

Assessing OSM Data Usefulness to Support SDGs Using Temporal Evolution and Frequency Analysis - Quantification of Affected Population After Earthquakes Using the Examples of Indonesia 2018 and Türkiye 2023

Nikolaos Kolaxidis

M.Sc. Geography

pd281@uni-heidelberg.de

Heidelberg University

15.05.2023

Keywords: OpenStreetMap data, Sustainable Development Goals, earthquakes, resilience, population estimation, geographic analysis

1 Introduction

The United Nations has set 17 Sustainable Development Goals (SDGs) to be achieved by 2030, ranging from eliminating poverty and hunger to promoting sustainable economic growth and reducing inequality [1]. However, achieving these goals can be challenging, especially in low-income countries where data is often scarce and funding for statistical offices is inadequate [2]. Although valuable and necessary, traditional sources of data like national statistical offices, government ministries and international organizations have several drawbacks. The data is costly to collect and therefore the collection of data infrequent, which results in quickly outdated data. Additionally, reports are usually made on a national scale, so small scale variations are often not captured sufficiently. Last but not least, the openness and accessibility of data strongly depends on the origin of the providers and national security levels and the like [3].

To address the aforementioned issues of data costs, accessibility, currentness and scarcity, new, non-traditional sources of data are being explored to measure progress towards achieving the SDGs [3, 4]. OpenStreetMap (OSM), an open data community and the largest Volunteered Geographic Information project in the world, has been shown to contribute to filling existing data gaps [5, 6]. The data collected has been widely utilized in various domains, such as urban planning, disaster management, and public health [3, 7, 8]. The community comprises both individual volunteers, which contribute on an individual level, as well as organized mapping communities including corporate and humanitarian organizations, which contribute to OSM collectively through organized mapathons [5]. This has resulted in a growing collection of geospatial data on buildings, roads, and other features, which can be used to track progress towards the SDGs and other problems like data coverage and accessibility [1].

To show the usefulness of such data in comparison to official data, a specific SDG is picked and an analysis designed. A recent event with a large impact and a huge media coverage was the earthquake in Türkiye and Syria on 6th Febru-

ary 2023 [9]. A matching goal would be SDG 11, which focuses on "making cities and human settlements inclusive, safe, resilient, and sustainable" [10]. In regard to disasters, supporting the SDG 11 with different aftermath analyses enables using the derived information to build the resilience of residents against further disasters and therefore helps minimizing possible impacts later on [1, 8, 11]

Due to the recency of the event and its impact, it seems suitable to assess whether the collection of OSM data is fit-for-purpose today to support the mentioned SDG and any humanitarian aid organizations. Important for such organizations is for example the number of affected people during the event to plan helping processes accordingly [8, 12, 13]. Specifically, this data can be used to monitor progress towards Target 11.5 of SDG 11, which aims to "reduce the number of deaths and people affected by disasters and substantially decrease the direct economic losses relative to global gross domestic product caused by [them]" by 2030 [10]. The analysis of affected population after disasters, such as earthquakes, can inform the development of disaster risk reduction strategies and policies that align with SDG 11 [12, 13]. This is highlighted by the United Nations Office for Disaster Risk Reduction (UNDRR), which notes that disaster risk reduction is a critical element of achieving the SDGs, particularly SDG 11 [14]. That applies as well to SDG 9, which aims to "build resilient infrastructure, promote sustainable industrialization and foster innovation" [15], but this will not be looked at further in this elaboration.

To be able to compare the results with other older situations, another similar event is picked by using the PAGER index. According to the USGS, the latest largest earthquakes with a PAGER alert level of "red" were the 2023 earthquake in Türkiye and Syria (analysis will only look at Türkiye) and the 2018 earthquake in Indonesia [9], specifically on the 28th September 2018. For both events a map with the affected areas can be seen in the appendix. The PAGER alert level "red" means that there are more than 1000 estimated fatalities and estimated costs of more than one billion US dollars and therefore shows earthquakes with severe impact [16].

The research question is whether the OSM data is suitable for such an analysis compared to official data and if there is any difference before and after the event to evaluate whether there have been any mapping activities which

could alter the results. In the following, the used data for the analysis is explained, the methodology proposed and results are shown. Afterwards it is discussed whether the data, especially the OSM data, is fit-for-purpose and potential problems are pointed out.

2 Data

The task requires different datasets, from population data for the statistical analysis to the affected areas (areas of interest (AOI)). They are explained below.

2.1 Population data

Population data can be derived from WorldPop, a research programme based in the School of Geography and Environmental Sciences at the University of Southampton. It is "a multi-sectoral team of researchers, technicians and project specialists that produces data on population distributions and characteristics at high spatial resolution" [17]. It provides many different datasets in raster or csv format including urban change, migration flows and development indicators among others.

In order to calculate population statistics, the dataset *population counts* is most suitable. It is offered as a raster format file covering whole countries and the years 2000 to 2020 with a spatial resolution of 100 meters among other types. The data is downloaded for Indonesia for 2018 to reflect the population during the earthquake event and Türkiye for 2020, the latest available.

2.2 Copernicus Emergency Management Service

Information about the events themselves like the extents or damage grades are often not provided fast enough to be useful during or directly after a disaster event. Therefore the European Union (EU) launched the Copernicus Emergency Management Service (EMS) in 2012 to have readily available systems when an event occurs to provide necessary data for humanitarian aid in mere hours or days through rapid mapping. In addition to that, it provides risk and recovery mapping long before an event for mitigation and risk assessment and after an event to help recovery processes with and without humanitarian organizations in the affected areas [18].

There are different datasets offers, most suitable for this analysis are the so called *grading products* which "evaluate the severity of the damage resulting from the event" [18]. They are available as packages including vector files with different file types and topics like the extent of the AOIs and building damage grade among others. The data is downloaded for both events accordingly.

2.3 OpenStreetMap data

To be able to locate the AOIs of the EMS and get additional information about administrative zones, a special OSM dataset was used. *OSM-Boundaries* provides a download interface for administrative boundaries such as country borders, state borders or equivalents from the Open-

StreetMap databases [19]. There are multiple levels with increasing spatial scale, but for this task the administrative level four is sufficient for it includes all AOIs without being too large.

In addition to that, buildings data has been downloaded from OSM as well using the *Ohsome API* offered by the Heidelberg Institute for Geoinformation Technology [20]. With this client/web service, data for specific locations and timestamps is queryable and the mapping history reconstructible. This helps using the data state before and after the events to evaluate possible contributions in that time and assess the usefulness of the data.

3 Methodology

The methodology to calculate the affected population is quite common, for it only is the combination of population raster data and the AOIs from the EMS vector files, but in order to compare OSM to EMS data over time, multiple additional steps are required. Figure 1 shows the used workflow during the analysis and with it the proposed methodology. The workflow has been done four times, for each event (Türkiye and Indonesia) and again pre and post event, one day before (pre) and four weeks after (post) the event accordingly.

First the datasets are preprocessed using a combination of scripts to extract the relevant data from the EMS packages, sending queries to the Ohsome API and multiple GIS tools to reproject, clip, merge and dissolve specific datasets. Having the datasets ready, the first calculation is the simple one with dissolved AOIs and the clipped population data using QGIS's *Zonal Statistics* tool (see point 1 in figure 1). Additionally the area for all AOIs together is calculated. Using the OSM-Boundaries dataset, the AOIs are merged in their respective regions and the same calculation is done to get the population and area per merged AOIs/regions (see point 2 in figure 1). This helps increasing the resolution of the data to better evaluate differences afterwards.

Next a grid with one kilometer cell size is created spanning the extent of the AOIs to be able to calculate the population count and buildings count on an even higher spatial resolution. The count of buildings per cell is done using the *Count Points in Polygon* tool in QGIS for both the OSM and the EMS building data. At this point it should be mentioned that centroids have been calculated for the OSM buildings, which are shipped primarily as polygon geometries, to make them comparable to the centroids used in the EMS dataset. This is useful because only the count of buildings is used in the greater calculations rather than the building geometries themselves. These could be used to compare the mapped area of buildings with other datasets, but that is not part of this analysis [6].

Having calculated the counts of OSM and EMS buildings per cell (see point 3 in figure 1), the same is done again with the OSM-Boundaries dataset and merged AOIs to get values to compare to on the same extent as before (see point 4 in figure 1). This is needed because during the temporal evolution analysis of the data, which builds on top of the

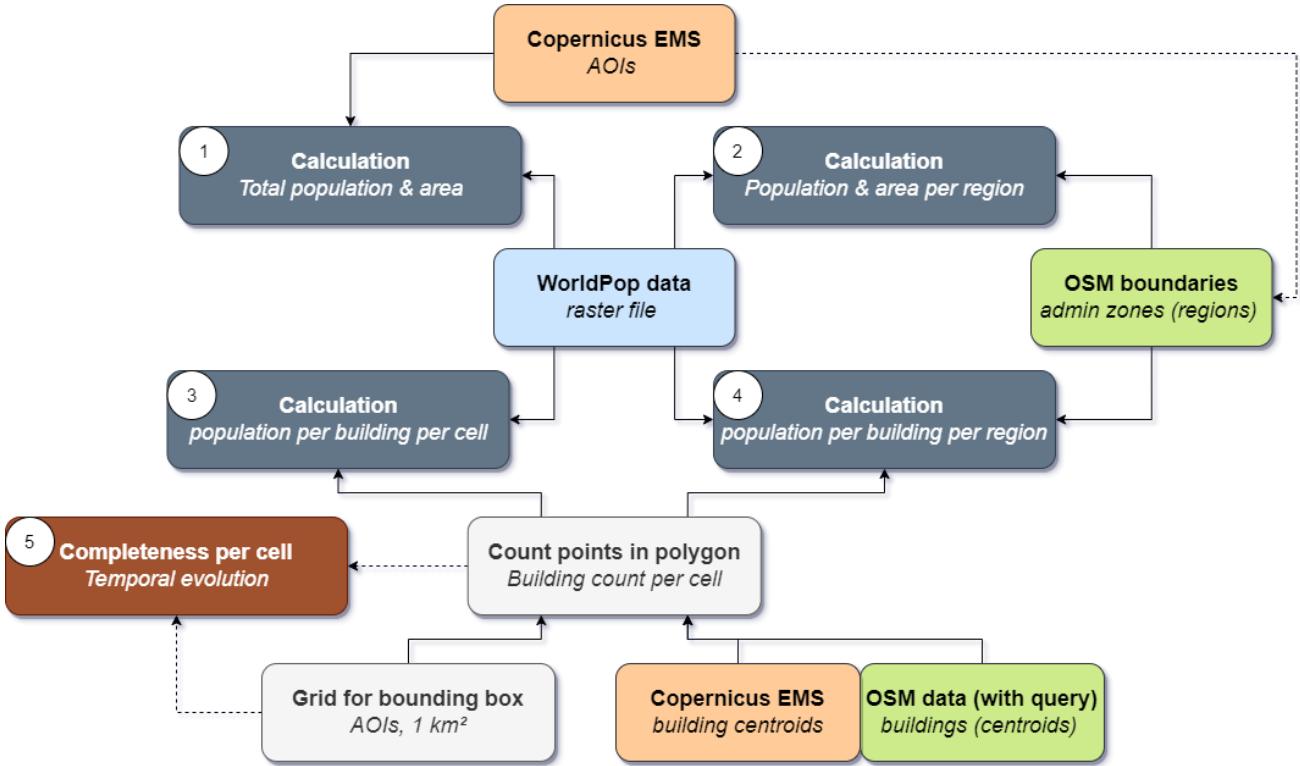


Fig. 1. Workflow of the analysis. Element coloring: Green, blue and orange are inputs, grey are intermediate steps, dark grey are calculations and brown is the completeness and temporal evolution analysis [own figure]

completeness calculation, the osm contributions are grouped by the regions. Speaking of, the completeness analysis is more of a comparison between the building counts of EMS and OSM, but calculated for pre and post event. With this the comparison of the frequency of the building counts or rather the count similarity is evaluated (see point 5 in figure 1). This is used because it is difficult and nearly impossible to find datasets with completely mapped buildings, for each and every building would have to be ground truthed [7, 6].

The frequency comparison is achieved by taking the two datasets from before, the grid and the building counts, and calculating statistics like difference and deviation of the counts. The resulting values are categorized into three comparing classes *less osm*, *similar* and *more osm* to show differences between OSM and EMS. In addition to that the history of OSM mapping contributions is looked at with the Ohsome Dashboard on <https://ohsome.org/apps/dashboard/> to take into account the temporal evolution of contributions made directly after the events. This can support claims about the effectiveness and positive effects of OSM data and the mapping community [6].

4 Results

The first calculation is intended to show the total population in the AOIs and their area respectively as seen in tables 1 and 2. The affected area during the Türkiye earthquake in 2023 is approximately three times larger, but the population in these areas is ten times larger than in the affected areas during the 2018 earthquake in Indonesia.

Table 1. Population and area per merged AOI in Türkiye

Region	Area (in km ²)	Population
Gaziantep	272	1,388,678
Diyarbakir	331	645,956
Kahramanmaraş	229	582,131
Şanlıurfa	125	527,429
Hatay	180	456,306
Malatya	141	352,833
Osmaniye	95	249,254
Adiyaman	148	233,690
Total	1522	4,436,277

Table 1 shows the population and area for the merged AOIs per region ordered by the population count for Türkiye in 2020 (latest available dataset as mentioned before). This will be used as a comparison base for the frequency and temporal evolution analysis afterwards.

Table 2 shows the same data for Indonesia in 2018. The regions are similar large but there are less regions than in the results for Türkiye 2023 as seen in 1. It seems that the population density is lower in the Indonesian regions, but that is not a topic to discuss in this elaboration.

Table 2. Population and area per merged AOI in Indonesia

Region	Area (in km ²)	Population
Palu	142	257,440
Sigi Regency	238	158,291
Donggala Regency	151	27,997
Total	532	443,728

Mean values for the calculation of the population per building per cell as explained in point 3 of the methodology can be seen in tables 3 and 4. Interesting is the fact, that there is no large difference between the OSM data pre and post the event, but both are lower than the EMS data. In the Türkiye datasets there is a large difference between the OSM data pre and post the event and the EMS data lies between them.

To generate spatial structured values, the same calculation is done with the OSM-Boundaries data as stated in point 4 in the methodology. The resulting values can be seen in table 3. Prior claims can be confirmed with these values as well.

Table 3. Population per building per region for Indonesia

Regions	EMS	OSM pre	OSM post
Donggala Regency	38.1	1.9	1.9
Palu	4.2	2.6	2.4
Sigi Regency	7.5	3.5	3.2
Mean per cell	38.6	20.8	19.6

The results of the same calculation for Türkiye can be seen in table 4. Here the distribution of the values is far more diverse for there is for example a very high EMS mean value in the Diyarbakir region and a comparable low value in the Osmaniye region. The regions Malatya and Şanlıurfa have comparable values in EMS and OSM pre data, but post event OSM data seems to always be much lower than the other two datasets except for the mentioned Diyarbakir region. Here only the EMS data values are high, pre and post OSM data is quite similar. This raises the questions why there are so huge differences in the building count between the regions and when did the OSM counts change.

For both events the results of the population per building per cell have been visualized in histograms in figures 2 and 3. The first thing to mention is that the x-axis limit is different due to the very different distribution of values along the full scale of the values. Additionally, the EMS histogram and OSM post histogram have a different course in Türkiye than in the Indonesia datasets. The largest difference can be seen in the temporal evolution of the OSM data

Table 4. Population per building per region for Türkiye

Region	EMS	OSM pre	OSM post
Gaziantep	1,601.7	134.9	9.2
Adiyaman	84.4	53.5	7
Diyarbakir	15,022.2	43.3	43.1
Kahramanmaraş	77.1	198.9	9.3
Malatya	102.5	108.8	8.4
Osmaniye	20.4	606.5	8.3
Şanlıurfa	88.5	89	7.5
Hatay	88.5	726.6	7.7
Mean per cell	371.4	594.5	49.1

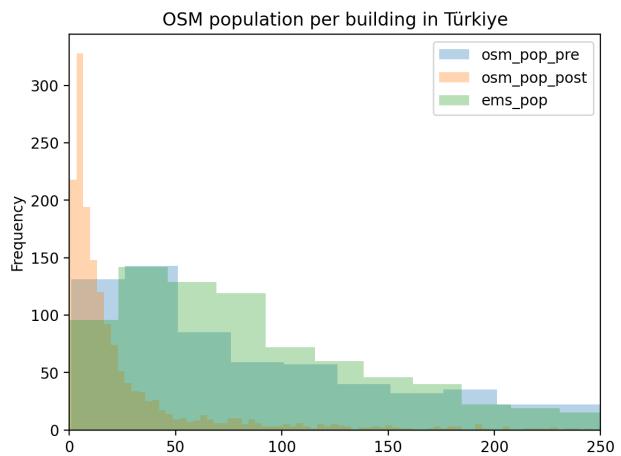


Fig. 2. Histogram of the population per building per region for Türkiye [own figure]

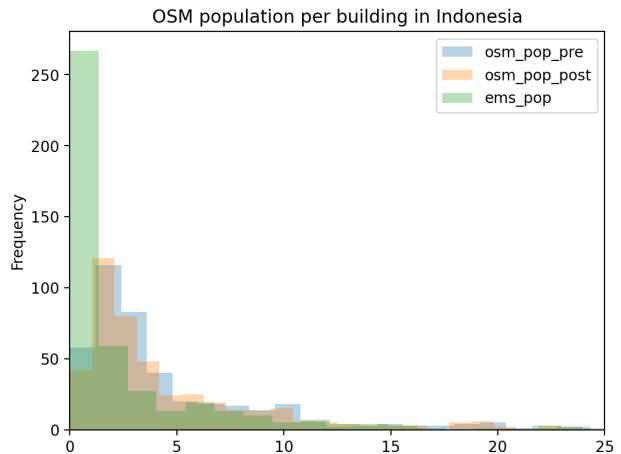


Fig. 3. Histogram of the population per building per region for Indonesia [own figure]

in Türkiye where from being mostly distributed from 0 to 200 people per building the counts evolved so as that the val-

ues decreased strongly to 0 up to 20 people per building. In Indonesia, the EMS values range between 0 and 1 person per building, which could hint towards either too many mapped buildings or buildings where no population is present.

To support this data and to derive data in order to answer the aforementioned question, the results from *Ohsome* are looked at in 4 and 5 as well as a map of 1:1 comparison between EMS and OSM values per cell in figure 10 in the appendix. It is noticeable that there are generally more cells in the post event data and a large shift towards cells with more OSM data. However, this does not apply to every AOI for example in Gaziantep and Diyarbakir large parts of the AOI shows cells with more OSM data during both times. On the other hand, AOIs like Hatay and Osmaniye show the shift towards cells with more OSM data as mentioned before in their boundaries.

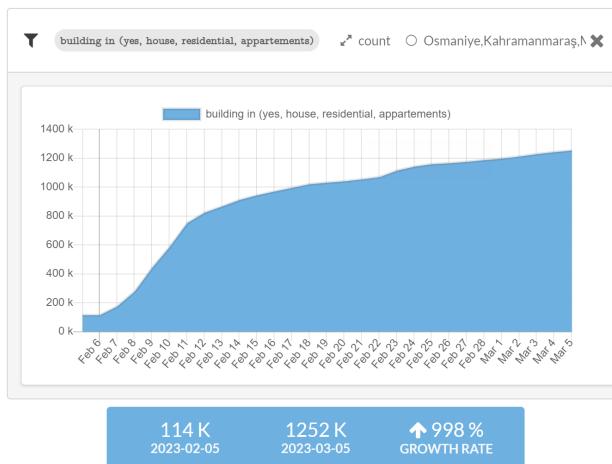


Fig. 4. OSM mapping activity from 2023-02-05 to 2023-03-05 for the AOI in Türkiye [adapted from Ohsome API]

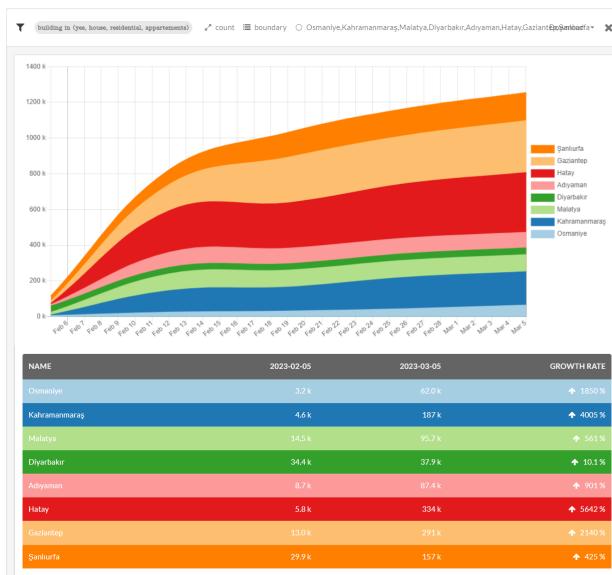


Fig. 5. OSM mapping activity from 2023-02-05 to 2023-03-05 grouped by regions in Türkiye [adapted from Ohsome API]

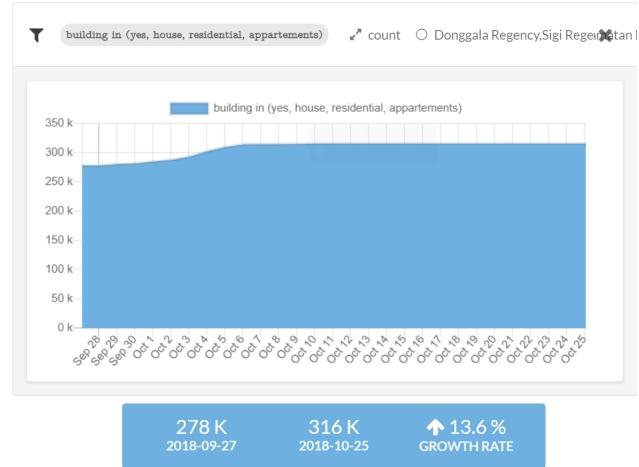


Fig. 6. OSM mapping activity from 2018-09-27 to 2018-10-25 for the AOI in Indonesia [adapted from Ohsome API]

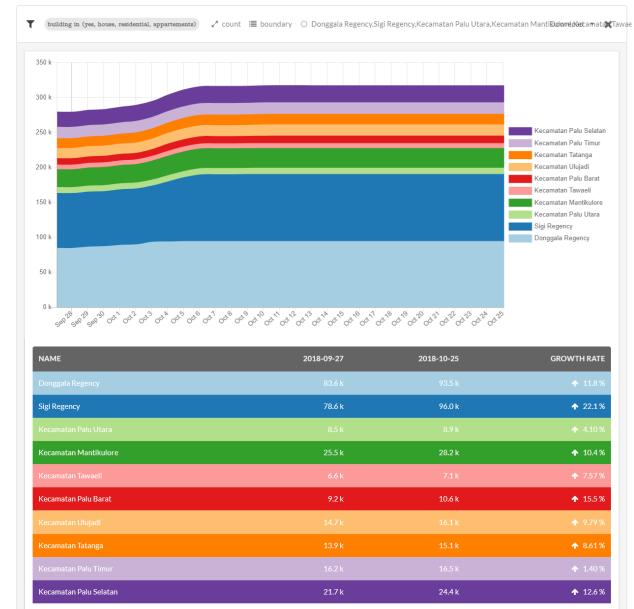


Fig. 7. OSM mapping activity from 2018-09-27 to 2018-10-25 grouped by regions in Indonesia [adapted from Ohsome API]

Using *Ohsome*, the history of OSM contributions can be shown in a histogram. In figure 4 it is clearly shown that there is a huge increase of contributions directly after the disaster event on the 6th February 2023. The growth rate is shown as nearly 1000 % in the four weeks time period after the event.

If grouped by regions as seen in figure 5, the growth rate is split strongly between the regions as claimed before. In Diyarbakir, which has been mentioned before as an AOI with a seemingly small shift from cells with less OSM to cells with more OSM data, the increase in contributions is only 10 %. In Hatay on the other hand, which has been mentioned as the opposite, the increase is nearly 5650 %.

To compare those results, the same procedure has been done for Indonesia. Again there is the map in figure 11 in the appendix which shows the 1:1 comparison between EMS and OSM values per cell as well as the Ohsome responses

in figure 6 and figure 7. Even on the smaller overall area of the AOIs the general cell count increases after the disaster event and there is a shift from cells with less OSM data to cells with more OSM data as well. Interestingly, the cells with similar building count increases in opposite to the data in Türkiye.

The largest shift towards cells with more OSM data is in the Palu region and the northern Sigi Regency region. The other regions all have more OSM data available prior to the event. The Ohsome data further confirms the claims in figure 6 as the total increase in OSM contributions is only 13.6 %, way less than in Türkiye. However, the contributions in total are at approximately 280 thousand, in direct comparison to Türkiye less, but due to the smaller AOI area still relatively similar.

If grouped by regions as seen in figure 7, it is shown that the differences in increase are not as diverse as in Türkiye (note: Kecamatan means district in Indonesian and all entries with the prefix "Kecamatan" form the Palu region mentioned above). The values range between 1 % to 15.5 % in the Palu region and between 1 % and 22 % for all AOIs in Indonesia.

Therefore there is a clear difference between the data in Türkiye and the data in Indonesia which will be discussed in the following section.

5 Discussion and conclusion

There are differences in the OSM contribution counts between the two events and in the population per building values per regions as well as per cells. One possible explanation could be that the OSM mapping community is often organized in groups which use mapathons to contribute data to specific regions after disaster events rather than constantly in a high spatial and temporal density all over the world [1, 8, 21]. Due to this, the mapping activity in a region increases strongly directly after an event. Therefore, regions where multiple events happen in shorter time like Southeast Asia, which are prone to disaster events so to speak, have usually already more complete mapping coverage than regions where events are not happening as often [7, 14, 22]. Contrary, developed countries like the ones in Europe are highly dense mapped even when not being as prone to disasters [7, 12]. This suggests that the differences are largely a combination of socio-economic origin and the proneness to disasters [7].

Is the OSM data fit-for-purpose?

As mentioned in Barron et al. [6], there are many different quality indicators to assess this question. In this elaboration primarily the comparison of the count frequency between EMS data and OSM before and after the event has been assessed. The results show a high OSM contribution count directly after the event and ongoing mapping activities even four weeks after the event. The EMS data was published early after the event and was updated only up to one week, which is not as frequently as the OSM data. This alone would suggest that the possibility of further contributing data to a freely accessible platform like OSM is very useful and helps humanitarian aid organizations in their work

even when the Copernicus data is funded by official institutions and has a probable higher quality [7, 18]. EMS seems to be a very good first estimation service for direct humanitarian aid, but OSM is better suited for later restoration processes and ongoing data collection without the need of large funding [7, 22].

Additionally to all statements, OSM provides different datasets to support others like administrative zones, which are available freely. Although they have to be checked for positional accuracy as mentioned in Barron et al. [6], they still provide a fast way to get good enough data for such an analysis. Because of the easy accessibility, this data is useful for a lot of applications and analyses and should often be considered. However, the lack of a possible counter-check with ground-truthing of the OSM data has to be considered if doing a highly real-life dependent analysis or planning.

Nevertheless, due to the constant updating and quality checking, it is very probable that the OSM data will get better and better in the future [7, 8]. Therefore the OSM community will play an even increasing role in humanitarian aid and shows the high usefulness of freely available, volunteer-driven data.

Overall, the use of open data communities like OSM presents a promising solution to filling data gaps and achieving SDGs, particularly in low-income countries where traditional data sources are inadequate [3, 8, 12]. By contributing to these communities, individuals and organizations can help improving the quality and accessibility of geospatial data, which is essential for tracking progress towards SDGs and help humanitarian aid organizations to plan their help processes accordingly.

At last a daring but almost confirmed thesis based on this elaboration: OSM can save lives and should always be considered a useful addition for every humanitarian aid application. Researching possibilities to enhance the OSM data is viable in the future, for doing analyses and saving lives alike.

References

- [1] Solis, P., and Zeballos, M., 2023. "Where YouthMappers connect to the SDGs". In *Open Mapping towards Sustainable Development Goals*, P. Solis and M. Zeballos, eds., 1st ed., Vol. 2. Springer, Cham, Switzerland, ch. 1.2, pp. 2–5.
- [2] United Nations Office for the Coordination of Humanitarian Affairs (OCHA) Centre for Humanitarian Data, 2023. The State of Open Humanitarian Data 2023. - URL: <https://centre.humdata.org/the-state-of-open-humanitarian-data-2023/> [as of 07.05.2023].
- [3] Fritz, S., See, L., Carlson, T., Haklay, M., Oliver, J., Fraisl, D., Mondardini, R., Brocklehurst, M., Shanley, L., Schade, S., Wehn, U., Abrate, T., Anstee, J., Arnold, S., Billot, M., Campbell, J., Espey, J., Gold, M., Hager, G., He, S., Hepburn, L., Hsu, A., Long, D., Masó, J., McCallum, I., Muniafu, M., Moorthy, I., Obersteiner, M., Parker, A., Weisspfugl, M., and West, S., 2019. "Citizen science and the united nations sustainable de-

- velopment goals". *Nature Sustainability*, **2**, pp. 922–930.
- [4] Sun, L., Chen, J., Li, Q., and Huang, D., 2020. "Dramatic uneven urbanization of large cities throughout the world in recent decades". *Nature Communications*, **11**(5366).
- [5] Anderson, J., Sarkar, D., and Palen, L., 2019. "Corporate Editors in the Evolving Landscape of OpenStreetMap". *International Journal of Geo-Information*, **8**(232).
- [6] Barron, C., Neis, P., and Zipf, A., 2014. "A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis". *Conflict and Health*, **18**(6), pp. 877–895.
- [7] Herfort, B., Lautenbach, S., de Albuquerque, J., Anderson, J., and Zipf, A., 2021. "The evolution of humanitarian mapping within the OpenStreetMap community". *Scientific Reports*, **11**(3037).
- [8] Scholz, S., Knight, P., Eckle, M., Marx, S., and Zipf, A., 2018. "Volunteered Geographic Information for Disaster Risk Reduction—The Missing Maps Approach and Its Potential within the Red Cross and Red Crescent Movement". *Remote Sensing*, **10**(1239).
- [9] USGS, 2023a. Latest Earthquakes. - URL: <https://earthquake.usgs.gov/earthquakes/map/> [as of 10.05.2023].
- [10] United Nations, 2023. Goal 11: Make cities inclusive, safe, resilient and sustainable. - URL: <https://www.un.org/sustainabledevelopment/cities/> [as of 07.05.2023].
- [11] Tan, M., 2023. "Building Resilience and Community-Based Disaster Risk Management (CBDRM): Experiences and Lessons from Communities in the Philippines". In *International Handbook of Disaster Research*, A. Singh, ed. Springer, Singapore.
- [12] Greenough, P., and Nelson, E., 2019. "Beyond mapping: a case for geospatial analytics in humanitarian health". *Conflict and Health*, **13**(50).
- [13] Humanitarian OpenStreetMap Team, 2023. Open Mapping for SDGs: Goal 11 - Sustainable Cities and Communities. - URL: <https://sdgs.hotosm.org/pages/part-iv-using-open-data-and-maps-to-meet-and-monitor-sdgs/sdg-goal-11/> [as of 13.05.2023].
- [14] UNDRR, 2015. Disaster risk reduction and resilience in the 2030 agenda for sustainable development. Reflection report, United Nations Office for Disaster Risk Reduction. - URL: <https://www.undrri.org/publication/disaster-risk-reduction-and-resilience-2030-agenda-sustainable-development>.
- [15] United Nations, 2023. Goal 9: Build resilient infrastructure, promote sustainable industrialization and foster innovation. - URL: <https://www.un.org/sustainabledevelopment/infrastructure-industrialization/> [as of 12.05.2023].
- [16] USGS, 2023b. PAGER Scientific Background. - URL: <https://earthquake.usgs.gov/data/pager/background.php> [as of 10.05.2023].
- [17] WorldPop, 2023. WorldPop. About. - URL: <https://www.worldpop.org/about/> [as of 09.05.2023].
- [18] European Union, 2023. The Emergency Management Service - Mapping. - URL: <https://emergency.copernicus.eu/mapping/ems/emergency-management-service-mapping> [as of 09.05.2023].
- [19] Ground Zero Communications AB, 2023. OSM-Boundaries. About. - URL: <https://osm-boundaries.com> [as of 09.05.2023].
- [20] Heidelberg Institute for Geoinformation Technology, 2023. OpenStreetMap History Data Analytics Platform. - URL: <https://heigit.org/big-spatial-data-analytics-en/ohsome/> [as of 12.05.2023].
- [21] Humanitarian OpenStreetMap Team, 2023. Humanitarian OpenStreetMap Team. - URL: <https://www.hotosm.org> [as of 13.05.2023].
- [22] Minghini, M., Sarretta, A., Lupia, F., Napolitano, M., Palmas, A., and Delucchi, L., 2017. "Collaborative mapping response to disasters through OpenStreetMap: the case of the 2016 Italian earthquake". *Geologia e Mineraria Ambientale*, **54**(2), pp. 21–26.

Appendix

The appendix will continue on the next page and will provide the large maps mentioned before.

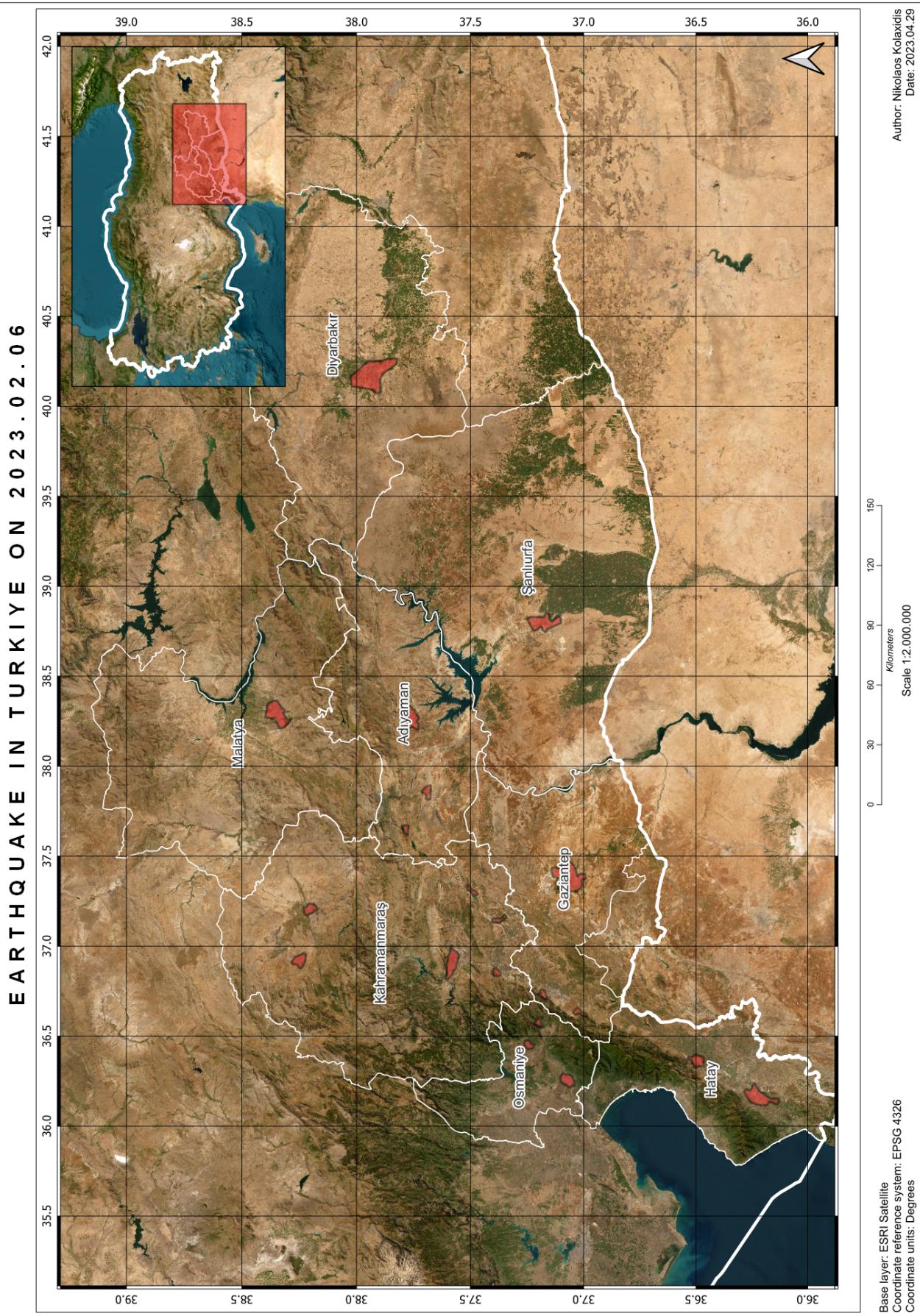


Fig. 8. Map of Türkiye with regions and AOIs during the earthquake [own figure]

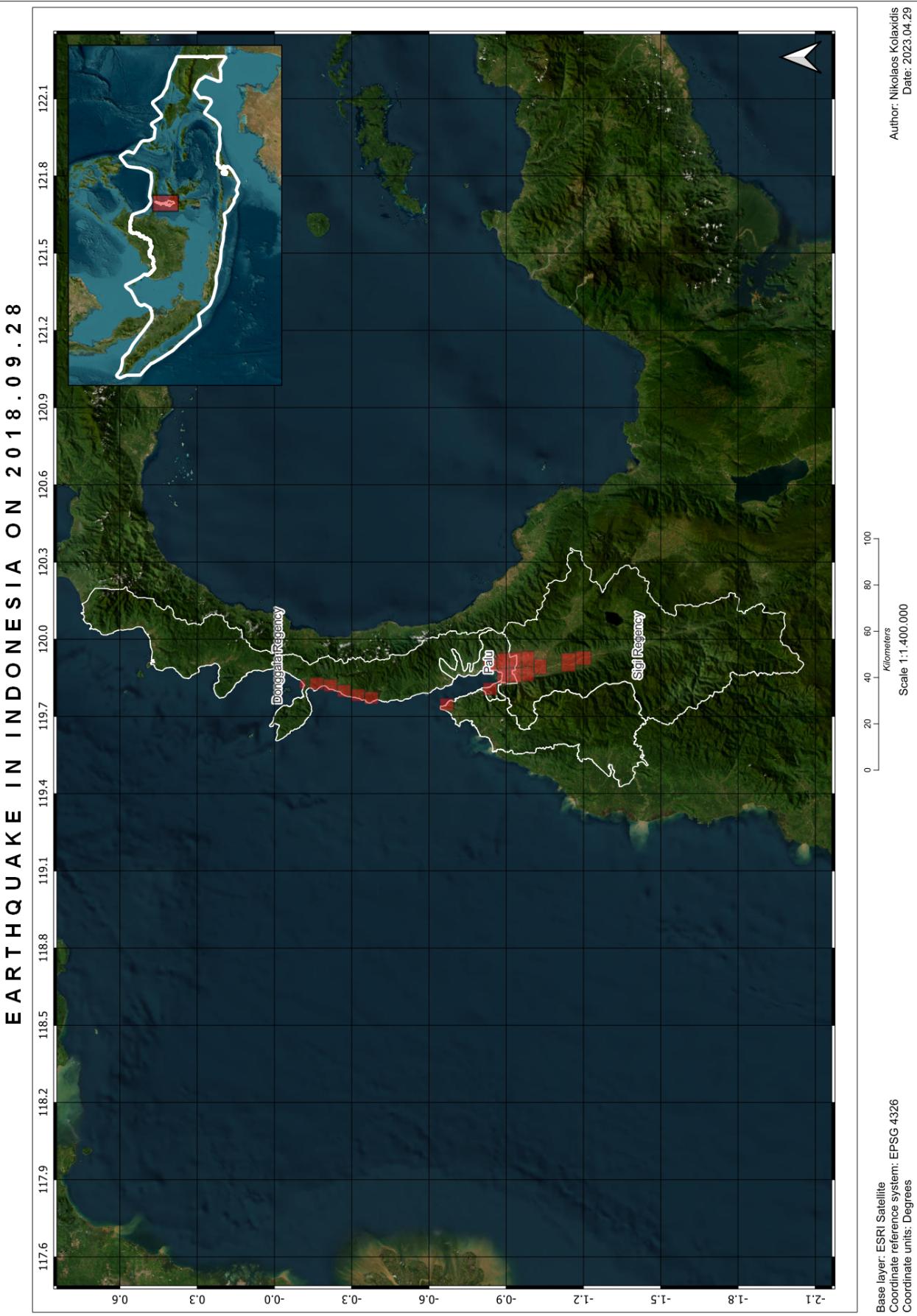


Fig. 9. Map of Indonesia with regions and AOIs during the earthquake [own figure]

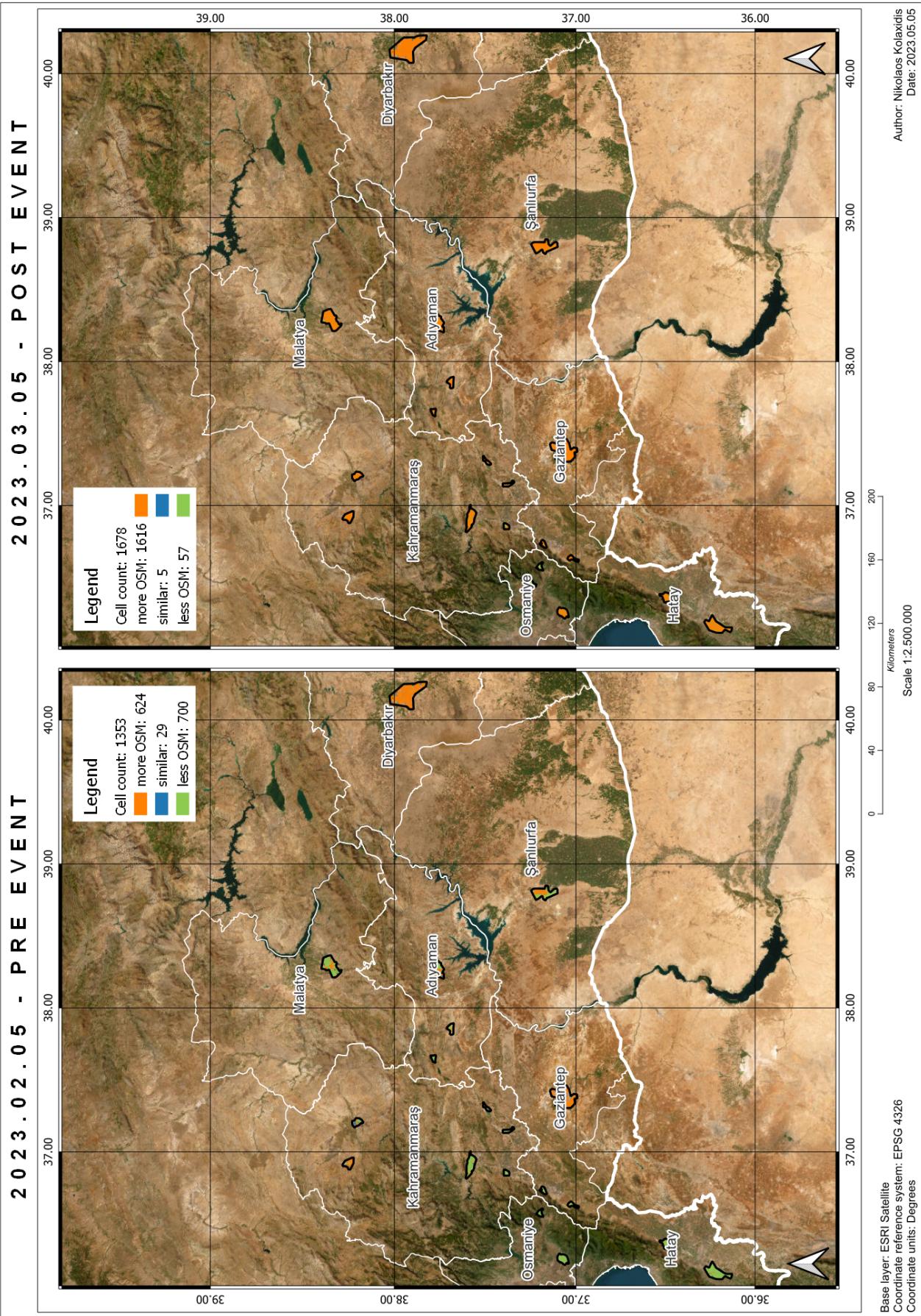


Fig. 10. Map of Türkiye with temporal evolution analysis per cell [own figure]

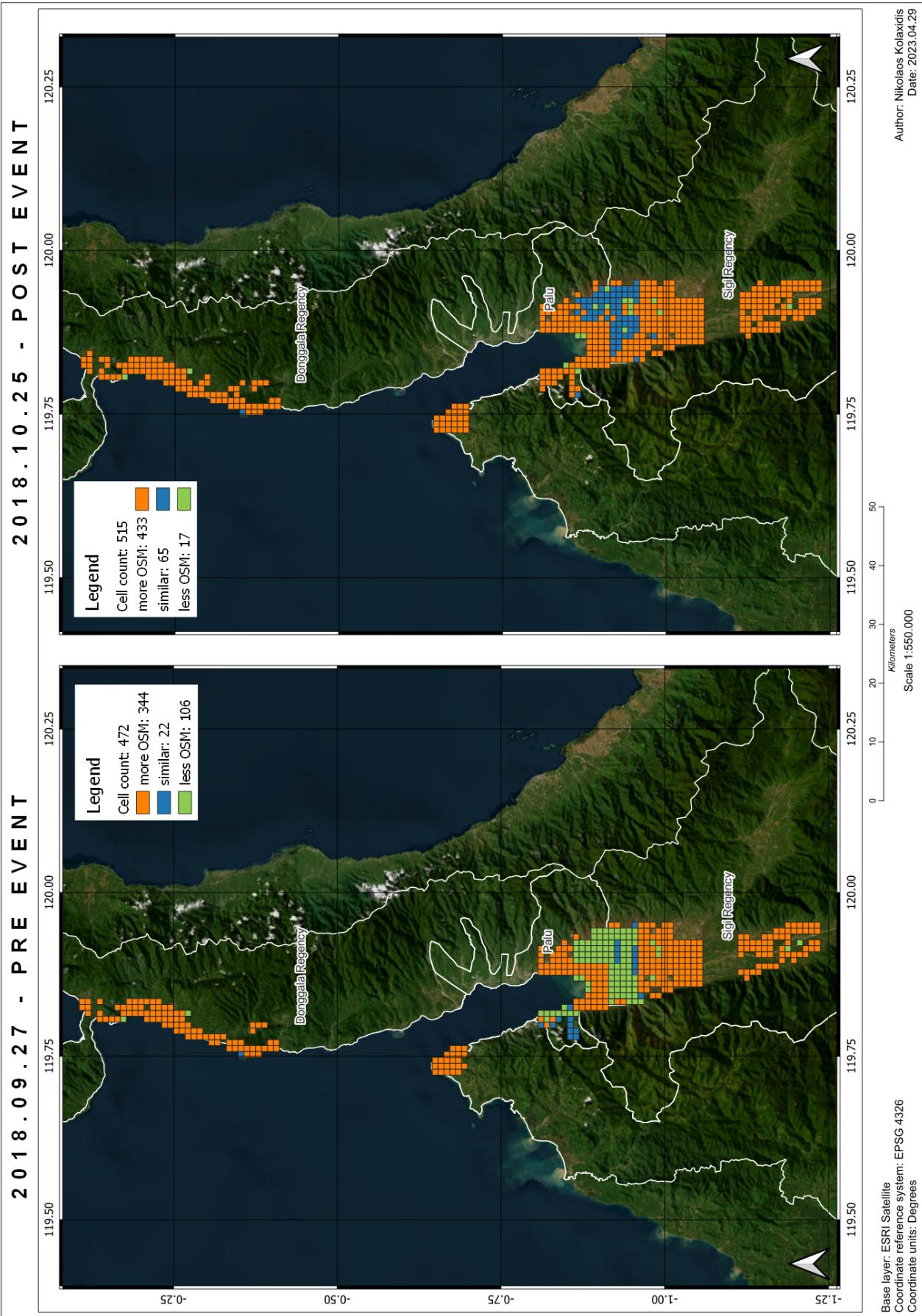


Fig. 11. Map of Indonesia with temporal evolution analysis per cell [own figure]