The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIX-B3-2026
XXV ISPRS Congress 2026 “From Imagery to Understanding”, Commission III, 4–11 July 2026, Toronto, Canada

Fusion of AlphaEarth Embeddings and Sentinel-1 Time-Series for Conflict-Related Urban Damage Mapping

Jakub Ślesiński, Kinga Karwowska, Magdalena Lewińska

Department of Imagery Intelligence, Faculty of Civil Engineering and Geodesy, Military University of Technology, Warsaw, Poland
(jakub.slesinski@wat.edu.pl, kinga.karwowska@wat.edu.pl, magdalena.lewinska@wat.edu.pl)

Keywords: change detection; SAR; AlphaEarth; war damage mapping; conflict monitoring

Abstract

Conflict-related urban damage mapping requires methods that remain reliable when field access is limited, cloud cover is frequent, and urban conditions vary strongly. Sentinel-1 SAR time series provide systematic all-weather coverage, but SAR-only indices are still sensitive to speckle, acquisition geometry, and heterogeneous urban background. AlphaEarth Foundations embeddings offer a complementary feature space that may stabilize change detection by encoding broader semantic context. This study evaluates scalar fusion of AlphaEarth embedding change and Sentinel-1 change indices for building-level conflict damage mapping.

We first compare multiple AlphaEarth-based change metrics and retain Earth Mover's Distance (EMD) as the most suitable embedding-space indicator. We then combine EMD with Sentinel-1 indices through simple scalar fusion rules and evaluate the results on 75,825 reference building footprints from seven Ukrainian cities with damage rates ranging from 0.2% to 33.7%. The strongest transferable method, EMD+PWTT_w40, achieves mean building-level AUC 0.676, outperforming both PWTT alone (0.657) and EMD alone (0.593). A lightweight learned fusion gives only a small within-city gain (mean AUC 0.686) and degrades under leave-one-city-out transfer (0.614). Sensitivity analysis further shows that the transferred fusion weight remains stable, with w_EMD = 0.4 staying within 0.0028 AUC of the city-specific optimum on average.

These results show that AlphaEarth and SAR capture complementary aspects of urban damage and that a simple scalar fusion rule can combine them effectively without sacrificing interpretability or transferability.

1. Introduction

1.1 Motivation and problem statement

Conflict-related urban damage mapping is increasingly required for humanitarian response, rapid situational awareness, reconstruction planning, and longitudinal assessment of war impacts on the built environment. In recent conflicts, including those in Ukraine and the Gaza Strip, analysts have had to identify damaged neighbourhoods and buildings under severe time pressure, often in conditions where field access is limited or impossible. Remote sensing is therefore a primary source of evidence, but the characteristics of the available sensors create a persistent trade-off between spatial detail, temporal revisit, and robustness to environmental conditions (Aimaiti et al., 2022; Scher and Van Den Hoek, 2025).

Very high resolution optical imagery remains essential for detailed interpretation, but it is not always available when needed and is constrained by cloud cover, acquisition gaps, viewing geometry, and licensing. Recent conflict-oriented optical approaches based on repeated high-resolution imagery and machine learning have shown strong potential, especially in Syria, but they also highlight issues of transferability, label scarcity, and operational cost when scaling to new cities or long time series (Mueller et al., 2021). For operational monitoring, Sentinel-1 C-band SAR time series provide a more consistently available alternative. Their all-weather, day-and-night acquisition capability makes them particularly attractive for repeated observations over active conflict zones. However, urban damage detection from SAR remains difficult at the building scale because backscatter changes are affected not only by destruction, but also by speckle, viewing geometry, local scattering conditions, and heterogeneous urban morphology (Aimaiti et al., 2022; Boloorani et al., 2021; Scher and Van Den Hoek, 2025).

Recent Earth observation foundation models also provide a new feature space for change analysis. AlphaEarth Foundations (AEF) embeddings are designed as compact, semantically informed representations learned from multi-source geospatial observations (Brown et al., 2025). Rather than encoding only instantaneous radiometric behaviour, these embeddings summarize broader spatial, temporal, and environmental context. This makes it possible to measure change in an embedding space that is less affected by local radiometric noise and seasonal background variation than conventional single-sensor indicators.

1.2 Limits of single-source change detection

Single-source SAR indicators can highlight abrupt structural change, especially in heavily damaged urban blocks, but in mixed land-cover settings their responses are often unstable. In dense cities, the same index can be influenced by layover, local incidence angle, changes in moisture, or background heterogeneity unrelated to conflict damage. As a result, SAR-only approaches can produce false positives in stable areas and inconsistent ranking of damaged buildings across cities with different morphology and damage prevalence.

AEF-only change metrics have the opposite profile. Because they operate in a semantically richer embedding space, they can reduce some of the background variability that affects conventional pixel-wise indices. They are therefore attractive for distinguishing genuinely altered urban fabric from broader land-cover fluctuations. Yet they are not optimized specifically for abrupt building collapse or facade damage, and on their own they may respond less sharply than dedicated SAR change indicators to sudden structural disruption.

Taken together, these strengths suggest treating AEF and SAR as complementary inputs rather than competing alternatives. SAR can provide high sensitivity to abrupt change, while AEF can contribute contextual stability and reduce background artefacts. The key question is whether this complementarity can be used in a fusion scheme that remains interpretable, transferable across cities, and practical in near-operational settings.

1.3 Research gap and paper objectives

Existing demonstrations of AEF and Sentinel-1 fusion for conflict-related damage mapping have been largely qualitative and case-study oriented. More broadly, the recent literature is split between high-resolution optical machine-learning approaches for conflict damage monitoring (Mueller et al., 2021) and SAR-based or InSAR-based mapping workflows that emphasize broad operational coverage (Aimaiti et al., 2022; Boloorani et al., 2021). It is still unclear whether the complementarity between embedding-space change and SAR indices holds across multiple cities, how fusion compares with strong single-source baselines, and whether a simple scalar rule is preferable to a lightweight learned model when transferability matters.

This paper addresses those questions through a quantitative evaluation built around building-level reference data for seven Ukrainian cities spanning heavy-, moderate-, and light-damage settings. We compare AEF-only, SAR-only, and fused methods using standard accuracy metrics, evaluate learned fusion with both within-city and leave-one-city-out protocols, and analyse the sensitivity of scalar fusion to the transferred weight parameter and selected post-processing options.

The paper therefore examines whether AEF and SAR remain complementary across diverse urban conflict settings and whether a simple scalar fusion rule can perform robustly without city-specific retraining.

1.4 Contributions

The main contributions of this paper are as follows:

- We transfer a set of established change metrics to the AlphaEarth embedding space and compare them systematically. Earth Mover's Distance (EMD) emerged as the most suitable indicator - somewhat unexpectedly, given that simpler Euclidean metrics performed well in the original spectral domains these measures come from.
- A family of scalar fusion operators combining AEF EMD with Sentinel-1 PWTT is formulated and evaluated, with weighted summation ($w\_EMD = 0.4$) identified as the strongest transferable variant.
- We provide a building-level benchmark across seven Ukrainian cities, 75,825 footprints, damage rates from 0.2% to 33.7%, which to our knowledge is the most geographically diverse quantitative evaluation of AEF-based change detection to date.
- Scalar fusion is compared against lightweight learned fusion under both within-city and leave-one-city-out protocols. The results argue against adding model complexity when transferability is the priority.

2. Study Areas and Datasets

2.1 Study design

The study is built around a multi-city validation experiment in Ukraine. Seven cities were selected to cover a broad spectrum of urban conflict conditions, damage prevalence, and geographic context: Irpin, Sievierodonetsk, Rubizhne, Avdiivka, Hostomel, Lysychansk, and Kramatorsk. All cities are evaluated with the same methodological pipeline, using AEF embeddings from 2021 and 2022 and city-specific Sentinel-1 pre/post-event stacks aligned to the corresponding reference date.

The selected cities span three broad damage regimes. Sievierodonetsk and Rubizhne represent heavy-damage cases, with damage prevalence between 24.3% and 33.7%. Irpin, Avdiivka, Hostomel, and Lysychansk represent moderate-damage cases, with damage rates between 7.4% and 13.7%. Kramatorsk represents a light-damage case, with only 0.2% damaged buildings in the reference inventory. This grouping matters because it tests the methods across both different urban forms and very different levels of class imbalance.

Across the seven cities, the evaluation includes 75,825 building footprints with binary damage labels. Reference dates range from 31 March 2022 to 21 September 2022, and the damage prevalence ranges from 0.2% to 33.7%. This design supports both per-city analysis and cross-city comparisons of robustness and transferability. Table 1 summarizes the study areas, reference dates and building counts used in the quantitative benchmark.

City	Reference date	Buildings	Damage rate (%)	Damage regime
Irpin	2022-03-31	7,242	11.3	Moderate
Sievierodonetsk	2022-07-27	5,970	24.3	Heavy
Rubizhne	2022-07-09	8,899	33.7	Heavy
Avdiivka	2022-09-20	7,262	8.7	Moderate
Hostomel	2022-03-31	4,326	13.7	Moderate
Lysychansk	2022-09-21	20,246	7.4	Moderate
Kramatorsk	2022-07-24	21,880	0.2	Light

Table 1. Study areas and reference data summary

2.2 Earth observation inputs

The optical-semantic component of the study is based on AlphaEarth Foundations embeddings (Brown et al., 2025). AEF provides annual 64-dimensional embedding fields at 10 m spatial resolution derived from multi-source Earth observation inputs, including Sentinel-1, Sentinel-2, Landsat, digital elevation information, climate descriptors, and land-cover contextIn this study, the embeddings are treated as semantically informed feature vectors rather than conventional spectral signatures. Change metrics are therefore computed in the embedding space between the pre-event and post-event annual fields, which in the Ukrainian validation setup correspond to 2021 and 2022.

The SAR component is derived from Sentinel-1 GRD time series. For each city, we construct pre-event and post-event stacks from VV/VH observations and retain a single relative orbit where possible in order to reduce geometric inconsistencies between

https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-1331-2026 |

acquisitions. This choice follows the operational logic of conflict monitoring with SAR: maintaining frequent coverage while minimizing variability caused by orbit-dependent imaging differences. Depending on the city and temporal window, the SAR stacks typically contain several to a dozen acquisitions per phase. In this respect, the study is closer to lightweight time-series damage mapping workflows based on open Sentinel-1 data than to approaches that rely on interferometric coherence or commercial VHR optical imagery alone (Canty et al., 2020; Boloorani et al., 2021; Ballinger, 2025).

The two sources complement each other both in the type of information they encode and in the sensing physics behind them. AEF embeddings integrate broader environmental and temporal context into a compact annual representation, while Sentinel-1 retains direct sensitivity to abrupt structural change at short revisit intervals. Their common 10 m working resolution also enables straightforward pixel-level combination before building-level aggregation (Gorelick et al., 2017).

2.3 Reference data

The primary evaluation layer is taken from the open PWTT Dataset (Ballinger, 2025), which provides UNOSAT-derived building footprints together with binary damage labels for conflict-affected cities. Related UNOSAT damage inventories have also been used as reference material in recent satellite-based assessments of war damage in Ukraine (Aimaiti et al., 2022; Scher and Van Den Hoek, 2025). In this study, these footprint labels are used strictly as reference data for evaluation. They are not used to train the proposed scalar fusion, and they provide a consistent building-level target across all seven cities.

For each city, the footprint dataset contains a binary class indicating whether the building is labelled as damaged in the reference inventory. This binary label is the target for the main building-level evaluation reported throughout the paper. Because the goal of the study is ranking and discrimination of damaged versus undamaged buildings, the building-level AUC is the primary metric, complemented by threshold-dependent measures such as F1, precision, recall, overall accuracy, and Cohen's kappa.

The PWTT Dataset is complemented by city-specific AOI geometries where available. For Avdiivka and Sievierodonetsk, polygon AOIs are available directly, whereas for the remaining cities the analysis extent is derived from the bounding box of the building footprints. This does not affect the evaluation labels, but it does define the processing extent used in Earth Engine.

3. Methodology

3.1 AlphaEarth change metrics

Let $\mathbf{e}_i^{pre}, \mathbf{e}_i^{post} \in \mathbb{R}^K$ denote the AlphaEarth embedding vectors for pixel i in the pre-event and post-event annual layers, respectively, with $K = 64$ in the present implementation. The basic embedding change vector is

$$\Delta\mathbf{e}_i = \mathbf{e}_i^{post} - \mathbf{e}_i^{pre} \tag{1}$$

Several standard change metrics commonly used in multispectral and hyperspectral change detection were transferred to the AEF embedding space (Lv et al., 2025). Rather than assuming that one metric would suit foundation-model embeddings best, we tested which measure of embedding change most clearly separates damaged urban fabric from stable urban and peri-urban areas. The tested family included Spectral Angle Mapper (SAM), Spectral Information Divergence (SID), Squared Euclidean Distance (SED), the magnitude of the Change Vector Analysis (CVA) vector, and Earth Mover's Distance (EMD). For SAM, we measure the angular difference between the two embedding vectors:

$$\mathrm{SAM}_i = \arccos\left(\frac{\mathbf{e}_i^{pre} \cdot \mathbf{e}_i^{post}}{\|\mathbf{e}_i^{pre}\|_2 \, \|\mathbf{e}_i^{post}\|_2 + \varepsilon}\right) \tag{2}$$

where ε is a small stabilizing constant. This metric emphasizes directional change in the embedding space rather than absolute magnitude. The Euclidean family is represented by

$$\mathrm{SED}_i = \|\Delta\mathbf{e}_i\|_2^2, \mathrm{CVA}_i = \|\Delta\mathbf{e}_i\|_2 \tag{3}$$

These metrics react to the overall size of the displacement in embedding space and therefore preserve strong contrast when large changes occur, but they can also remain sensitive to broad background shifts. SID was treated as an information-theoretic alternative, following its standard use on normalized feature vectors interpreted as pseudo-distributions. This provides a complementary perspective on whether the relative composition of the embedding channels changes between the two annual states.

The main AEF metric retained for the subsequent fusion experiments was EMD. In the Earth Engine implementation, annual AEF mosaics are loaded from the GOOGLE/SATELLITE_EMBEDDING/V1/ANNUAL collection, clipped to the AOI, and compared using the built-in spectral distance operator for EMD. The raw output is divided by the number of embedding bands and then robustly normalized to the $[0,1]$ range using 2nd and 98th percentiles within the AOI:

$$\mathrm{EMD}_i^{norm} = \mathrm{clip}_{[0,1]}\left(\frac{\mathrm{EMD}_i - P_2}{\max(P_{98} - P_2, \varepsilon)}\right) \tag{4}$$

This percentile normalization is important because the dynamic range of the raw distance map varies across cities and AOI extents. In Figure 1, EMD gives the clearest separation between damaged districts and stable surroundings, whereas the other metrics respond more strongly to land-cover boundaries, agricultural texture, and seasonal background variation. For that reason, EMD was adopted as the primary AEF-based input in the remainder of the study, although the initial multi-metric comparison remains important for motivating this selection.

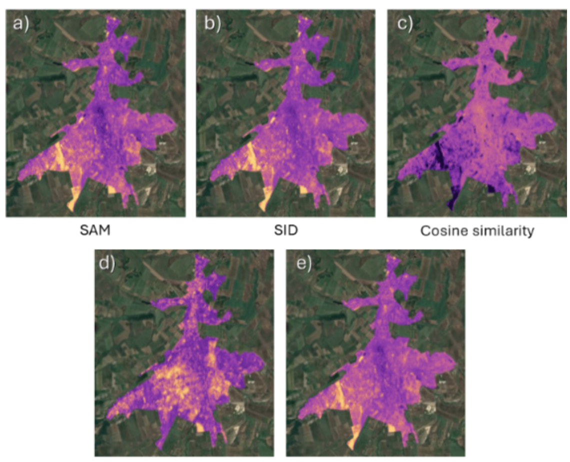

Figure 1. Qualitative comparison of AlphaEarth-based change metrics over Bakhmut

3.2 SAR change indices

The SAR branch is based on Sentinel-1 GRD VV/VH time series and provides the high-frequency structural change cues that AEF alone does not explicitly encode. Two SAR-only indices were used as baselines: a REACTIV-style coefficient of variation (CV) measure and the Pixel-Wise Welch's T-Test (PWTT) introduced by Ballinger (2025). Both are intended to highlight abrupt conflict-related changes, but they capture different temporal behaviours and therefore serve as complementary reference points for the proposed fusion. The broader methodological context is the growing use of statistical time-series analysis on open Sentinel-1 archives within Earth Engine and related cloud environments (Canty et al., 2020).

The REACTIV-style metric follows the coefficient-of-variation logic for SAR amplitude time series over urban areas (Colin Koeniguer and Nicolas, 2020). Sentinel-1 images are first restricted to the dominant relative orbit over the AOI to reduce geometry-driven inconsistencies. For each acquisition, VV and VH backscatter are converted to amplitude form and, when required, masked to built-up areas using a Dynamic World built-up probability mask computed over the pre-war period (Brown et al., 2022). The CV score is then computed separately for VV and VH and averaged:

$$\mathrm{CV}_i = \frac{1}{2}\left(\frac{\sigma(\mathrm{VV}_i)}{\mu(\mathrm{VV}_i)+\varepsilon} + \frac{\sigma(\mathrm{VH}_i)}{\mu(\mathrm{VH}_i)+\varepsilon}\right) \tag{5}$$

In the notebook implementation this raw quantity is further normalized using a stack-size-aware scaling, so that the final REACTIV-style score remains comparable across cities with different numbers of Sentinel-1 acquisitions. The resulting map is again clipped to [0,1]. This metric is sensitive to temporal instability in the SAR signal and can therefore respond strongly to damaged zones, but it may also react to heterogeneous urban texture, wet surfaces, or other non-conflict temporal changes.

PWTT is designed to measure the statistical significance of the shift between pre-event and post-event SAR stacks. For each AOI, Sentinel-1 GRD_FLOAT images are filtered by instrument mode, polarization availability, orbit pass, and then again reduced to the most common relative orbit. A Lee filter is applied to reduce speckle, after which the VV and VH channels are log-transformed and split into pre-event and post-event collections. For each polarization, the Welch-type standardized difference between post-event and pre-event means is computed as

$$t_{i,b} = \frac{\left|\mu_{i,b}^{post} - \mu_{i,b}^{pre}\right|}{\sqrt{\frac{(\sigma_{i,b}^{pre})^2}{n_{pre}} + \frac{(\sigma_{i,b}^{post})^2}{n_{post}} + \varepsilon}} \tag{6}$$

where $b \in \{\mathrm{VV}, \mathrm{VH}\}$. The final pixelwise PWTT score is the maximum over the two polarizations:

$$\mathrm{PWTT}_i = \max\left(t_{i,\mathrm{VV}}, t_{i,\mathrm{VH}}\right) \tag{7}$$

The PWTT map is then smoothed with a 30 m circular median filter and normalized to [0,1] with the same 2nd/98th percentile scheme used for EMD. PWTT is expected to be more directly sensitive than REACTIV-style CV to abrupt structural shifts between the two time windows. However, it still remains vulnerable to residual orbit mismatch, local scattering geometry, and context-specific background effects. In the remainder of the paper, REACTIV-style CV and PWTT are treated both as single-source baselines and as ingredients for fusion, with PWTT emerging as the stronger SAR baseline overall.

3.3 Scalar fusion

The key methodological question is whether AEF and SAR can be combined with a simple scalar rule, rather than a city-specific trained model, without losing interpretability or transferability. All candidate fusion maps are constructed from normalized single-source inputs in the common [0,1] range, which makes the operators directly comparable across cities and prevents one source from dominating only because of scale. We tested four scalar fusion families. The first is a weighted linear combination:

$$F_i^{lin}(w) = w\,\mathrm{EMD}_i + (1-w)\,S_i \tag{8}$$

where S_i is either REACTIV-style CV or PWTT. The principal configuration used in the article is EMD+PWTT_w40, corresponding to $w = 0.4$ for EMD and 0.6 for PWTT. This choice gives slightly more weight to the SAR branch because it is more sensitive to abrupt damage, while still preserving a substantial contribution from the more stable AEF branch.

The second family is a multiplicative soft-AND operator:

$$F_i^{prod} = \mathrm{EMD}_i \cdot S_i \tag{9}$$

This form tends to suppress isolated high responses that appear in only one modality and therefore favours locations where both AEF and SAR indicate change. The third family uses the geometric mean,

$$F_i^{geom} = \sqrt{\max(\mathrm{EMD}_i, \varepsilon)\max(S_i, \varepsilon)} \tag{10}$$

which behaves similarly to the product but is less aggressively shrinking. The fourth family uses the quadratic mean,

$$F_i^{quad} = \sqrt{\frac{\mathrm{EMD}_i^2 + S_i^2}{2}} \tag{11}$$

which tends to preserve stronger responses from either source while still blending them smoothly. In the evaluation notebook these operators were instantiated for both EMD+CV and EMD+PWTT, producing the set of scalar baselines summarized later in the results section.

Conflict-damage monitoring often has to be deployed across multiple cities with limited labelled data, little time for retraining, and a need for transparent methods. A fixed scalar fusion rule can be computed directly from the two input maps, interpreted visually, and transferred without estimating city-specific parameters. This reasoning is consistent with earlier work showing that combining complementary features can stabilize high-resolution change detection (Wang et al., 2018) and with broader recent remote-sensing literature emphasizing feature fusion as a practical route to more robust change mapping (Lv et al., 2025). If a single weight selected on one benchmark remains close to optimal across multiple cities, then the resulting method is much easier to justify as a practical workflow than a more complex learned alternative.

3.4 Learned and adaptive fusion baselines

To test whether the simple scalar fusion is leaving substantial performance on the table, we also evaluated a lightweight learned fusion baseline based on logistic regression. We included logistic

https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-1331-2026

regression only to test whether slightly more flexible combinations of the same features could improve performance and transfer.

For each city, the logistic model is trained on the building-level scores extracted from the three base inputs EMD, REACTIV_CV, and PWTT. In addition to the raw scores, we include a small set of interaction and quadratic terms so that the model can represent modest non-linear combinations without becoming a complex black-box alternative. These predictors are standardized and passed to a Logistic Regression classifier with class_weight='balanced', lbfgs optimization, and max_iter=1000. Two evaluation modes are reported later. The first is within-city stratified 5-fold cross-validation, intended to estimate the best-case gain of a lightweight learned fusion when local labels are available. The second is leave-one-city-out (LOCO) transfer, in which the model is trained on seven cities and tested on the held-out eighth city. This provides a direct comparison between a fixed scalar rule and a simple learned fusion under realistic cross-site transfer.

3.5 Evaluation protocol

All pixel-level change maps were evaluated at building level, which is the primary target scale of the paper. For each method and city, the continuous score image was reduced over the PWTT/UNOSAT building polygons using the mean value within each footprint, producing one score per reference building with binary label class \in \{0,1\}.

The principal discrimination metric is building-level AUC because the task is ranking damaged versus undamaged buildings under strongly varying class imbalance. We also report the best F1 score from the precision-recall curve together with the corresponding precision, recall, overall accuracy, Cohen's kappa, and threshold.

The evaluation has three layers. First, we report direct per-city performance of the single-source baselines and scalar fusion methods on the full building inventory. Second, for the learned-fusion baseline we perform within-city stratified 5-fold cross-validation. Third, we run leave-one-city-out transfer experiments in which the logistic model is trained on seven cities and tested on the held-out city. Uncertainty for learned fusion is summarized through non-parametric bootstrap confidence intervals on AUC. Robustness is assessed primarily through a fusion-weight sweep for EMD+PWTT, which tests whether the transferred weight remains close to the city-specific optimum.

4. Results

4.1 Qualitative complementarity of AEF and SAR

The qualitative maps show that the two modalities contribute different strengths. In Avdiivka, PWTT produces the sharpest localized responses over the most visibly disrupted urban blocks, indicating high sensitivity to abrupt structural change between the pre-event and post-event Sentinel-1 stacks. However, it also shows scattered responses outside the main damage core, especially along heterogeneous urban edges and in mixed background areas. REACTIV-style CV reacts even more broadly to temporal instability and therefore highlights a larger amount of background texture.

By contrast, EMD in the AEF space is smoother and more semantically selective. Its strongest responses tend to align with broader zones of transformed urban fabric rather than isolated noisy pixels, and it suppresses part of the background variability that is more evident in the SAR-only maps. This produces cleaner regional contrast, but on its own EMD is less sharply focused than PWTT on abrupt building-scale disruption.

These contrasting behaviours are illustrated in Figure 2. The three individual source maps show distinct error modes and strengths: REACTIV-style CV is the broadest and noisiest, PWTT is the sharpest but still background-sensitive, and EMD is the most spatially coherent but less event-focused. This comparison explains why the fusion experiments focus on PWTT as the stronger SAR input while using EMD to add contextual stability.

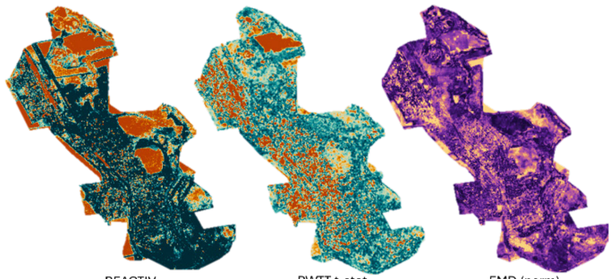

Figure 2. Individual change indices from Sentinel-1 and AlphaEarth for Avdiivka

4.2 Single-source baselines

The single-source baseline comparison shows that PWTT is the strongest individual method overall. Averaged across the seven cities, PWTT reaches a mean building-level AUC of 0.657 and a mean F1 of 0.294, compared with 0.593 and 0.257 for EMD, and 0.572 and 0.245 for REACTIV-style CV. PWTT therefore serves as the principal SAR baseline for the subsequent fusion analysis.

The city-level results also show that EMD contributes genuinely complementary information rather than simply acting as a weaker signal. EMD is the strongest single-source baseline in one heavy-damage case, Rubizhne, where it reaches AUC 0.693 compared with 0.683 for PWTT. In that city, the AEF representation already ranks damaged buildings effectively in heavily transformed urban fabric and therefore provides a strong basis for subsequent fusion.

REACTIV-style CV is the least stable baseline. It performs competitively in selected cities such as Sievierodonetsk and Kramatorsk, but in several others it is near-random or strongly threshold-unstable, with AUC values around 0.50 in Irpin and Hostomel. This supports the decision to treat REACTIV-style CV primarily as a secondary SAR comparator rather than the main fusion partner.

The results also show clear variation across damage regimes. In moderate-damage cities, PWTT is consistently stronger than EMD as a single-source method, for example in Avdiivka (0.624 vs 0.586), Hostomel (0.692 vs 0.610), and Lysychansk (0.609 vs 0.559). In the light-damage case of Kramatorsk, AUC remains relatively high for PWTT (0.726), but F1 drops to 0.056 because the damaged class is extremely sparse (see in Table 2). This anticipates the importance of reporting both ranking-based and threshold-based metrics.

City	EMD	REACTIV_CV	PWTT
Irpin	0.568	0.502	0.676
Sievierodonetsk	0.579	0.580	0.591
Rubizhne	0.693	0.584	0.683
Avdiivka	0.586	0.581	0.624
Hostomel	0.610	0.500	0.692
Lysychansk	0.559	0.582	0.609
Kramatorsk	0.559	0.675	0.726
Mean	0.593	0.572	0.657

Table 2. Building-level performance of single-source baselines across 7 cities

The qualitative CV-based fusion examples in Figure 3 provide a bridge between the single-source baselines and the broader fusion analysis. They do show some reduction of isolated activations relative to REACTIV alone, but they remain visibly more diffuse and less stable than the corresponding PWTT-based fusion maps. This is consistent with the quantitative results, where CV-based fusion rarely emerges as the strongest transferable option.

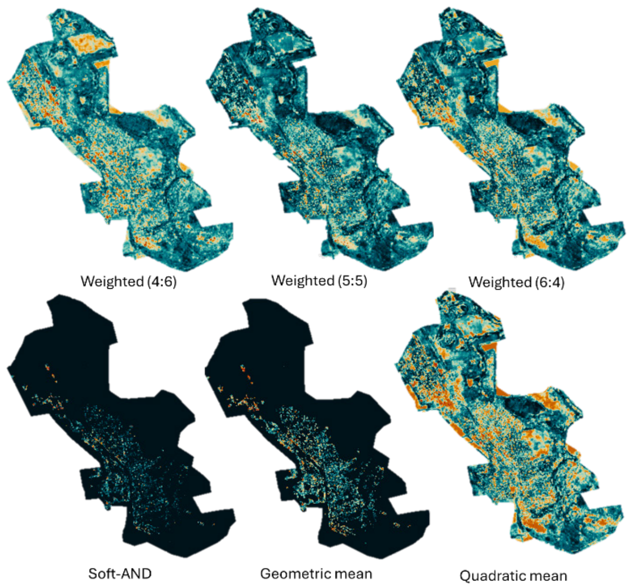

Figure 3. Fusion of AlphaEarth EMD with Sentinel-1 CV for Avdiivka using different scalar fusion rules

4.3 Scalar fusion performance across cities

Across the benchmark, combining AEF EMD with PWTT improves on the best single-source baseline on average and in most cities. Among the selected transferable methods, EMD+PWTT_w40 achieves the highest mean AUC at 0.676 and the highest mean F1 at 0.308. It is followed very closely by EMD_PWTT_quad with mean AUC 0.674 and mean F1 0.307. Both outperform PWTT alone (0.657 mean AUC) and clearly exceed EMD alone (0.593 mean AUC). By contrast, the multiplicative and geometric variants are less competitive on average, with mean AUC values of 0.656 for EMD*PWTT and 0.648 for EMD_PWTT_geom.

The strongest case-study improvement appears in Avdiivka, where EMD+PWTT_w40 reaches AUC 0.656, compared with 0.624 for PWTT and 0.586 for EMD. This is the clearest example of the intended complementarity: PWTT contributes sensitivity to abrupt change, while EMD reduces part of the non-damage background variability. Hostomel shows a similar but slightly smaller gain, with EMD+PWTT_w40 reaching 0.710 versus 0.692 for PWTT and 0.610 for EMD. Rubizhne provides the strongest heavy-damage example, where EMD+PWTT_w40 reaches 0.736, clearly above both EMD (0.693) and PWTT (0.683), and with an accompanying F1 increase to 0.588.

The per-city results reveal the same pattern. EMD+PWTT_w40 is best in Avdiivka, Rubizhne, and Hostomel, while EMD*PWTT is marginally best in Sievierodonetsk (0.619). In Lysychansk, the quadratic fusion is the strongest AEF+PWTT variant, with AUC 0.624 compared with 0.609 for PWTT and 0.559 for EMD. In Kramatorsk, EMD_PWTT_quad only narrowly exceeds PWTT, indicating that under very low damage prevalence the fusion does not materially change the ranking. Even where the gain is small, the AEF+PWTT family remains competitive and avoids the stronger instability seen in the CV-based alternatives.

The PWTT-based fusion family is illustrated more directly in Figure 4. Compared with the CV-based examples in Figure 3, the corresponding EMD+PWTT maps are generally sharper over the damaged urban core and less dominated by diffuse background response. The weighted and quadratic variants are especially useful because they preserve the stronger localized contrast of PWTT while damping part of the isolated activations that remain visible in the SAR-only map.

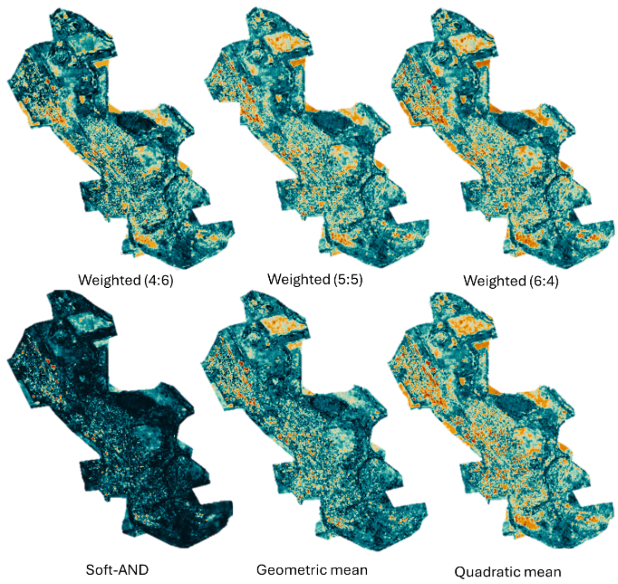

Figure 4. Fusion of AlphaEarth EMD with Sentinel-1 PWTT t-statistic for Avdiivka using the same scalar fusion rules

The Avdiivka case study in Figure 5 shows more clearly what this gain looks like in spatial terms. When the PWTT and EMD+PWTT_w40 maps are compared against the UNOSAT building-damage layer, PWTT alone produces several localized high-response clusters that are not well supported by the damaged-building footprints and spill more visibly into areas dominated by intact buildings. The fused map preserves the strongest responses over building clusters labelled as damaged, but suppresses part of these unsupported hotspots and reduces the visual tendency to flag entire micro-areas as changed. In that sense, the empirical improvement of EMD+PWTT_w40 is not only a small gain in average AUC, but also a more spatially specific pattern of response relative to the building reference (see in Table 3).

https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-1331-2026 | © Author(s) 2026. CC BY 4.0 License.

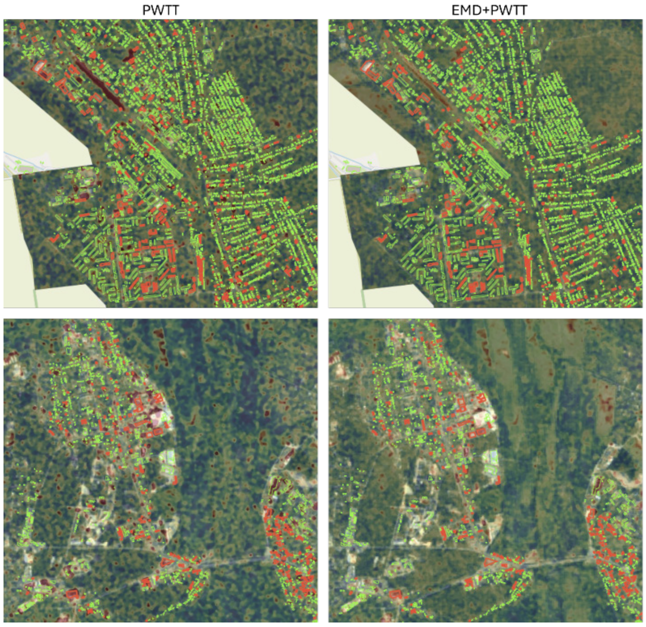

Figure 5. Visual comparison of PWTT (left) and EMD+PWTT_w40 (right) for Avdiivka with building-reference overlay

Overall, the gain does not come from a complex operator, but from a small set of simple AEF+PWTT fusion rules, especially the weighted sum and quadratic mean.

City	EMD	PWTT	E+P w40	E*P	E-P g	E-P q
Irpin	0.568	0.676	0.671	0.631	0.622	0.666
Sieviero-donetsk	0.579	0.591	0.614	0.619	0.614	0.607
Rubizhne	0.693	0.683	0.736	0.736	0.735	0.735
Avdiivka	0.586	0.624	0.656	0.647	0.642	0.655
Hostomel	0.610	0.692	0.710	0.678	0.672	0.705
Lysychansk	0.559	0.609	0.623	0.605	0.599	0.624
Kramatorsk	0.559	0.726	0.724	0.673	0.653	0.726
Mean	0.593	0.657	0.676	0.656	0.648	0.674

Table 3. Selected scalar fusion results across cities. E+P w40 denotes weighted EMD+PWTT fusion with w_EMD = 0.4; E*P denotes multiplicative fusion; E-P g and E-P q denote geometric and quadratic mean fusion, respectively.

4.4 Learned fusion, uncertainty, and transferability

The logistic-regression baseline provides a useful upper bound on what can be gained from lightweight supervised fusion with local labels, but that gain remains modest. Averaged across the seven cities, the best scalar method for each city reaches mean AUC 0.677, compared with 0.686 for the within-city logistic model evaluated by stratified 5-fold CV. Under leave-one-city-out transfer, however, the mean AUC drops to 0.614, below the fixed scalar alternatives. Learned fusion does not fail outright, but its benefit is mostly local: once transfer across cities matters, the simpler scalar rule is more reliable.

4.5 Sensitivity and robustness

The sensitivity curves in Figure 6 show that the transferred choice w_EMD = 0.4 is not always the exact per-city optimum, but it remains consistently close to it. Across the seven cities, the mean AUC gap between the best city-specific weight and the fixed w_EMD = 0.4 configuration is only 0.0028. Moreover, in six of the seven cities the local optimum lies within the interval 0.2 to 0.6, which supports the idea that a single transferred weight is meaningful across varied urban contexts.

The individual cities illustrate the same point. Avdiivka peaks at w_EMD = 0.6 with AUC 0.662, only 0.0057 above the fixed-weight result of 0.656. Irpin peaks at w_EMD = 0.2 with a somewhat larger but still moderate gap of 0.0094. Rubizhne peaks at w_EMD = 0.5, only 0.0015 above the fixed-weight result, and Lysychansk reaches its optimum exactly at w_EMD = 0.4. Overall, the local optima remain concentrated in a relatively narrow range and do not suggest any need for aggressive city-specific retuning.

The weight sweep also shows that performance changes gradually with the EMD contribution rather than abruptly. A single transferred parameter therefore appears sufficient for the main operational workflow: w_EMD = 0.4 remained within 0.0028 AUC of the city-specific optimum on average.

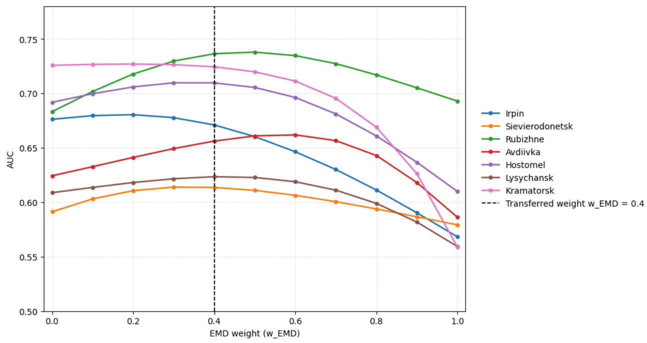

Figure 6. AUC as a function of the EMD fusion weight across validation cities

5. Discussion

5.1 Why fusion helps

AEF and SAR capture different but compatible aspects of damage. PWTT is the stronger single-source baseline overall because it reacts directly to abrupt structural change between the pre-event and post-event Sentinel-1 stacks, but it is also more vulnerable to heterogeneous urban background and local acquisition effects. EMD in the AlphaEarth embedding space plays a different role: it is spatially smoother, more semantically selective, and in some heavy-damage settings it competes directly with the SAR baselines. The fused maps improve ranking because they combine two only partly overlapping error patterns rather than relying on either source alone.

This interpretation is consistent with the city groups observed in the benchmark. In heavy-damage cities, EMD alone already captures broad urban transformation and can slightly exceed PWTT. In moderate-damage cities, PWTT remains the stronger single-source method, but EMD+PWTT_w40 improves it consistently enough to produce the clearest mean gain. The method is therefore strongest precisely in the regime most relevant to operational urban damage screening, where sharp SAR sensitivity benefits from contextual stabilization by the AEF branch.

The Avdiivka visual comparison supports the same reading at the level of local map behaviour. Relative to the UNOSAT building layer, PWTT alone shows several compact hotspots that are only weakly supported by damaged-building footprints and extend more visibly into intact building blocks. The fused map does not simply weaken the whole response surface; instead, it preferentially suppresses part of these unsupported activations while retaining stronger responses over the main damaged clusters. This is consistent with the idea that the AEF branch acts as a semantic stabilizer for SAR-driven change evidence rather than as a generic smoothing term.

5.2 Operational implications and limitations

The main practical advantage of the workflow is its simplicity: once the two source maps are ready, fusion requires no training and depends on a single transferred parameter. The learned baseline confirmed this — logistic regression added only a marginal within-city gain (AUC 0.686 vs. 0.677) and dropped to 0.614 under leave-one-city-out transfer. When transferability matters, the scalar rule is simply the safer choice.
Several limitations are worth stating plainly. Class imbalance is the most persistent one — Kramatorsk shows this clearly, where a respectable AUC coexists with an F1 of 0.056. The reference labels also carry their own uncertainty: UNOSAT damage inventories are reliable enough for comparative benchmarking, but they are not field-verified at the building level, which is why we describe the Avdiivka visual comparison in terms of supported and unsupported responses rather than confirmed errors.

Two further issues are harder to resolve cleanly. Spatial smoothing was tested but produced inconsistent results across cities, so we dropped it rather than risk overfitting the pipeline to specific urban morphologies. Our Earth Engine PWTT implementation also appears conservative relative to published dataset variants — we noticed this during cross-checks but did not resolve it within the scope of this study, and it likely leaves performance on the table. The Gaza analyses face a related issue: the reference data there require further harmonization before they can support the same kind of systematic evaluation, so we leave that for future work.

6. Conclusions

Fusing AlphaEarth embedding change with Sentinel-1 PWTT improves building-level damage discrimination across all seven Ukrainian cities in the benchmark. The preferred scalar variant, EMD+PWTT_w40, reached mean AUC 0.676 - a modest but consistent gain over PWTT alone (0.657) and a clear one over EMD alone (0.593). The more interesting finding, in our view, is how little tuning this requires. A fixed weight of w_EMD = 0.4, chosen without city-specific optimization, stayed within 0.003 AUC of the per-city optimum on average. That kind of stability is not guaranteed in multi-city transfer experiments, and it suggests the two sources are genuinely complementary rather than redundant. The learned fusion results push in the same direction. Logistic regression added a small gain locally (0.686 vs. 0.677) but dropped to 0.614 under leave-one-city-out transfer, worse than the fixed scalar rule.
What remains unresolved is the SAR branch. Our Earth Engine PWTT implementation appears conservative relative to published variants, and improving it is probably the most direct path to better results. Extending the benchmark beyond Ukraine and doing so with better-verified reference data than binary UNOSAT labels would also strengthen any conclusions about generalizability.

References

Aimaiti, Y., Sanon, C., Koch, M., Baise, L.G., Moaveni, B., 2022. War Related Building Damage Assessment in Kyiv, Ukraine, Using Sentinel-1 Radar and Sentinel-2 Optical Images. Remote Sensing 14(24), 6239. https://doi.org/10.3390/rs14246239

Ballinger, O., 2025. Open access battle damage detection via Pixel-Wise T-Test on Sentinel-1 imagery. Remote Sensing of Environment 331, 115025. https://doi.org/10.1016/j.rse.2025.115025

Boloorani, A.D., Darvishi, M., Weng, Q., Liu, X., 2021. Post-War Urban Damage Mapping Using InSAR: The Case of Mosul City in Iraq. ISPRS International Journal of Geo-Information 10(3), 140. https://doi.org/10.3390/ijgi10030140

Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J., Czerwinski, W., Pasquarella, V.J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Hanson, C., Guinan, O., Moore, R., Tait, A.M., Do, M.T., Stern, A., Huang, C., Pollard, K., et al., 2022. Dynamic World, Near real-time global 10 m land use land cover mapping. Scientific Data 9, 251. https://doi.org/10.1038/s41597-022-01307-4

Brown, C.F., Kazmierski, M.R., Pasquarella, V.J., Rucklidge, W.J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L.L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., Kohli, P., 2025. AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data. https://doi.org/10.48550/arXiv.2507.22291

Canty, M.J., Nielsen, A.A., Conradsen, K., Skriver, H., 2020. Statistical Analysis of Changes in Sentinel-1 Time Series on the Google Earth Engine. Remote Sensing 12(1), 46. https://doi.org/10.3390/rs12010046

Colin Koeniguer, E., Nicolas, J.-M., 2020. Change Detection Based on the Coefficient of Variation in SAR Time-Series of Urban Areas. Remote Sensing 12, 2089. https://doi.org/10.3390/rs12132089

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017: Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment 202, 18-27. https://doi.org/10.1016/j.rse.2017.06.031

Lv, Z., Zhang, M., Sun, W., Lei, T., Benediktsson, J.A., Liu, T., 2025. Land cover change detection with hyperspectral remote sensing images: A survey. Information Fusion 123, 103257. https://doi.org/10.1016/j.inffus.2025.103257

Mueller, H., Groeger, A., Hersh, J., Matranga, A., Serrat, J., 2021. Monitoring war destruction from space using machine learning. Proceedings of the National Academy of Sciences of the United States of America 118(23), e2025400118. https://doi.org/10.1073/pnas.2025400118

Scher, C., Van Den Hoek, J., 2025. Nationwide conflict damage mapping with interferometric synthetic aperture radar: A study of the 2022 Russia-Ukraine conflict. Science of Remote Sensing 11, 100217. https://doi.org/10.1016/j.srs.2025.100217

Wang, G.H., Wang, H.B., Fan, W.F., Liu, Y., Chen, C., 2018. Change detection of high-resolution remote sensing images based on adaptive fusion of multiple features. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-3, 1689-1694. https://doi.org/10.5194/isprs-archives-XLII-3-1689-2018