Supplementary Material

An open-source tool for mapping war destruction at scale in Ukraine using Sentinel-1 time series

Olivier Dietrich[1*], Torben Peters[1], Vivien Sainte Fare Garnot[2], Valerie Sticher[3], Thao Ton-That Whelan[4], Konrad Schindler[1], Jan Dirk Wegner[2]

[1]Photogrammetry and Remote Sensing, ETH Zurich, Zurich, Switzerland.
[2]EcoVision Lab, Department of Mathematical Modeling and Machine Learning, University of Zurich, Zurich, Switzerland.
[3]Department of Political Science, University of Zurich, Zurich, Switzerland.
[4]International Committee of the Red Cross, Geneva, Switzerland.

*Corresponding author(s). E-mail(s): odietrich@ethz.ch;

Supplementary Note 1 Detailed Comparison with PWTT

The Pixel-Wise T-Test (PWTT), proposed in[1], is a statistical approach designed to identify changes by comparing the mean values of two Sentinel-1 amplitude time series, adjusted by their standard deviation. We use it as a performance baseline since it utilizes the same data and has similar specifications, in particular, the scalability to large regions and the ease of deployment on GEE.

The PWTT can be calibrated to various conflict settings by optimizing a threshold value to fit existing damage assessments like our UNOSAT labels. For Ukraine in particular, the authors reported an optimal cutoff value of 1.63, based on data from eight different cities. However, their reported results cannot be directly compared to ours as they used different temporal ranges and different training examples for the *undamaged* category: they assume that every building footprint without a damage label from UNOSAT is intact, disregarding the fact that damages could have occurred after the date of the UNOSAT analysis, or could have been overlooked.

For a meaningful comparison, we keep the recommended, optimal cutoff value but rerun PWTT using our time periods and our own, stricter definition of negative labels. We report the corresponding results in Table S1. For readability, we repeat our results in Table S2. It should be noted that Chernihiv and Kharkiv, which represent 61% of our test set, were used by the original authors during the calibration of the PWTT. Nevertheless, our method consistently outperforms the baseline across all metrics.

Label	Precision	Recall	F_1	Accuracy	AUC
Damaged	61.9	76.3	68.4	75.5	75.7
Undamaged	85.6	75.1	80		

Table S1: **Evaluation metrics for PWTT.** Quantitative performance of the PWTT approach on our test set, with the cutoff value of 1.63 as recommended for Ukraine. All metrics are given as percentages.

Label	Precision	Recall	F_1	Accuracy	AUC
Damaged	67.1	84.6	74.9	80.3	81.3
Undamaged	90.5	78.0	83.8		

Table S2: **Evaluation metrics for our approach.** Quantitative performance of our approach on our test set, with confidence threshold 0.5 (reprinted from **??**). All metrics are given as percentages.

Supplementary Note 2 Choice of threshold

There is an inherent trade-off between precision and recall, and the optimal threshold depends on the specific application. To keep false alarms at a low enough rate, we set a target precision of 90%, for which we find an optimal confidence threshold of 0.655. That value has been used to generate all metrics in the present paper unless stated otherwise. With this threshold, the user can expect a recall of 46.5%. In practice, users of our tool can dynamically adjust the confidence threshold to suit their specific needs. For instance, one could start with a high threshold to discover hotpots with a very high likelihood of destruction, then gradually decrease the threshold to increase the recall and better assess the extent of destruction. Fig. S1 illustrates how our metrics vary with different thresholds.

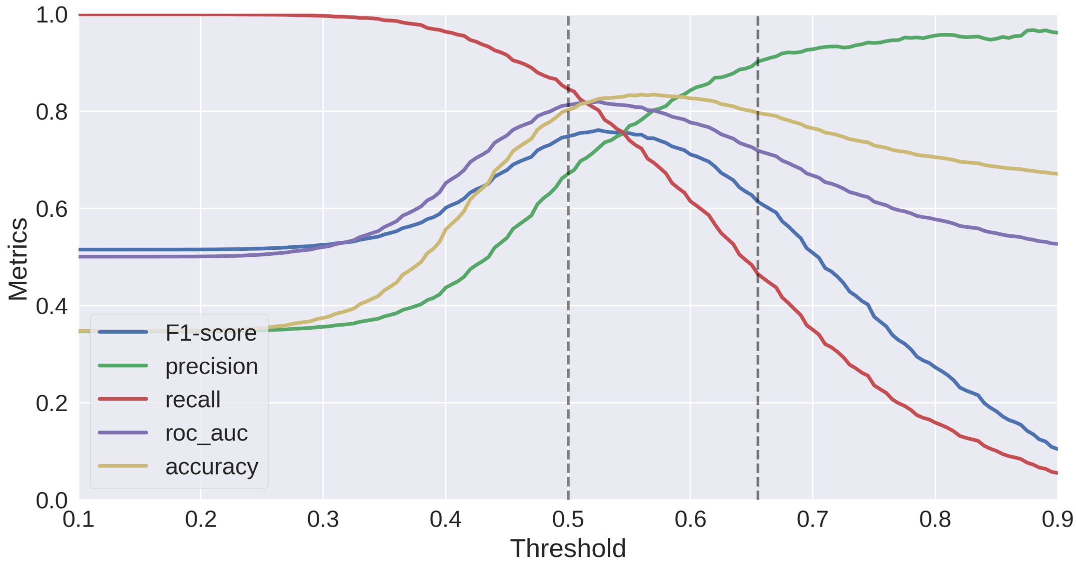

Fig. S1: **Performance of the model for different thresholds.** Pixel-level performance of our method on our test set, for different confidence thresholds. We set a target desired precision of 90% and found a confidence threshold of 0.655. The two vertical lines indicate the thresholds 0.5 and 0.655.

Supplementary Note 3 Temporal Evolution of Damage

The series of maps in Figure Fig. S2 illustrates the progression of building damage over the first two years of the conflict. The maps clearly delineate the primary locations of battles and the frontline, highlighting the areas most affected as the conflict persists.

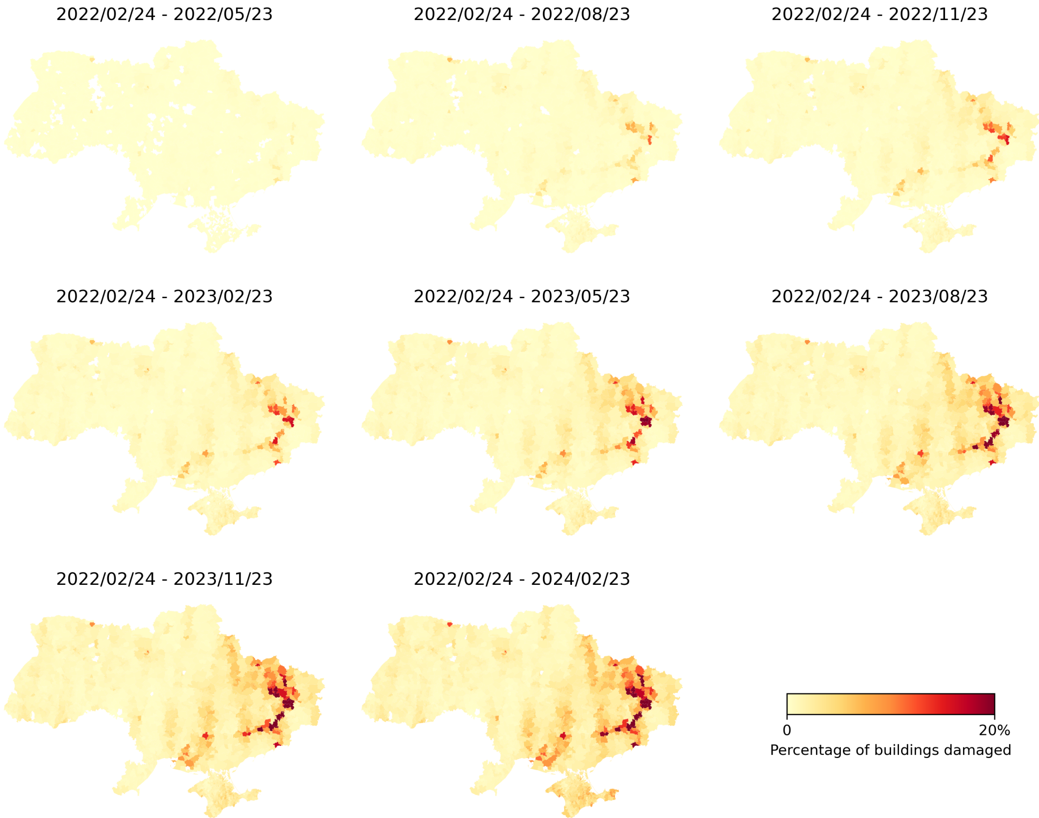

Fig. S2: **Evolution of damage distribution in Ukraine over time.** Percentage of buildings likely damaged for each period, cumulated since the beginning of the war and aggregated by *hromadas*. The predictions were thresholded at 0.655, and only buildings larger than $50\,\mathrm{m}^2$ were considered.

Supplementary Note 4 Limitations of Building Footprints

The building footprints provided by Overture Maps[2], which for Ukraine were compiled from OpenStreetMap[3] and Microsoft[4], inevitably come with limitations. For instance, OpenStreetMap data may be obsolete due to the demolition of older structures (Fig. S3.A) or the construction of new ones (Fig. S3.B). While Microsoft footprints serve as a valuable complement to OpenStreetMap data, they were automatically retrieved with a deep neural network from Bing Maps imagery, which may not always offer recent images, and suffers from gaps due to cloudy or outdated acquisitions, for example in Kherson (Fig. S3.C). Moreover, it is not uncommon for large non-building structures, e.g. containers (Fig. S3.D), boats (Fig. S3.E) or even aircraft (Fig. S3.F), to be wrongly identified as buildings.

Fig. S3: **Quality issues of existing building footprints.** All images cover 500×500 m^2 and are displayed alongside the corresponding building footprints sourced from Overture Maps. Only buildings larger than 50 m^2 are depicted. (A) Outdated buildings in Sevastopol. (B) Absent shopping mall near Chernihiv. (C) Missing footprints in Kherson. (D) Containers in Odessa. (E) Boats in Zaporizhzhia. (F) Antonov International Airport, Hostomel. Satellite images were obtained from ESRI/Maxar Technologies (A and C), Bing/Maxar Technologies (B, E, and F), and Google/Maxar Technologies (D).

Supplementary Note 5 UNOSAT Dataset

Our reference dataset has been created by combining all existing UNOSAT datasets for Ukraine[5]. We combined them into 18 distinct areas of interest (AOIs). For our purposes, we only use the two strongest labels, *destroyed* and *severely damaged*. For the sake of completeness, we also report the third one here, *moderately damaged*. For Mariupol and Kharkiv, for which multiple analyses were conducted by UNOSAT, we keep only the latest label assigned to a point, assuming that in the case of repeated assessments, the annotations were progressively refined. Table S3 summarizes our AOIs and the number of labeled locations in each of them.

AOI ID	Locations	UNOSAT product ID	# Destroyed	# Severely damaged	# Moderately damaged	Total
UKR1	Mariupol Azovstal	3371, 3300 3358	365	2184	3102	5651
UKR2	Vorzel Hostomel Irpin Bucha Moschun	3356 3359 3360 3363 3417	647	1116	670	2433
UKR3	Rubizhne Lysychansk Sievierodonetsk	3414 3444 3446	700	3182	1041	4923
UKR4	Volnovakha	3415	263	541	66	870
UKR5	Avdiivka	3443	36	505	86	627
UKR6	Chernihiv	3354	266	374	256	896
UKR7	Kharkiv	3357, 3455, 3454	56	168	241	465
UKR8	Borodyanka	3355	43	55	58	156
UKR9	Makariv	3403	18	64	21	103
UKR10	Mykolaiv	3404	33	66	14	113
UKR11	Shchastia	3405	4	12	1	17
UKR12	Sumy	3406	3	6	10	19
UKR13	Trostianets	3407	12	29	6	47
UKR14	Kramatorsk	3408	5	25	8	38
UKR15	Okhtyrka	3413	10	20	9	39
UKR16	Kherson Antonivka	3436 3435	6	90	9	105
UKR17	Kremenchuk	3437	1	7	3	11
UKR18	Melitopol	3442	3	19	13	35
	Total		2471	8463	5614	16548

Table S3: **Overview of ground truth** Summary of the 18 distinct areas of interest (AOI) in Ukraine, merging data from 29 UNOSAT products. The table lists counts of destroyed, severely damaged, and moderately damaged labels per area, totaling 16,548 labeled damage instances, or 10,934 after excluding the moderately damaged ones.

Supplementary Note 6 Ablation studies

We have performed an ablation study to support our design choices and understand the impact of various settings on our model. We kept the same train/test split and label definitions as described in **??**. To reduce the computational cost, we opted for a pixel-wise comparison to the UNOSAT labels, contrary to **??**, where we ran the inference densely over the entire area of interest and allowed for shifts up to ± 1 pixel in x and y to account for geo-location uncertainty of the labels.

The F_1-score of our optimal configuration, which utilizes both VV and VH Sentinel-1 bands, all seven features, 50 decision trees, and extracts the time series pixel-wise, is 0.72. Fig. S4 summarized the outcome for different settings. We observe the following characteristics:

Input bands: Our tests confirmed that VV polarization provides more relevant information for detecting destruction compared to VH. However, using both bands together produced the best results, indicating that they complement each other to some degree.

Input data: We tried replacing Sentinel-1 SAR data with Sentinel-2 optical images[6]. Pixel-wise Sentinel-2 time series were extracted from the Level-2A collection available on GEE[7], from which we filter out cloudy pixels using GEE's CloudScore+ mask[8]. Various band combinations were tried, but incorporating Sentinel-2 data never enhanced model performance. This was expected, as our pipeline was designed for Sentinel-1 properties, such as stable pre-event signals and a focus on temporal context over spatial context. Intuitively, it is harder to distinguish intact buildings from damaged ones by looking at the brightness change at an individual pixel location. Also combining Sentinel-1 and Sentinel-2 led to poorer results than using Sentinel-1 alone, likely due to the confounding effect of a six-fold increase in features, most of which contain little information for the task.

Model: We evaluated alternative classifiers available in GEE, i.e., Boosted Trees[9] and SVM[10]. Boosted trees showed weaker performance while taking approximately 3× longer to run. We were unable to obtain results for the SVM: when applied to our data it quickly exceeded the computational budget allowed by GEE.

Number of trees: We tested models with 10, 25, 50, 75, and 100 trees. Performance gains were minimal beyond 50 trees, so we chose 50 as the optimal number, balancing accuracy and computational cost. Above 100 trees, we again hit GEE's bound on the compute budget.

Features: We tried different configurations, from a simple combination of the mean and standard deviation to the entire set of seven features. We observed consistent gain as the number of features increased.

Extraction window: Instead of extracting the time series pixel-wise, we experimented with taking the spatial mean over a 3×3 neighborhood to suppress noise. The resulting heatmaps are smoother, but this comes at the cost of missing many small-scale and sparse damage events.

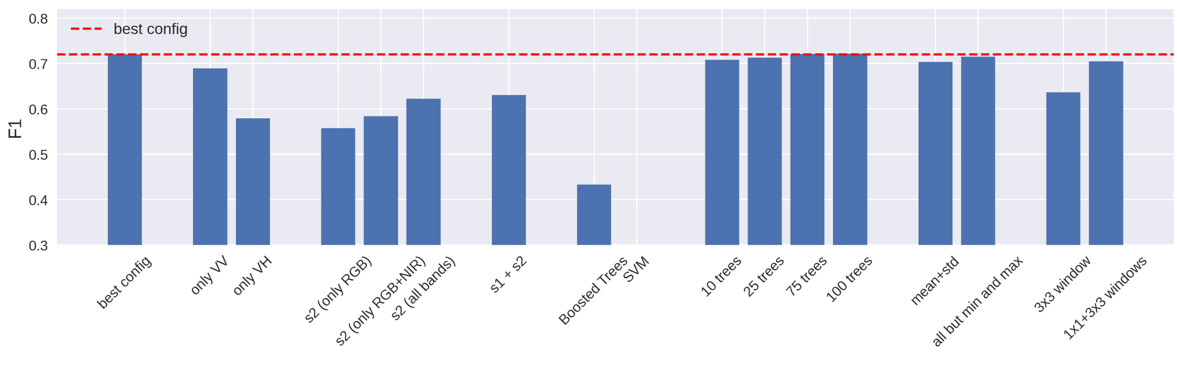

Fig. S4: **Results of the ablation studies.** F_1-score for different configurations. The dashed red line indicates the best configuration.

Supplementary Note 7 Changes in Backscatter due to Destruction

The backscatter signal from a building is influenced by numerous variables, including building characteristics (e.g., geometry, orientation, surface texture, material composition), satellite parameters (e.g., frequency, polarization, incidence angle, orbit direction), and environmental factors (e.g., vegetation, soil moisture, snow, or heavy rainfall). Intuitively, one might expect that intact structures yield stronger specular returns, with flat surfaces and right-angle corners acting as corner reflectors, whereas structural damage would disrupt these patterns, producing a more diffuse signal. However, several studies indicate that damage signatures in SAR data are highly variable and depend on the type and extent of destruction[11,12]. Damaged buildings can even exhibit backscatter signals as high as undamaged structures, or even higher[13].

Our experiments corroborate these findings. Fig. S5 presents pixel-wise backscatter time series that illustrate how conflict-related destruction appears in the data. High-resolution pre- and post-destruction satellite imagery from Google Earth is provided for reference. We observe that the backscatter amplitude following destruction can vary widely – increasing, decreasing, or remaining largely stable – depending on the location, the satellite orbit, and the polarization. These examples emphasize the need to utilize all available orbits in order to maximize reliability.

Fig. S5: **Example of Sentinel-1 time series** Sentinel-1 backscatter time series at selected UNOSAT locations for VV and VH polarizations, with VHR satellite images from Google Earth showing pre- and post-destruction conditions. Vertical dashed lines indicate the date of Russia's full-scale invasion (black) and the dates of the displayed pre-destruction (green) and post-destruction (red) images. For readability, only two orbits per location are shown, even if in some locations up to four orbits are available. Amplitude can decrease (B, C), increase (A, C) or remain relatively stable (B, D) after destruction. The same event may manifest differently depending on the flight direction (B), or even on the incidence angle within the same flight direction (C). Basemaps from Google Earth/Maxar Technologies.

Supplementary Note 8 World Bank Estimates

In this section, we compare our estimates, with a total of ≈400,000 buildings damaged between Feb 2022 to Feb 2024, to figures provided by the World Bank (WB) in their third Rapid Damage and Needs Assessment report[14].

According to the WB report, a total of 2,106,920 housing units were damaged by December 2023. Among these, 547,010 were destroyed, 679,382 sustained medium damage and 880,528 suffered minor damage. Importantly, the WB figures refer to housing units (HU), not buildings. A single-family home counts as one HU, but multi-family buildings may contain dozens of HU, making a direct comparison to our building-based estimates difficult.

In addition to damage estimates, the report provides the geographical distribution of damage cost. The WB attributes over 75% of the total damage cost to the following oblasts: Donetska (30 %), Kharkivska (27 %), Luhanska (12 %), and Kyivska (8 %). Another 18 % is distributed among Mykolaivska (4 %), Chernihivska (4 %), Khersonska (4 %), Zaporizka (3 %), and Dnipropetrovska (3 %). We use this spatial distribution of damage cost as a proxy to compare with our own estimates.

Fig. S6 shows the differences between our estimates and those of the WB. Excluding the *other* oblasts outside the direct conflict zone, both distributions align quite well, with indeed the majority of damage concentrated in Donetska and Kharkivska. However, our model estimates considerably more damage than the WB for the *other* oblasts, i.e., the ones further from the frontline. It appears that construction and other non-conflict-related changes dominate in those regions, which our model is unable to distinguish from building damages. This highlights that there is a trade-off between, on the one hand, comprehensive screening to detect also unexpected conflict-induced destruction far from the known frontline; and, on the other hand, accurate assessment by ruling out areas that are a priori unlikely to be affected.

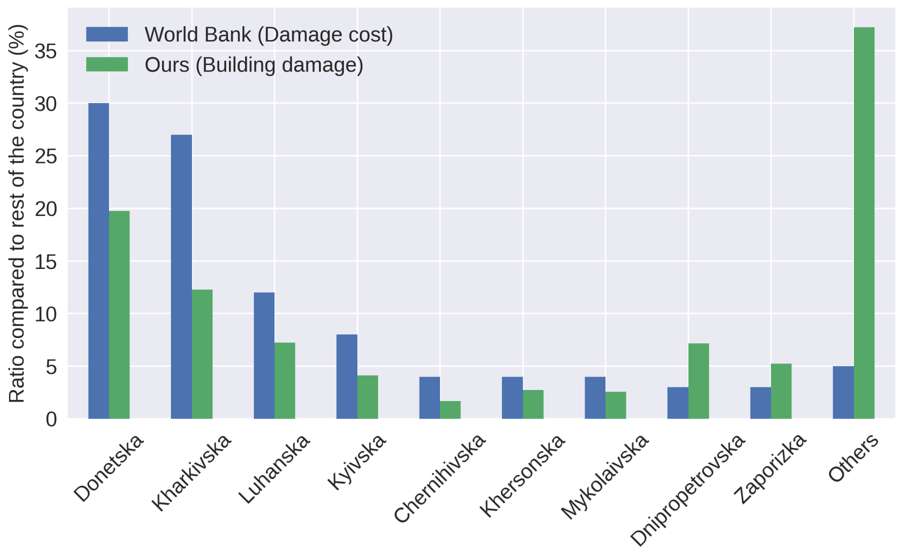

Fig. S6: **World Bank estimates vs Ours.** Comparison of the geographical distribution of the World Bank estimates against ours. The World Bank provides spatial distribution for their damage cost estimates, while our spatial distribution shows buildings likely damaged.

Supplementary References

[1] Ollie Ballinger. *Open Access Battle Damage Detection via Pixel-Wise T-Test on Sentinel-1 Imagery*. 2024. DOI: 10.48550/ARXIV.2405.06323. URL: https://arxiv.org/abs/2405.06323.

[2] Overture Maps Foundation. *Open Map Data*. Accessed: 2024-03-14. URL: https://overturemaps.org.

[3] OpenStreetMap contributors. *OpenStreetMap*. Accessed: 2024-03-14. URL: https://www.openstreetmap.org.

[4] Microsoft. *Global ML Building Footprints*. Accessed: 2024-03-14. URL: https://github.com/microsoft/GlobalMLBuildingFootprints.

[5] UNOSAT. *United Nations Satellite Center*. Accessed: 2024-03-14. URL: https://unosat.org/about-us.

[6] European Space Agency Copernicus Programme. *Sentinel-2*. Accessed: 2024-10-31. URL: https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2.

[7] Google. *Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A*. Accessed: 2024-10-31. URL: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED.

[8] Valerie J. Pasquarella et al. "Comprehensive quality assessment of optical satellite imagery using weakly supervised video learning". In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Ed. by IEEE. IEEE, June 2023, pp. 2125–2135. DOI: 10.1109/cvprw59228.2023.00206. URL: http://dx.doi.org/10.1109/CVPRW59228.2023.00206.

[9] Haifeng Li. *Smile - Statistical Machine Intelligence and Learning Engine*. Accessed: 2024-03-14. 2014. URL: https://haifengl.github.io.

[10] Chih-Chung Chang and Chih-Jen Lin. "LIBSVM: A library for support vector machines". In: *ACM Transactions on Intelligent Systems and Technology* 2.3 (2011), pp. 1–27.

[11] Simon Plank. "Rapid Damage Assessment by Means of Multi-Temporal SAR — A Comprehensive Review and Outlook to Sentinel-1". In: *Remote Sensing* 6.6 (May 2014), pp. 4870–4906. ISSN: 2072-4292. DOI: 10.3390/rs6064870. URL: http://dx.doi.org/10.3390/rs6064870.

[12] Pinglan Ge, Hideomi Gokon, and Kimiro Meguro. "A review on synthetic aperture radar-based building damage assessment in disasters". In: *Remote Sensing of Environment* 240 (2020), p. 111693. ISSN: 0034-4257. DOI: 10.1016/j.rse.2020.111693. URL: http://dx.doi.org/10.1016/j.rse.2020.111693.

[13] Dominik Brunner, Guido Lemoine, and Lorenzo Bruzzone. "Earthquake Damage Assessment of Buildings Using VHR Optical and SAR Imagery". In: *IEEE Transactions on Geoscience and Remote Sensing* 48.5 (May 2010), pp. 2403–2420. ISSN: 1558-0644. DOI: 10.1109/tgrs.2009.2038274. URL: http://dx.doi.org/10.1109/TGRS.2009.2038274.

[14] World Bank. *Ukraine - Third Rapid Damage and Needs Assessment (RDNA3): February 2022 - December 2023*. World Bank Group. Accessed: 2024-10-31. Washington, D.C., 2023. URL: http://documents.worldbank.org/curated/en/099021324115085807/P1801741bea12c012189ca16d95d8c2556a.