The NISAR neighborhood excess terciles are cut within the position block and never across it, so that position can never become the contrast being measured. A sensor disagreement stratum isolates the cells where the radar and the optical products contradict each other, which is where a visit changes what we believe rather

than confirming it. A NISAR flagged stratum serves the secondary estimand, NISAR precision. A Copernicus labeled stratum anchors the fourth estimand, the model assisted damaged total.

Twelve of those strata survive at the budget the itinerary allows, in five cities, against 21 in seven cities before the schedule cut. The arithmetic that forces this is simple. A city can afford the integer part of its record capacity divided by the cluster size m , each surviving stratum needs at least one drawn cell, and no city reaches six cells. Armenia and Pereira and Dosquebradas keep the three band excess contrast but give up position blocking, so their contrast is no longer matched on neighborhood context. Manizales keeps one cell in one stratum and has no contrast at all. Armenia's last step was forced not by budget but by the 20 to 1 within city weight ratio cap, which six strata breached at 20.34.

The order in which strata were surrendered is recorded and it is the reverse of the estimand ordering, so the primary contrast is always the last thing given up: the Copernicus anchor first, then the NISAR flagged stratum, then sensor disagreement, then position blocking, then the middle excess band, and the contrast last.

Take-away: the six position by excess strata survive in no city, so the primary contrast is estimable in Armenia, Pereira and Dosquebradas, Quimbaya and Sevilla as a pooled high against low comparison and not as the matched one the frame was built for, and it is not estimable in Manizales at all.

12.3 The itinerary is a hard constraint, and the frame shares records rather than adding them

The adjudicated Schedule sheet commits 550 records over ten survey days. The same sheet carries a header total of 610 distinct records, which is the sum of the objective success criteria read as distinct buildings and was written before the day grid was fixed. It exceeds the day grid by 60 records, 10.9 percent. We size the frame on 550, because a frame sized on 610 would claim records on days that do not exist, and because the 20 September mid mission cut rule already provides for the gap by releasing about 100 records of aspiration.

A building surveyed for an objective and for the frame is one record and not two, because objectives RS-12, RS-15, RS-16 and RS-17 are mandatory on every record taken. The frame's claim is therefore a subset of the 550 and never an addition to it. Of the 550, 300 are pinned to a named non probability building list or a station buffer and can carry no inclusion probability; the frame claims 169 as shared with the objectives and leaves 381 objective only.

The consequence is that the capacity entering the allocation is not a city's day target but the part of it the objectives have not already pinned. Pereira and Dosquebradas has 120 free records across four days, Armenia 50, Sevilla 40, Quimbaya 30, Manizales 10, Cali 0 and Quibdo 0.

Take-away: the frame walks the itinerary that already exists and adds no visit to any day, which is why its 169 records sit inside 550 and not beside them.

12.4 The allocation: Neyman with unequal costs, inside the itinerary

Records are allocated by Neyman allocation with unequal costs, with the number of units in stratum h proportional to the product of the stratum size and the stratum standard deviation divided by the square root of the per unit cost. The rule is Neyman (1934, Journal of the Royal Statistical Society 97, 558 to 625) as given for variable cost by Cochran (1977, Sampling Techniques, section 5.5). The stratum standard deviation is built from the mapped damage quantity used as an allocation prior after Padilla et al. (2017, Remote Sensing of Environment 203, 240 to 255). The competing estimands are reported alongside the primary rather than optimized jointly, after Wagner and Stehman (2015, Remote Sensing of Environment

168, 126 to 133) and Stehman (2012, Remote Sensing Letters 3, 111 to 120), whose point is that no single allocation serves every estimand at once.

The cost term is in team minutes per record and every term comes from the itinerary or from a frame file. It is 6 minutes for the grade, plus the cell set up of 20 minutes and the hop to the cell divided by the m records the frame takes there, plus the daily mobilization of 45 minutes per city day amortized over the city's scheduled records, plus the 20 September inter city leg amortized over the 310 records it unlocks at 6.19 minutes per record, plus the 15 September arrival amortized over all 550 at 3.49 minutes per record. The hop is twice the stratum's own median distance to the density core at 20 km per hour, which is what makes the cost vary within a city and not only between cities. The resulting per record cost runs from 15.88 minutes in Pereira and Dosquebradas to 22.25 in Armenia.

The cost term is second order and we say so with a number. Holding the total fixed and ignoring the itinerary, it moves Pereira and Dosquebradas up 10.9 percent and Armenia down 6.3 percent. The schedule moves the same allocation by an order of magnitude more: Cali falls from 374 records on the 900 record backbone plus reserve basis to zero, Manizales from 136 to 10, and Quimbaya and Sevilla rise from 17 each to 30 and 36 because they hold a scheduled day and a scheduled half day.

A factor of two stress test on all six cost terms qualifies what may be said about the cost model, and the qualification is not cosmetic. The full five city cost ordering survives 9 of 12 one at a time moves and only 20 of 64 joint moves, so the ordering is not a result and is not quoted as one. What survives is the distinction the allocation actually uses: Pereira and Dosquebradas is the cheapest city per record in 12 of 12 one at a time cases and 48 of 64 joint cases, and its gap to the next cheapest is 27 percent where the three cities behind it sit within 2.4 percent of one another. The sign of the cost effect is stable in 64 of 64 joint cases for Pereira and Dosquebradas and for Armenia, and unstable for Quimbaya in 8 of 64 and for Manizales and Sevilla in 16 of 64 each, so those three shifts are not quoted to one decimal without the caveat.

Take-away: the cost term is worth having and it is not what reshaped this design, and the only part of the cost model robust to a factor of two is that Pereira and Dosquebradas is much the cheapest city per record.

12.5 The draw, the weights, and the two domains

Cells were drawn by PPS systematic sampling within stratum, with size the cell's frame buildings, without replacement, under seed 20260905. Inside each drawn cell, each of four within cell classes is drawn at the conditional probability equal to the smaller of one and the class quota divided by the cell's building count, with randomized rounding, so the expected rate is exactly that probability and the probability actually used is recorded on the row. The design weight is the reciprocal of the product of the two stages.

The draw yields 21 cells and 169 buildings, with design weights from 17.85 to 9,058.41 and an effective sample size of 57.9. Because the second stage rounds randomly, the realized sample size is itself random: over 3,000 independent draws it averages 165.5 with a standard deviation of 5.69 and a range of 145 to 189, and the 169 in force is one draw from that distribution.

The within city weight ratio cap of 20 to 1 is met everywhere, in expectation and at the worst of 3,000 draws, with the worst city Pereira and Dosquebradas at 12.09. The frame wide Kish deff cap of 1.50 is not met and cannot be met under this itinerary. Its expected value over 3,000 draws is 3.191 with a standard deviation of 0.162 and a probability of meeting the cap of 0.000; the value realized in the recorded draw, 2.921, sits at the 5.8 percentile of that distribution, which is why the expected figure and not the realized one is quoted first.

The breach is entirely between cities and no coarsening of the within city allocation reaches it. Every within city Kish deff lies between 1.23 and 1.40, and the floor with every city allocated exactly

proportionally inside itself is 3.209, above the value in force. Sevilla is 1.9 percent of the design population and takes 23.7 percent of the sample because it holds a whole scheduled day, while Manizales is 28.7 percent and takes 5.3 percent because 50 of its 60 records are pinned; Sevilla's sampling rate is 69 times Manizales'. Weight trimming would meet the cap by introducing bias and is not used.

The Kish deff of a domain is computed on that domain's units alone, so the operational answer is to report two domains and never pool them. Armenia with Pereira and Dosquebradas carries 90 records at an expected Kish deff of 1.3965, with a probability of 0.0063 of breaching the cap on any given draw; Quimbaya with Sevilla carries 70 records at 1.2579 with a probability of 0.0000. Manizales meets the cap on its own but is a single PSU city and carries no within stratum variance.

Take-away: the two domains are the largest populations this sample can speak for inside the weight cap, and a single frame wide figure from these 169 records is not defensible.

12.6 The acceptance test against simple random sampling

The design was tested against 600 replicates at 169 visits on seven synthetic truth fields and four estimands, with a degenerate replicate counted as a miss. Two of the truth fields are clustered on an arbitrary offset tiling of the UTM plane that has nothing to do with the strata, the cells or the cities, and one is deliberately anti correlated with the product score, so the test cannot flatter a design tuned on that score.

Three comparisons are reported and none may be quoted without the other two.

Comparison arm	What it is	Design wins on the primary contrast	Worst standard error ratio
Equal visits SRS over the design population	the acceptance bar in force, not executable under this itinerary	no, loses on four of seven truth fields	1.3589
Schedule constrained SRS at the same per city record counts	the alternative the field lead could actually choose	yes, on all seven	0.8398
Equal cost SRS, 55 records for the same team minutes	the bar implied by allocating on cost	yes, on all seven	0.7600

Source: `catalogue/s1\_screen/out\_frame\_v3/acceptance\_test.csv`, rows with estimand `primary\_contrast`, columns `se\_ratio\_design\_over\_srs`, `se\_ratio\_design\_over\_srs\_city` and `se\_ratio\_design\_over\_srs\_equal\_cost`, maxima over the seven truth fields.

The equal cost arm needs its accounting stated. The design spends 3,329 team minutes on its 169 records at its own recorded per stratum cost. An unstratified sample of the same design population lands about one building per cell, so it pays the cell set up and the hop once per record rather than amortizing them over m, at 60.45 minutes per record, and the same budget therefore buys 55 records.

Visit equivalence, the number of SRS visits that would buy the same precision, is reported per truth field and never averaged. On the equal visits basis it runs 273, 228, 173, 152, 140, 137 and 92 against the design's own 169 visits; on the equal cost basis it runs 295, 266, 190, 169, 142, 100 and 95 against the 55 records the same budget buys. The minimum is the number to quote. The mean of the first column is 171 and would read as though the design were worth more than its own visits on every field, which it is not: on the truth field anti correlated with the score it is worth 92 of its own 169. The field the allocation was tuned on ranks second of seven rather than first, and the best field is one of the two clustered on the arbitrary offset tiling, so the advantage comes from the matched within cell contrast and not from tuning.

Take-away: this design is efficient for the itinerary it is stuck with and is not efficient against an unconstrained sample of the same size, and both statements are recorded.

12.7 The estimators, the interval, and what this sample cannot do

Every estimate from the survey is a weighted one, a ratio of weighted totals with a linearized residual for the variance, following the estimators of section 8 and Stehman (2014, International Journal of Remote Sensing 35, 4923 to 4939) for the case where the strata differ from the classes being estimated. The estimator is unbiased at this sample size: over 3,000 independent draws the absolute bias is at most 0.0019 and at most 0.026 design standard errors on all seven truth fields, with the largest a t statistic of 1.45 against Monte Carlo error.

The variance is a between PSU variance with the 250 m cell as the PSU, computed over collapsed variance groups rather than over the design strata, after Wolter (2007, Introduction to Variance Estimation, section 2.5). Collapsing is required because the schedule leaves several strata with a single drawn cell: 4 of the 12 drawn strata carry a collapsed group and the design has 12 degrees of freedom. The group is written into the `variance\_group` column of the allocation, cell and building files and the field analysis must use it; using the stratum instead returns an interval that is too narrow.

The interval needs more than that, and this is the part of the design that failed its own audit and was corrected. At 169 records the design based variance estimator understates the true sampling standard deviation of the primary contrast by a factor of 1.111 to 1.433 across the six non degenerate truth fields, and the nominal 95 percent interval built from Student's t at 12 degrees of freedom covers at 0.842, 0.802 and 0.808 on the three truth fields that carry signal. A single multiplier does not repair it, because the estimated standard error is unstable rather than merely biased low: on the anti correlated field its fifth percentile is 0.00478 against a median of 0.05712.

The rule adopted is a stated modeling choice, calibrated on the seven synthetic truth fields. Multiply the design standard error by 3.00 in place of Student's t at 2.179, and report the result as an interval of its measured coverage, 0.858 to 0.997, never as an exact 95 percent interval. The multiplier is the knee of the measured coverage curve, the smallest multiplier past which widening the interval by half again buys less than one coverage point.

No multiplier reaches 0.95, because a replicate in which the variance estimator returns zero is a miss that no width rescues. Each estimand therefore has a coverage ceiling, and for two of them the ceiling is fatal: NISAR precision and NISAR recall have a worst field ceiling of 0.030, so this sample supports no interval at all on either. That is a direct consequence of Cali and Quibdo leaving the design population, since they held the Copernicus certainty stratum and most of the frame's optically comparable cells. All NISAR quantities here remain provisional.

Setting this out as a list of what may and may not be claimed:

Claim	Status
The high against low neighborhood excess contrast, in Armenia with Pereira and Dosquebradas, and in Quimbaya with Sevilla, reported separately	supported, at an interval of measured coverage 0.858 to 0.997
The same contrast, matched on neighborhood context	not supported in any city; position blocking did not survive the budget
The same contrast in Manizales	not supported; one cell, one stratum, no arms
A single frame wide accuracy, precision, recall or damaged total	not supported; the frame wide Kish deff is 3.191 against a cap of 1.50
Any number about Cali or Quibdo	not supported; neither city is in the design population
An exact 95 percent interval on anything	not supported; intervals are quoted at the calibrated multiplier and labeled by measured coverage
Any interval on NISAR precision or NISAR recall	not supported; coverage ceiling 0.030

Source: `catalogue/s1\_screen/out\_frame\_v3/README.md`, the recommendation section, and `frame\_v3\_log.json`, blocks `weights`, `interval\_rule` and `design\_population`.

Take-away: this sample answers one question well in two groups of cities, and the honest list of what it cannot answer is longer than the list of what it can.

12.8 What the field team has to do, and what would recover the rest

Three rules govern the field work and each of them, if broken, removes a class of result rather than a single record. Never substitute a building: every one of the 169 was drawn with a known inclusion probability, and replacing an unreachable building with a convenient neighbor falsifies that probability for the whole stratum, not just for that row. If a building cannot be reached or graded, the reason is recorded and the team moves on, because the non response is data and there is no reserve list in this design. The inclusion probability and design weight are printed on the row before departure and are never edited afterwards, and the row is completed in the cell rather than from memory. And outside a producer mask is no evidence and never undamaged, so the absence of a product assessment is never recorded as the absence of damage.

Three schedule changes would recover most of what the itinerary cost, and none of them alters the design. Quibdo on 26 September is worth 40 records: the 74 Copernicus affected buildings all sit inside the AOI04 center, which is exactly Quibdo's Copernicus certainty stratum, so working that named list in an order drawn by the frame's second stage returns Quibdo and the mission's best optical evidence to the design population. Manizales on 23 September is worth 40 records: one station buffer alone holds 725 footprints and the objective fixes the buffer rather than the building, so drawing the buffer buildings inside frame cells raises Manizales from 10 free records to 50 and stops it being a single PSU city. Cali on 25 September is worth 20 records in the one part of the day not tied to a named list, and the larger fix is a second Cali day. The first two together would raise the sample from 169 to about 249 and would add records to the two cities with the lowest sampling rates, which moves the between city Kish deff toward the cap rather than away from it.

Take-away: the design as it stands is defensible in two domains and no wider, the three field rules are what keep it defensible, and the highest value change available to the mission is to let the frame order Quibdo's Copernicus list on 26 September.

Sources: `catalogue/s1\_screen/out\_frame/allocation.csv`,
`catalogue/s1\_screen/out\_frame/frame\_log\_v2.json`, `catalogue/s1\_screen/out\_frame\_v3/README.md`,
`catalogue/s1\_screen/out\_frame\_v3/acceptance\_test.csv`,
`catalogue/s1\_screen/out\_frame\_v3/allocation.csv`, `catalogue/s1\_screen/out\_frame\_v3/cost\_model.md`,
`catalogue/s1\_screen/out\_frame\_v3/cost\_stress.csv`,
`catalogue/s1\_screen/out\_frame\_v3/frame\_v3\_log.json`,
`catalogue/s1\_screen/out\_frame\_v3/sample\_buildings.csv`,
`catalogue/s1\_screen/out\_frame\_v3/schedule\_reconciliation.csv`,
`catalogue/s1\_screen/out\_frame\_v3/visit\_equivalence.csv`,
`catalogue/s1\_screen/out\_frame\_v3/weight\_diagnostics.csv`,
`report/EEFIT\_Colombia\_2026\_Field\_Frame.xlsx`,
`report/EEFIT\_Colombia\_2026\_RemoteSensing\_Objectives.xlsx`,
`report/targets/FAME\_V3\_ADJUDICATION.md`,
`report/targets/FAME\_V3\_SCHEDULE\_ADJUDICATION.md`.