

CITIZEN-GENERATED GEODATA FOR NATURAL PARKS USE ANALYSIS: INSIGHTS FROM FACEBOOK IN THE INSUBRIA REGION

D. Oxoli^{1*}, M. A. Brovelli¹

¹ Dept. of Civil and Environmental Engineering, Politecnico di Milano, Como Campus, Via Anzani 42, 22100 Como, Italy - (daniele.oxoli, maria.brovelli)@polimi.it

Commission IV, WG IV/4

KEY WORDS: Natural Parks, Citizen-generated Geodata, Social Media, Space-Time Patterns

ABSTRACT:

Green areas such as natural parks provide citizens with a number of health and leisure benefits, often accessible with a few minutes of travel from urban centres. Moreover, the natural heritage enclosed in most green areas plays a pivotal role also in the economic integrity of these territories by driving local growth thanks to the establishment of tourism activities. In this context, the monitoring of both visitors and dwellers fluxes, as well as destination preferences, is key to provide land managers with critical information to shape local management and promotion strategies. This paper presents a preliminary investigation on the use of citizen-generated geodata -provided by Facebook- to empower the generation of space- and time-resolved insights into people fluxes in natural parks through a comparison with neighbouring urbanized areas. The Insubria region, a historical-geographical area between Northern Italy and Southern Switzerland, is considered a case study. Facebook users' population and movements data are analysed to identify trends and metrics on fluxes and support the estimation of the recreational and tourism value of natural parks. Results are presented as graphs and summary statistics and discussed according to their possible integration into territorial management and promotional practices.

1. INTRODUCTION

Green areas located within or nearby heavily urbanized territories provides countless benefits to communities including leisure, health and psychological supports to citizens who have access to them (Markevych et al., 2017). Furthermore, where green areas include natural parks or valuable natural heritage, local economic growth is likely to occur thanks to ventures connected to outdoor activities and tourism, which in turn is key for generating political support for green areas area (Di Minin et al., 2015). Nevertheless, promotion and management of wide green areas such as natural parks are often resource-demanding while remaining key to their conservation and marketing. Critical to most promotion and management operations is the availability of data enabling to assess human-environment relationships (Oleśniewicz et al., 2020), including e.g. visitors' fluxes, that are pivotal to the economic integrity of these areas but also a primary stress of the ecological heritage thereof. In view of the above, the consideration of novel data sources, such as citizen-generated and social media data, to hold up green areas monitoring and planning may provide a cost-effective mean to fill informational gaps on the actual people's interactions with green areas (Hausmann et al., 2018) by stimulating and supporting tailored promotion and management strategies (Wood et al., 2013).

To that end, this study preliminary investigates applications of citizen-generated geodata, namely records of users' population and movements provided by Facebook (Facebook, 2021), as a complementary data source for the monitoring of natural parks. The main purpose is to infer people fluxes (visitors and dwellers) through space and time by attempting also to derive metrics on natural parks recreational and tourism value, through a comparison with neighbouring areas. The target data source

provide the study with space- and time-resolved observations from which trends were inferred by means of exploratory statistical analyses.

The study focuses on the Insubria region (see Figure 1), a historical-geographical area that stretches between Northern Italy and Southern Switzerland, where the potential of tourism in the green areas is not fully exploited. This because of the fragmented political and management context of the area that extends across different local and national jurisdictions. The study is developed in the framework of the INSUBRIPARKS project, funded by the Interreg program of the European Union, which aims at increasing tourism attractiveness of the Insubria region through the provision of physical infrastructure as well as integrated marketing and management strategies for the Insubria natural parks (Oxoli et al., 2019).

The paper continues as follow. Section 2 introduces the data sources and datasets considered in the study. Data processing is described in Section 3. Section 4 includes overviews and discussion of the achieved results whereas conclusions and future directions of the work are reported in Section 5.

2. CITIZEN-GENERATED GEODATA: THE FACEBOOK DATA FOR GOOD INITIATIVE

The analysis of population and mobility trends across the Insubria region was performed by exploiting aggregated users' population and movements data provided by the Facebook Data for Good initiative (Facebook, 2021). The main purpose of this initiative is to empower researchers and policymakers with near-real-time and georeferenced information on a global scale to address, primarily, humanitarian issues such as disaster response and disease prevention. Data are derived by the users'

* Corresponding author

generated records stream into the Facebook social media platform that includes users' location and conversation topics. Privacy protection is ensured for the distributed data by spatial aggregation of individuals records. The initiative started in 2018 while a significant increase in its users' community has been registered due to the COVID-19 pandemic, for which specific datasets have been developed and released. Beyond emergencies response, these data are promising also for a wider range of applications connected to human-environment interactions by supplying space- and time-resolved information representative of a large community, such as the one of Facebook that today counts 2.8 billion active users worldwide.

Focusing on the proposed study, two among the available Facebook Data for Good dataset were considered and the data collection on the Insubria region was activated upon a specific request of the authors to the data provider. The first dataset is called *Facebook Population Maps* and consists of aggregated information of people using the Facebook app on their mobile devices (with location history turned on) and provides counts of users that stay within a specific location in a defined time interval. The second is called *Movements Maps* and, under the same assumption of the Population Maps, provides counts of users' movements between locations in a defined time interval (Heo et al., 2020). Aggregated records of both population and movements are available with a time resolution of eight hours. The raw data are provided in CSV format.

Data locations refer to pixels of geographic grids derived from the Bing tile map system (Maas et al., 2019) in the WGS84 reference system. The spatial grid resolution for the Insubria region was approximately 0.8 km for the population data and 1.7 km for movements data (see Figure 1). The different spatial resolution is due to the computational time required by the Facebook servers to generate the two datasets within the defined time unit of eight hours. Movements counts imply pairwise analysis between locations (i.e. between origin and destination pixels) while population counts are performed looking at a single location. Inevitably, the former operation is more complex than the latter thus resulting in a lower resolution for movements data. In fact, the servers automatically increase the spatial unit size to conclude the operation on time. The available resolution also depends on the extent of the study region for which a collection is activated. This means that the larger the study region, the lower is the resolution of the datasets for the same tradeoff explained above.

The period May 6th - December 1st 2020 was considered for this study. The analysis dataset was composed of two collections of 627 CSV files (one every eight hours) for both the users' population and movements (see figures 2 and 3). Users' movements for a single pixel (in every eight-hour time slot) were computed as the sum of movement records having as destination the pixel, minus the records having as origin the pixel but a different destination. In this way, a single value for each pixel was considered by producing a dataset comparable to the users' population one.

Additional spatial layers were used to perform the analysis. In particular, the 2018 CORINE Land Cover (CLC) raster grid at 100 m resolution (Büttner et al., 2017) was considered as a reference to assess users' population and movements trends by principal land cover classes. A vector layer of the Insubria natural parks was included to assess trends within these areas and compare them with land cover classes trends (see Figure 1).

Finally, the 2015 Global Human Settlement Population (GHS-POP) raster grid at 250 m resolution (Schiavina et al., 2019) was used to estimate the resident population in the study area. These estimates were used to approximately deduce the representativeness of the average Facebook users' population with respect to the actual one (see Table 1).

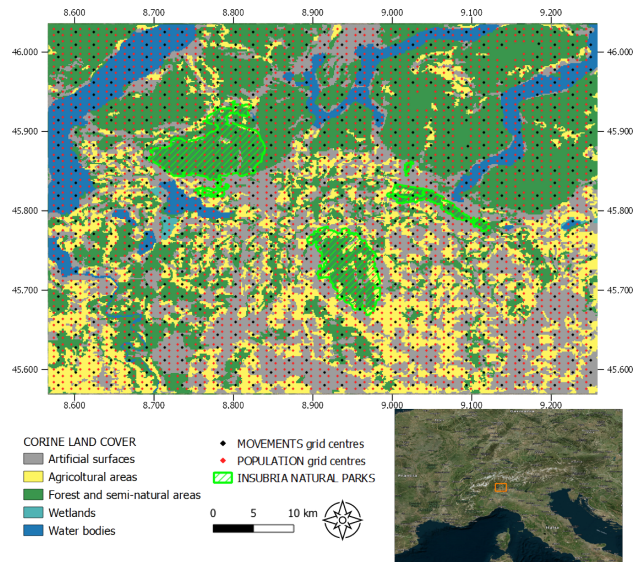


Figure 1. The Insubria region (Coordinate Reference System: WGS84) with the CLC classes and natural park areas highlighted. Black and red spots represent centroids of the Facebook data grids respectively for the users' movements and population datasets. (Basemap: Bing Satellite © Microsoft).

3. DATA PROCESSING

The data processing focused on Facebook data aggregation both on time (i.e. averages by days of the week) and space, by setting apart observations according to the reference CLC class or natural park areas in which they were recorded (see figures 4 and 5). Additionally, computation of relative changes between user's population and movements during weekdays and weekends was included in the analysis to explicitly investigate the use of natural parks as a destination for leisure activities whereby comparisons with e.g neighbouring urbanized or agricultural areas (see Table 2).

Time aggregations were performed using Python data analysis libraries, such as Dask (<https://dask.org>), Pandas (<https://pandas.pydata.org>), and GeoPandas (<https://geopandas.org>). Spatial aggregation of records was achieved by assign to each Facebook data grid pixel the corresponded CLC class (or its belonging to natural park areas) through overlay operations with the CLC raster grid and the vector layer of natural parks using QGIS (<https://qgis.org>). Only CLC level-one classes were considered (Büttner et al., 2017) which are artificial surfaces, agricultural areas, forests and semi-natural areas, wetlands, and water bodies. Wetlands class was later excluded from the analysis because it covers a negligible portion of the study region (see Figure 1).

Finally, the size in percentage of the Facebook users' population with respect to the actual resident population was computed exploiting the GHS-POP raster grid that supplied an estimation of the people count in the study region (Schiavina et al.,

2019). This information depicts the approximate representativeness of the Facebook users' population sample that is functional to the discussion of the analysis results. The choice of using this particular dataset format instead of traditional census data is supported by two reasons. The main reason is that the boundaries of the study region (as well as the ones of the CLC classes and natural parks patches) do not overlap with any existing census tracks, which is the reference area unit adopted by the national statistic bureaus to distribute census statistics. The second reason is that the Insubria region is located across the international border between Italy and Switzerland so affecting the availability of consistent census data covering the whole area.

The information on representativeness of the Facebook users' population sample, computed separately for each CLC class and for natural park areas is included in Table 1. Results of the analysis are presented and discussed in the following section.

Artificial surfaces	3,3%
Agricultural areas	9,0%
Forest and semi-natural areas	12,1%
Water bodies	14,6%
Natural parks	15,4%
Whole area	4,4%

Table 1. Representativeness of the average Facebook users' population, expressed as a percentage of the estimated resident population in the Insubria region per CLC classes and natural park areas. The resident population is estimated using the 2015 GHS-POP raster grid at 250 m resolution (Schiavina et al., 2019).

4. RESULTS AND DISCUSSIONS

Results include time-series graphs and summary statistics of users' population and mobility trends across the Insubria region. Experiments were run by grouping movements timeseries according to the pixels underlying CLC classes or natural park areas, as well as recording time such as days of the week. Trends were investigated in terms of patterns and percentage changes to provide insightful comparisons between diverse extension and population abundances characterizing Insubria region CLC classes and natural park areas.

Full time-series of the two considered variables are reported in figures 2 and 3. Summer holidays effect, between July and August, both on users' population and movements is clearly visible as a drop in the time-series graphs. Only water bodies areas showed a marked opposite trend which denotes the attractiveness of this peculiar territorial feature especially for summertime recreational and tourism activities.

A widespread co-location of the largest users' movements records within all CLC classes and natural park areas was observed during weekends, as shown in Figure 4. Namely, average users' movements during weekends (Saturday and Sunday) increase above 10% than the weekdays' baseline (Monday to Friday) within all CLC classes and natural park areas, as reported in Table 2. These increases are higher in water bodies and natural park areas providing insight into their appeal than e.g. artificial and agricultural areas for recreational weekends activities. By considering weekdays and weekends for users' population records, trends are instead different between CLC classes and natural park areas, as shown in Figure 5. The sharp decreases

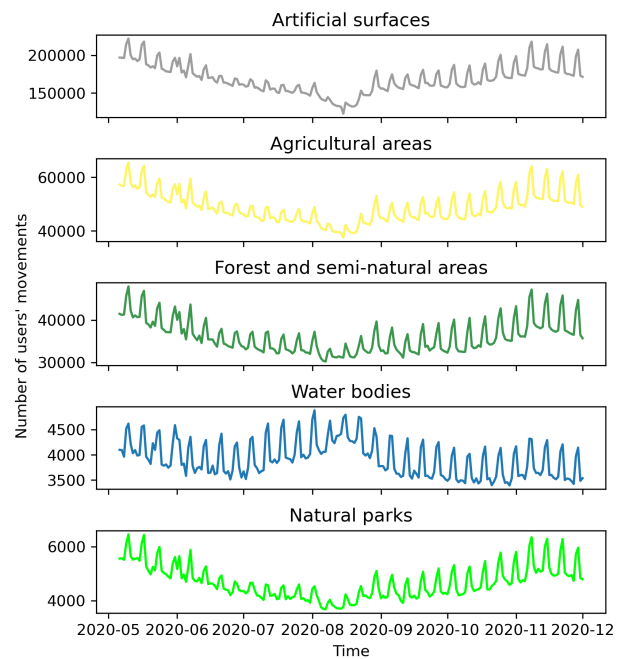


Figure 2. Time series of total users' movements observed in each CLC class and natural park areas of the Insubria region. Records are provided every 8 hours.

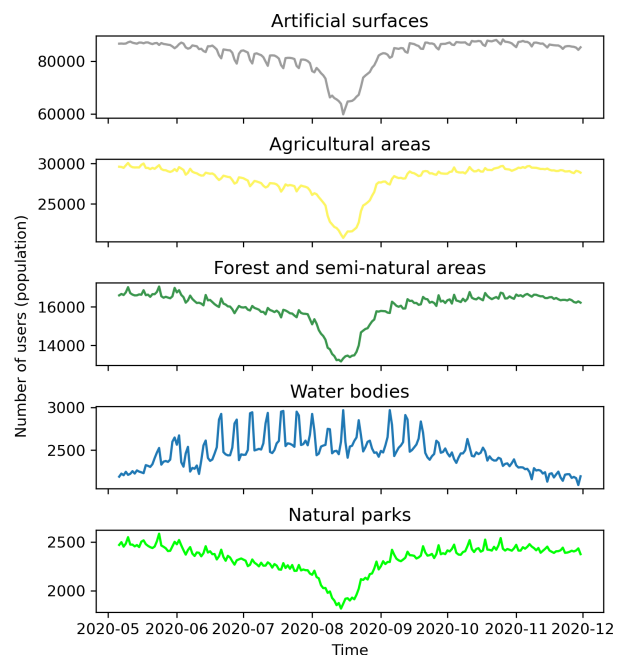


Figure 3. Time series of average users' population observed in each CLC class and natural park areas of the Insubria region. Records are provided every 8 hours.

registered in the artificial surfaces CLC class and (mostly during the transition between weekdays and weekend) in agricultural and forest areas, outline outflows of people from those places. As for the movements, the users' population showed an opposite trend in water bodies but also in natural park areas. This provides evidence of the recreational and tourism value also for natural parks which was less apparent observing users'

movements only.

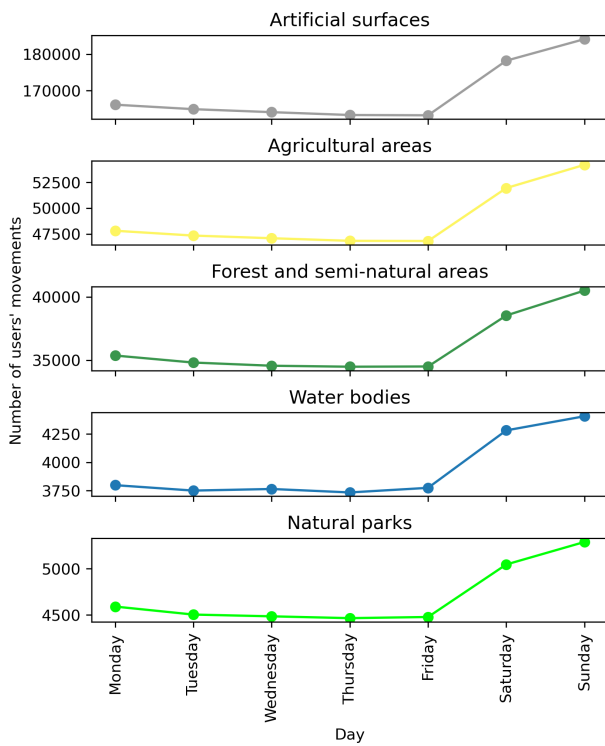


Figure 4. Average users' movements by day of the week in each CLC class and natural park areas of the Insubria region.

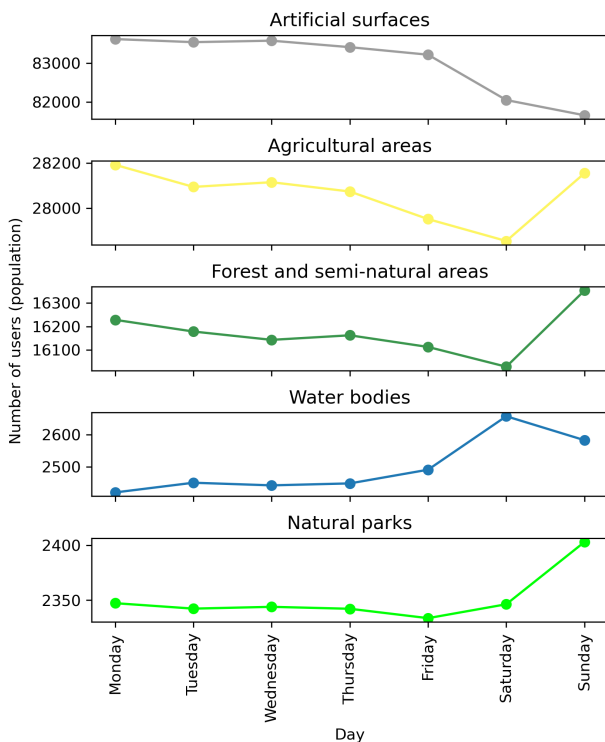


Figure 5. Average users' population by day of the week in each CLC class and natural park areas of the Insubria region.

In principle, quantitative measures of people fluxes could be derived by the analysis of actual values of the time-series, coupled with the computed representativeness of the Facebook users'

	Relative changes between weekdays and weekends	
	Population	Movements
Artificial surfaces	-1,9%	10,2%
Agricultural areas	-0,3%	12,4%
Forest and semi-natural areas	0,2%	12,9%
Water bodies	6,9%	13,8%
Natural parks	1,4%	13,9%
Whole area	-1,16%	11,01%

Table 2. Relative changes in percentage of average users' population and movements in weekends with respect to weekdays.

population sample (Oleśniewicz et al., 2020). However, the aim of this preliminary study was to examine the suitability of citizen-generated geodata, such as the Facebook users' population and movements records, as a complementary asset to territorial management, by searching for proof of their utility in the characterization of space-time people's fluxes. Results provided a substantial positive answer to the underlying research question.

5. CONCLUSIONS

In this paper, the use of citizen-generated geodata for the analysis of space-time trends of people fluxes in the Insubria region was investigated. In particular, Facebook users' population and movements data were exploited to infer underlying differences in the use of natural park areas with respect to neighbouring urbanized areas and other landscapes within the study region.

Preliminary results outlined that the considered citizen-generated geodata are informative for space-time trends characterization and capable of supply metrics of people fluxes across territories that are difficult to retrieve using traditional monitoring tools, such as surveys (Hausmann et al., 2018). The representativeness of the adopted data remains questionable due to the size of the Facebook users' sample which represents only a small portion of the resident population. Furthermore, tracking of individuals is not meant by the Facebook data thus making it hard to discern between Insubria region dwellers and external visitor fluxes. However, the availability of information at high space and time resolutions on destination preferences, as well as seasonal/weekly people fluxes variations, may open new rooms toward effective and evidence-based local territorial management actions.

It is worth noticing that the study region was affected by COVID-19 mobility restrictions mainly during the fall season between mid of October and December. However, milder restrictions to both national and international travel were ongoing also during the spring and summer periods. Hence, the data on users' population and mobility might be biased if compared with the pre-pandemic scenario. Investigation on differences with the pre-pandemic period was not carried out in this work due to lack of data, while it may be considered for future developments of the study.

Finally, the development of user-friendly data management tools for data collection, storage and processing remain key to allow also not expert data consumers to take advantage of these modern data resources. To that end, future directions of this work will focus also on testing alternative open-source data tooling such as the Open Data Cube (<https://www.>

opendatacube.org) to ease recursive data operations, including space-time aggregations, which were largely exploited in this preliminary study.

ACKNOWLEDGEMENTS

This research has been funded by the project “INSUBRI-PARKS” within the Interreg Co-operation Programme 2014 - 2020 (ID 605472). The paper is based on the datasets made available by Facebook through the Facebook Data for Good initiative (<https://dataforgood.fb.com>).

REFERENCES

- Büttner, G., Kosztra, B., Soukup, T., Sousa, A., Langanke, T., 2017. CLC2018 technical guidelines. *European Environment Agency: Copenhagen, Denmark*, 25.
- Di Minin, E., Tenkanen, H., Toivonen, T., 2015. Prospects and challenges for social media data in conservation science. *Front. Environ. Sci.*, 3, 63.
- Facebook, 2021. Facebook data for good. <https://dataforgood.fb.com>. Accessed 29.01.2021.
- Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., Di Minin, E., 2018. Social media data can be used to understand tourists’ preferences for nature-based experiences in protected areas. *Conserv. Lett.*, 11(1), e12343.
- Heo, S., Lim, C. C., Bell, M. L., 2020. Relationships between Local Green Space and Human Mobility Patterns during COVID-19 for Maryland and California, USA. *Sustainability*, 12(22), 9401.
- Maas, P., Iyer, S., Gros, A., Park, W., McGorman, L., Nayak, C., Dow, P. A., 2019. Facebook disaster maps: Aggregate insights for crisis response & recovery. *KDD*, 19, 3173.
- Markevych, I., Schoierer, J., Hartig, T., Chudnovsky, A., Hystad, P., Dzhambov, A. M., De Vries, S., Triguero-Mas, M., Brauer, M., Nieuwenhuijsen, M. J. et al., 2017. Exploring pathways linking greenspace to health: Theoretical and methodological guidance. *Environ. Res.*, 158, 301–317.
- Oleśniewicz, P., Pytel, S., Markiewicz-Patkowska, J., Szromek, A. R., Jandová, S., 2020. A model of the sustainable management of the natural environment in national parks—A case study of national parks in Poland. *Sustainability*, 12(7), 2704.
- Oxoli, D., Cannata, M., Terza, V., Brovelli, M., 2019. Natural Heritage Management and Promotion Through Free and Open Source Software: a Preliminary System Design for the Insubri-parks Project. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, 4214, 179–183.
- Schiavina, M., Freire, S., MacManus, K., 2019. GHS population grid multitemporal (1975, 1990, 2000, 2015) R2019A. *Eur. Comm. JRC*.
- Wood, S. A., Guerry, A. D., Silver, J. M., Lacayo, M., 2013. Using social media to quantify nature-based tourism and recreation. *Sci. Rep.*, 3(1), 1–7.