Point-by-point response to reviewers' comments

Reviewer #1

Comment 1.1. In this paper, the authors utilize satellite imagery and computer vision models to estimate internal displacement patterns during the Russia-Ukraine war. They collect over 1000 high-resolution images of urban areas in Ukraine from 2019 to 2022 and analyze spatial-temporal changes in car counts as a proxy for internal displacement. Their findings suggest an East-to-West movement of cars following the onset of the conflict, indicating potential patterns of internal displacement. However, they acknowledge limitations in data sparsity and the inherent constraints of satellite imagery and computer vision models. Despite these challenges, the study presents an innovative approach to estimate internal displacement and highlights opportunities for future research in this area.

The authors of this paper have conducted a serious, highly consistent, and clearly structured piece of work. I can recommend the acceptance of this paper once the authors address the following points:

Response 1.1. Thanks for highlighting the important contributions of our paper and acknowledging its scientific merit.

Comment 1.2. "While there have been wide-ranging international efforts to measure the scale of cross-border population flows, estimation of within-country migration flows are particularly prone to inaccuracies, lack of timely updates, and lack of disaggregation, e.g., by age (Checchi et al, 2017; Ratnayake et al, 2022)." -> I suggest including other aspects in the list, such as the use of data from different institutions, the lack of consistency in many of them, etc.

Response 1.2. We thank the reviewer for the valuable suggestion. We do agree that insufficient and inconsistent data collected from disparate sources often complicate the process even further. Please note that we present a more detailed discussion on this aspect in the fourth paragraph of the Introduction section (i.e., lines 74-84 of the revised manuscript).

Comment 1.3. A fundamental aspect is when the authors argue "Although public authorities have a good understanding of the population flow to and from receiving

countries, less is known about the Ukrainians' displacement pattern at the sub-national level (Rowe et al, 2022)." -> This aspect is crucial, and it is understood that different regions of the country do not carry out these counts officially, right? Therefore, it should be noted that there are no intermediary institutions from which to extract this data reliably.

Response 1.3. Yes, this is indeed a relevant aspect to be highlighted. We included a more explicit statement about this particular point in the revised manuscript. Please refer to lines 71-73 to see the changes.

Comment 1.4. "A large body of work has looked at the potential use of anonymized call detail records (CDR) from mobile phone operators to monitor mobility patterns, in particular in the context of communicable diseases, and most recently during the COVID-19 pandemic (Williams et al, 2015; Wesolowski et al, 2012, 2016; Oliver et al, 2020)." -> Not only that. Some studies in this regard measure communication patterns within particular communities such as Syrian refugees in Turkey, where mass displacements or settlements within their communities are understood because they maintain their own communication patterns among their own communities. This may be happening in the case of mass displacements in Ukraine in a war scenario, where local communities communicate with each other. I recommend seeing the work of Bakker et al. (2019) on the study of mobility and communication patterns of Syrian refugees in Turkey, who were fleeing from the conflict zone.

Response 1.4. Yes, we agree with the reviewer on this matter. Please note that we have reorganized the paragraph where we highlighted some case studies using CDR data. The revised manuscript version now includes a reference to Bakker's et al. (2019), as well as a more recent work conducted by Shibuya et al. (2024) to illustrate the use of CDR data in the context of internal displacement trends. We kindly refer the reviewer to lines 91-97 to see our modifications.

Comment 1.5. "While CDR data is undoubtedly useful, procuring access is a major hurdle, as are difficulties in quickly transferring a methodology from one country to another." -> Why the authors argue this? In theory, the CDR traces are independent of the country.

Response 1.5. Access to CDR needs to be negotiated with each operator individually. Apart from hindering fast and global access, this also means that data from different operators, if it can be obtained, are not necessarily directly comparable as operators

might apply different processing and aggregation protocols, and as they might have to comply with different national regulations. We expanded on this in lines 102-106.

Comment 1.6. "However, data coverage was much better during the months succeeding the start of the war than during the pre conflict period, especially for cities heavily involved in the conflict (e.g., Kharkiv, Donetsk, Mariupol, Odessa, and Ivano-Frankivsk; Fig. B3). This likely reflects operational adjustments by the satellite operator in response to the demand for images from the conflict areas."

Response 1.6. Please note that the reviewer did not make an explicit comment here.

Comment 1.7. About the dataset, the authors said that "we collected a total of 1009 very-high-resolution satellite images between 2019-2022, spanning 61 cities across Ukraine. Of these images, only 534 images remained for analysis after undergoing our data post-processing pipeline, with most cities remaining with less than 10 images throughout the time series (Fig. B2)." -> The authors should explain more in detail where they got these images (satellite program), spatial and temporal resolution, etc.

Response 1.7. We downloaded satellite images from Maxar's SecureWatch Platform, which has been recently rebranded as Maxar Geospatial Platform (MGP), https://www.maxar.com/maxar-intelligence/products/mgp-pro. The collected images have a ground sampling distance (i.e., spatial resolution) that varies between 0.3m, 0.4m, and 0.5m. However, this platform does not warrant a fixed temporal resolution for image collections where revisit times may range from days to weeks and months.

We kindly note that more detailed information about satellite imagery collection and processing was already mentioned in Section 4.3. In addition, we provided a more detailed explanation in Section 4.2 about how we identified the study region and areas of interest.

For improved clarity, however, we now added a guiding sentence in the Introduction section where we refer the reader explicitly to the Material & Methods section for more details. Please see lines 126-127 to see our changes.

Comment 1.8. "Mariupol, one of the most heavily affected cities by the war, suffered a massive drop in the number of cars during the first month succeeding the war when compared with the same month and region prior to the COVID-19 pandemic (Fig. 1a,b).

The opposite trend was observed for the city of Uzhhorod on the Slovakian border, with a substantial increase in the number of vehicles during the months succeeding the outbreak of the war when contrasted to the pre-war period (Fig. 1c,d)." -> Could you hypothesize that in a more threatened city, residents tend to hide their cars more in private garages or even in other secure spaces (The authors argue after that “Visual inspection suggests that such cases occurred mainly in dense residential areas where cars are likely hidden inside garages.”? Could we assume that many people could leave the country in their car, but not entire families?

Response 1.8. We thank the reviewer for this thought-provoking question. It is very likely that residents underwent several behavioral changes for increased protection, such as hiding their cars. This nevertheless remains highly speculative as there could be many reasons for such changes.

For instance, one might hide the car to protect it from damage, or to protect oneself from danger by avoiding attention. Moreover, there might be behavior changes due to the war that affect the pre-war link between (number of visible cars) and (number of people living in the area). For example, one might assume that, when fleeing, the cars could be more packed. Also, if a family has two cars, they might well offer their “unused’’ cars to friends and family to make sure that every car leaves at full capacity. As such, the drop in cars could be an underestimation of the drop in population. At the same time, the cars that remain might well become more hidden/protected. This could consequently lead to an overestimation of the population displacement. The precise link is thus hard to estimate.

We have emphasized this aspect more in the revised manuscript version. Please see lines 374-379 to see our changes.

With respect to the second question, it is unclear what the reviewer means by ‘family’, as it could extend to relatives like uncles, aunts, cousins, etc., all of which are likely to have their own cars. If the reviewer refers to the latter case, then we could certainly think of such a scenario.

Indeed, following a recent report published by the IOM (2024), most of the short- and long-term returnees in Ukraine were driven by family reasons (e.g., meeting relatives and friends). This therefore indicates that many Ukrainian families got separated during the war, with some members remaining in the country and others fleeing to nearby countries.

- IOM, 2024. Ukrainians and Third-Country Nationals Crossing Back to Ukraine - 2023 Regional Analysis. Accessed via: *https://dtm.iom.int/sites/g/files/tmzbdl1461/files/reports/Ukraine%20Response%20-%20Crossing%20Back%20to%20Ukraine%20-%202023%20Regional%20Analysis.pdf*

Comment 1.9. "At the quarter-of-year resolution, our results suggest an increasing and progressive trend of car displacement from East-to-West throughout the year" -> Can we observe these same patterns in those conflict regions where the majority of the population is pro-Russian?

Response 1.9. We certainly did observe some distinctive patterns in the eastern region of Ukraine. Following our results, we noticed that oblasts located closer to the eastern portion of the country were usually associated with an outflow of cars. Curiously, Luhansk and Kharkiv were the only oblasts that depicted an opposite trend. Given their proximity to the Russian border, this could suggest that this region has either received Ukrainians from nearby regions that eventually immigrated to Russia, or that it acted as a host to Russian immigration. We kindly note that these aspects are included in the manuscript. Please refer to Section 2.1 (lines 173-189).

Comment 1.10. The authors say that "in dense residential areas where cars are likely hidden inside garages." Could we talk about different patterns of behavior between spaces depending on their level of urbanization? After that, they argue “That is, if for a given area in a city the number of visible cars drops by, say, 40% compared to 2019, we assume that the population in that area has also dropped by 40%.” -> If we assume a different pattern of behavior depending on the level of urbanization, this correlation could be significantly different.

Response 1.10. Yes, we could undoubtedly expect different behavioral responses. As we briefly highlighted in lines 203-208 (section 2.2), we hypothesize that factors such as car ownership rates, street parking availability, and access to and quality of public transportation are likely to vary within and between cities.

That said, we agree with the reviewer that a fundamental limitation of the *ratio method* is the fact that it cannot account for changes in urbanization level. However, given the narrow time window in which we are conducting the analysis (2019-2022), it is very unlikely that a city (or an area within the given city) underwent significant urbanization, even more so when considering the 2 years of COVID-19 pandemic that lead to global economic recession.

Please note that we have mentioned more explicitly this issue now in the revised manuscript version. We kindly refer the reviewer to lines 216-221 to see our changes.

Comment 1.11. Finally, I suggest to the authors to incorporate a paragraph discussing the particularities of this region and whether the results obtained can be extrapolated to other regions. For example, there are certain specific factors in some regions such as certain regions in the east with a heavier pro-Russian population that may determine an E-W pattern, or the fact that factors such as the greater or lesser weight of population by sex or age may determine certain results in certain cities. In this sense, it is common in social science studies to incorporate references to studies such as "Scale, context, and heterogeneity: the complexity of the social space".

Response 1.11. It is indeed likely that the link between the displacement of cars and the displacement of people will be heterogeneous across regions. Concerning differences across different regions in Ukraine, unfortunately, we do not have sufficient insight into variation in the context and how it might translate to choosing (not) to leave the region by car. Concerning differences across international regions, we currently give the example of the “caminantes” leaving Venezuela (line 302) where we expect our approach to fail completely. Future survey-based research could help to shed more light on which contextual factors govern the choice of mode of transport in humanitarian disasters.

##########

Reviewer #2

Comment 2.1. This is a complex and substantial work, addressing a very dynamic component that might not be easily detected with remote sensing. It is a brave and dedicated effort. Analyzing dynamic aspects involving people, especially their movements, is challenging to define from satellite data. Quantitative methods often rely on capturing specific moments within this dynamism, but approximations can certainly be made. This paper demonstrates how satellite data can correspond with human movement during wartime. I strongly support the idea, as it addresses an important issue and illustrates where remote sensing approaches can aid in understanding complex political situations.

Response 2.1. Thanks for emphasizing the challenging nature of the problem and commending the proposed approach.

Comment 2.2. However, there are a few issues with the paper that need to be addressed. The paper does not follow the guidelines provided by the publisher and should be slightly revised. For example, the abstract should be no more than 200 words, but it currently

exceeds 300 words. Additionally, the length of the manuscript surpasses the allowed number of pages.

Response 2.2. We thank the reviewer very much for identifying the formatting issues. Please note that we have made the necessary amendments to follow the journal's guideline. This included:

- Shortening of the abstract (<200 words; see lines 13-40)
- Shortening of the keyword list (see lines 42-43)
- Removal of all footnotes
- Standardizing the references

We would nevertheless like to note that our current manuscript does not meet the length criteria suggested by the journal (11 typeset pages compared to 38 pages). We believe that this is rather a subjective measure, as our main text is below the journal's suggested word counts (4.500 words compared to roughly 3900 words).

Given that the journal does not impose limits on word count or page numbers (according to their own word), we will leave it up to the Editor to decide whether our manuscript needs to be shortened.

Comment 2.3. Apart from these points, I suggest a few small corrections, listed below (also see additional comments in the pdf version of the manuscript):

- Fig. 3: There is an error in part b) of this figure. Firstly, part a) is duplicated, and the title of the graph is incorrect (it should indicate the presence of snow).

Response 2.3. We thank the reviewer very much for pointing this out. We made all necessary corrections, whereby the reviewer can follow our updates in Figure 7 in the revised manuscript. Note that we conducted some minor arrangements to align with the journal's formatting layout, and as such figure and table orderings slightly changed.

We would also like to state that all minor comments made in the pdf version of the manuscript have been addressed. Below we highlight our changes to each comments:

Reviewer	Response
Include reference in the statement: ''*As of December 2023, over 28 million border crossings from Ukraine have been*	Kindly note that the appropriate reference was mentioned at the end of the referred sentence (UNHCR, 2023b; IOM, 2023).

registered since the start of the war (24 February 2022), [...]" - *Introduction section*	We have nevertheless updated the numbers to follow the latest official reports. Please refer to lines 62-66 of the revised manuscript to see the changes.
What is the motivation of this study? - Introduction section	As commented in an earlier paragraph, we aimed to evaluate whether internal displacement patterns can be estimated from the spatial-temporal changes in car counts detected from satellite imagery (see lines 122-124). To do so, we used the ongoing Russia-Ukraine war as a practical case study (see lines 150-151).
I don't think this is relevant. I believe the only argument you can add here is why you didn't use other - freely available and temporally denser images (like sentinel2). The reviewer refers to the following sentence in the Discussion section: ''*While our study focuses on one country, Ukraine, and one satellite image provider, Maxar, we believe that the computer vision approach for detecting and counting cars will generalize to other geographic contexts that are similar in terms of car, building, road types, as well as to other providers of very high-resolution satellite imagery (30- 50cm).*"	We respectfully disagree to some extent with the reviewer on this matter. Irrespective of satellite imagery provider, we believe that it is important to mention that our method can be generalized to other geographical contexts. If this would not be the case, then this could be a crucial limitation of our proposed method. Thus, we decided to keep this sentence in the revised manuscript version. We do, nevertheless, agree that the former manuscript version lacked additional information on why we did not use other, freely available, satellite imagery providers. We acknowledged this point more directly in the revised manuscript. Please see lines 481-487 (section 4.3).

Comment 2.4. Highlight the pros (e.g., transparency) and cons of your approach for detecting people displacement compared to other established methods for this statistic.

Response 2.4. Please note that we exposed extensively the pros and cons in our Discussion section.

As stated in lines 319-328, a major advantage of our proposed method relies on the fact that it could reduce the life-treating risks of the on-ground enumerators. In addition, with a fully automated pipeline and frequent imagery update, a method like ours could provide near real-time population estimates, which are essential during the acute phase of the humanitarian crisis. Lastly, our method is fully transparent, unlike the often biased black-box algorithms deployed by most social media platforms.

Along lines 329-379, we further exposed a handful of limitations, including: (i) price to acquire high-resolution images, (ii) limited availability of ground-truth data and its effect on AI models, and (iii) constraints in the satellite images (e.g., spatio-temporal coverage, weather-related obstructions, etc.).

Comment 2.5. Explain why you used Maxar images instead of freely available Sentinel-2 data. Since spatial resolution plays an important role in your study, this should be exposed in the text.

Response 2.5. The proposed approach for estimating internal displacement depends very much on the assumption that we can reliably detect individual cars in satellite images. Despite its better spatial coverage and cadence, Sentinel-2 images have a spatial resolution of 10 meters per pixel, which is too low for detecting small objects like cars. Most cars will occupy less than a pixel, making it hard to distinguish them from the background. Therefore, we resorted to Maxar images which offer higher spatial resolution (i.e., 0.3-0.5 meters per pixel), needed for reliable car detection results in this proof-of-concept study. Should there be significant advancements in object detection methods for detecting small, sub-pixel objects such as cars in low-resolution satellite images, then freely available Sentinel-2 images would certainly become a better data source for our use case.

That said, we note that this point is now more emphasized in the revised manuscript version. Please refer to lines 354-361 and 481-487 to see our changes.

Comment 2.6. Use consistent values for all variables across all graphs to facilitate easier comparison. For instance, in Figure 6, the width of the bars and the time intervals between cities vary, making it difficult to compare them intuitively.

Response 2.6. Thank you for your suggestion. The figure has been accordingly updated, and can be referred to in Figure 5 in the revised manuscript version.

Comment 2.7. Consider using additional data sources for verifying the number of people living in a city, such as household electricity usage.

Response 2.7. Conceptually, there could indeed be additional data sources for estimating how many people are residing in a city at a given time. For example, internet-connected smart electricity meters could be used, as could city-level electricity consumption data. Even sewage volume can provide signals about the population size (https://www.theguardian.com/commentisfree/2010/dec/20/boundary-changes-census-millions-lose-out).

Unfortunately, all of these data sources are difficult to obtain. Furthermore, city-level electricity consumption will be heavily affected by commercial and industrial use, which might not be an ideal signal for population presence. As such, we see the exploration of these and other additional data sources for something to explore in future work.