Turkish Journal of Earth Sciences

http://journals.tubitak.gov.tr/earth/

Research Article

Turkish J Earth Sci
(2023) 32: 1041-1051
© TÜBİTAK
doi:10.55730/1300-0985.1892

Exploring crowdsourcing accountability for mapping Antarctica: a case study using 5 years of social media data

Ayşe Giz GÜLNERMAN*
Land Registry and Cadastre Department, Ankara Haci Bayram Veli University, Ankara, Turkiye

Received: 22.05.2023 • **Accepted/Published Online:** 01.11.2023 • **Final Version:** 28.11.2023

Abstract: The continent of Antarctica is one of the most challenging regions in terms of generating and updating geodata. Extensive research has been conducted on geodata acquisition, primarily focusing on earth observation satellites and local surveys in polar regions. Earth observation satellites offer limited spatial, temporal, and semantic data, while local surveys are constrained by polar regions' size and challenging conditions. This article delves into the crowdsourced data, aiming to significantly enhance geodata contribution in polar regions beyond current methods. To our knowledge, this study is the first of its kind to address social media data, a form of crowdsourcing, specifically for the Antarctic continent. Encompassing 5 years of social media data collection, the study uniquely presents original insights by investigating data reliability based on user activity levels and user movement consistency. The primary outcome of the study is that the activity level of users negatively correlates with spatial behavior consistency. This indicates that dominant user influence has led to inconsistent content manipulation. However, while the overall rate summary indicates a high inconsistency ratio for the active group, there still exists consistent behavior within these groups. A tight method to discriminate these reliable data generators with consistent behavior should be aimed as proposed in this study to prevent valuable data loss. This study contributes to reliable data scrutiny techniques from social media data in a general sense while providing a glimpse into the spatial quality of data generated specifically for Antarctica. This glimpse will enable future assessments of data collected in Antarctica for reliability checks and offer benefits in terms of processing workload and result accuracy in preprocessing steps for text, image, and spatial-based data processing.

Key words: Polar monitoring, crowdsourcing, new forms of geodata, geo-social media

1. Introduction

Antarctica stands as one of the most challenging areas for earth scientists to gather data. Due to its harsh climate, limited transportation, and uninhabited geography, current mapping technologies are restricted to remote sensing (Baumhoer et al., 2018; Yirmibesoglu et al., 2022; Gulher and Alganci, 2023). Additionally, during the Antarctic summer season, measurements can be taken in limited areas using methods like aerial photogrammetry and field survey techniques (Isler et al., 2021; Pina and Vieira, 2022; Selbesoglu et al., 2023). However, these methods, which require ground truth validation and/or enable mapping in constrained areas, are deficient in terms of data integrity, spatial accuracy, and semantic information. Hence, there is a need for alternative mapping methods that can contribute to Antarctic geographic information (Dong et al., 2022).

This study addresses crowdsourcing as an alternative data source that could contribute to Antarctic geographic information. Crowdsourced data platforms offer significant contributions in situations of data scarcity for mapping activities, whether in updating existing data or generating new data (Ghermandi and Sinclair, 2019; Brovelli et al., 2020). As a relatively new mapping technique, crowdsourcing primarily relies on various digital platforms integrated with global navigation satellite systems (GNSS), enabling nonexperts in mapping to provide data as volunteers (Goodchild, 2007; Singleton et al., 2017).

Geo-crowdsourcing is commonly categorized into three classes: social media, citizen science, and peer-production (See et al., 2016; Ballatore and De Sabbata, 2018; Gulnerman et al., 2021). These classes are based on the distinctions among volunteers (conscious, unconscious) and platform designs (social network-based, blog-based, map-based). In this study, we focus on social media crowdsourcing. Similar to all crowdsourced data, social media data (SMD) are primarily used to extract purposeful and meaningful information. However, social media data naturally tends to possess characteristics of being mostly unstructured, unreliable, and uncertain (Ballatore and Zipf, 2015; Basiri et al., 2019; Gulnerman et al., 2020).

In regions abundant with data through social media, density-based spatial information inferences can be con-

* Correspondence: ayse.gulnerman@hbv.edu.tr

ducted with high accuracy. However, these methods face a challenge in areas where data scarcity restricts contributions to a limited number of users (volunteers), such as Antarctica, where data-dominant users can pose a problem. Removing the data from dominant users is employed as a solution in social media research. In such cases, a portion of valuable data from regions with limited data could potentially be lost through this approach. Therefore, this study uniquely focuses on examining the spatial reliability of social media data related to the Antarctica region by considering user activity and spatial footprints. The objective is to assess spatial data consistency, contribute to user-based reliability calculations, and enhance the data extraction technique within preprocessing to maximize the evaluation of social media data.

Monitoring a region using social media data is more commonly encountered in urban areas, where it is largely based on semantic information. In urban settings, applications for inferring information about mobility, land use, and public health are prevalent (Jurdak et al., 2015; Gao et al., 2017; Gulnerman, 2021). In contrast, the focus shifts in regions like Antarctica. This article represents the initial work of our project focused on geo-monitoring the Antarctica region using social media data. The project overarchingly aims to generate, verify, and update topographic data in the Antarctica region, as well as track glacier changes and monitor the movements of animals. To our knowledge, the idea of geo-monitoring polar regions with social media platforms has not yet been researched. There are only four studies related to social media data for the Antarctic region. The one recommending social media platforms to promote their projects in polar regions (LaRue et al., 2020). Another one focuses on the tourists' perception retrieval from a social media platform (TripAdvisor) for "Antarctic Peninsula" (Frame et al., 2022). Similarly, the last one utilizes Chinese social media platforms ("Zhihu" and "Mafengwo") to indicate the environmental impacts of tourists over Antarctica (He and Liu, 2023).

Antarctica, the southernmost region of the world, lacks permanent human habitation. Nonetheless, each year, thousands of individuals, including tourists, researchers, and workers, venture to the polar regions (IAATO, 2022) and utilize crowdsourcing platforms to share their experiences and observations. The social media data examined in this study consist of posts from tourists, researchers, workers, and seasonal inhabitants who visit the Antarctica continent each year. Alternatively, these social media data can also comprise posts from bots or individuals manipulating content. This experimental study investigates this manipulation through various techniques using a 5-year Twitter dataset to filter unreliable datasets to contribute to overarching goals in future studies.

The contents of the subsequent sections are as follows: The second section discusses the dataset used, how it was acquired, filtered, and the methodology for measuring data consistency. The third section presents details of the data and analysis results. In the fourth section, the obtained results are discussed further. The fifth section presents the study's contributions to future research, and its limitations.

2. Materials and methods

The methodology employed in this study encompasses four main steps. Firstly, data is acquired from a social media platform, serving as the initial data collection process. Subsequently, the acquired data is filtered based on predefined spatial boundaries, narrowing down the dataset to the desired geographic area. Thirdly, the activity level of users is determined by considering tweet counts within the Antarctic region and the global area, and the spatial distribution of data based on data generators' (users') activity levels are visualized. Finally, the spatial consistency of inbound and outbound tweets is assessed to determine the coherence of the data, to identify any potential presence of manipulated content in terms of activity levels.

2.1. Acquiring data from an SMD

Various methods exist for retrieving SMD, including stream harvesting via an application programming interface (API), web scraping, and downloading from data pools. The Social Media Lab at Ryerson University has compiled a curated list of tools designed for retrieving and manipulating SMD[1]. These tools typically come in the form of web-based or desktop programs, and some may offer limited access to the data for free.

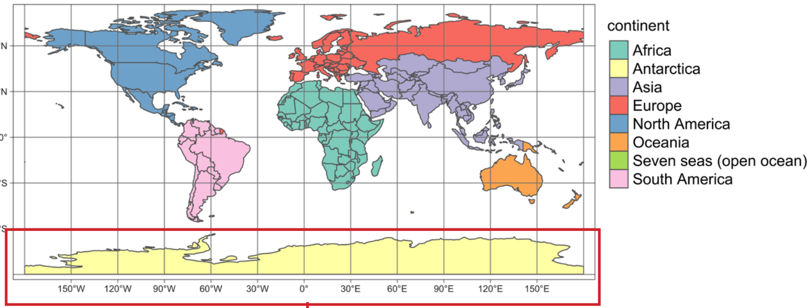
For the purpose of this study, we utilized the Geo Tweets Downloader program[2], which employs stream harvesting to retrieve Twitter data. It is important to note that using this program requires Twitter API credentials, which were previously available for academic use at no cost until February 2023. However, as of that date, the use of all API types has become subject to a fee.

2.2. Filtering data by spatial boundaries

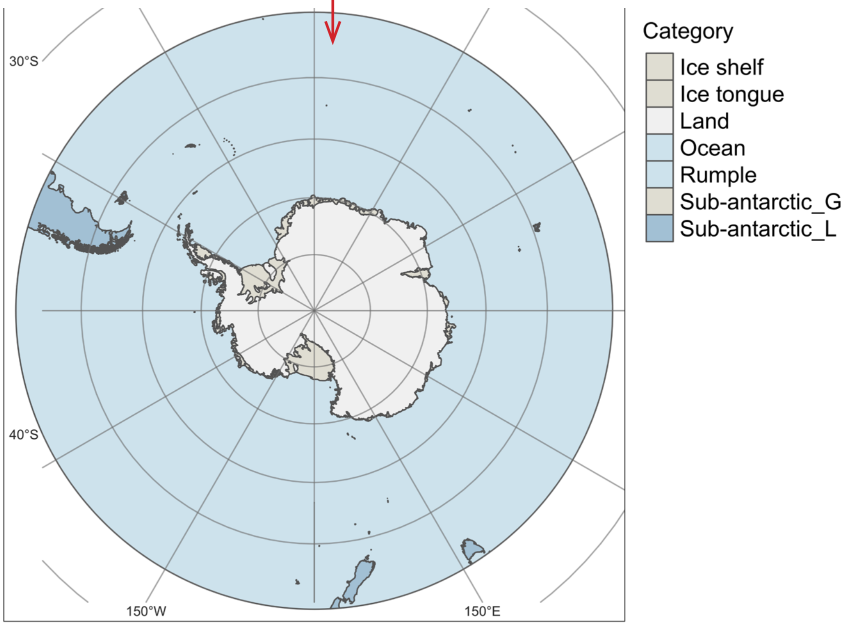
The data collection process involved using the Geo Tweets Downloader tool, which allows for the retrieval of all geotagged data obtained from the Twitter API. To narrow down the dataset, a bounding box (bbox) filter was applied. Specifically, the data retrieval bbox was defined as –180.00, –90.00 for the bottom left coordinates and 180.00, 90.00 for the top right coordinates (Figure 1a). As a result, tweets spanning the entire global region be-

[1]Toronto Metropolitan University. Social Media Research Toolkit [online]. Website https://socialmedialab.ca/apps/social-media-research-toolkit-2/5 [accessed 01 May 2023].

[2]Geo-tweet-downloader [online]. Website https://github.com/nagellette/geo-tweet-downloader [accessed 01 May 2023].



(a) World (EPSG:4326)



(b) Antarctica (EPSG:3031)

Figure 1. Study area.

tween 2016 and 2021 were obtained. To focus specifically on content related to Antarctica, spatial data filtering was applied. For this purpose, the bbox for Antarctica (Figure 1 (b)) was defined as –180.00, –90.00 for the bottom left coordinates and 180.00, –63.27 for the top right coordinates[3].

2.3. User activity level assignment and spatial data distribution

In the exploratory analysis of the filtered data based on the Antarctica bbox, the contribution rates of users (data producers) are determined, and their spatial footprints are visualized accordingly. In the first step, the number of tweets within the filtered data is calculated for each user.

The tweet counts data are highly skewed, therefore after conducting several empirical tests and relying on intuition, we defined the activity level intervals within Antarctica bbox (inbound) manually. The spatial distributions of data produced by users, based on their inbound activity levels, have been visualized for the Antarctic continent. For visualization, the coordinate reference system (CRS) was defined as WGS 84/Antarctic Polar Stereographic (EPSG:3031). This CRS was chosen because it provides a complete representation of the spatial relationships within the Antarctic continent. Displaying the spatial distribution of data in terms of inbound activity levels provides a visual opportunity to interpret the spatial data generation tendencies of the activity levels.

In the second step, a second dataset is prepared for reviewing the global activity of the users who tweeted at least once within the Antarctic bbox. This dataset is retrieved from the harvested dataset mentioned in section 2.1. which is significant for understanding whether the varying inbound activity levels of users hold true on the global scale (outbound).

[3]Country-bounding-boxes [online]. Website https://gist.github.com/graydon/11198540 [accessed 01 May 2023].

Thus, it allows the investigation of activities' susceptibility to regional, global, or different semantic characteristics. It examines global activity levels under each of the inbound activity levels. For each inbound activity level group, outbound activity levels were determined using the Jenks Natural Breaks Classification algorithm (Jenks, 1967).

This method employs the following two iterative steps for class determination:

1. Calculation of class variances: The mean of values within each class is computed, and the squared deviations of each value from this mean are calculated. These calculated values are summed for each class. The average of the summed values for each class is utilized as an indicator of the relationship between classes. The variance of each class inversely provides information about the proximity of values within the class. The magnitude of the average of these values indicates the extent of inconsistency among general class intervals.
2. Determination of new class intervals: Using a shifting technique, the new class interval is determined based on the previous class interval. All possible combinations of class intervals are established in the first step of the calculation. This two-step process continues until the minimum average variance value is found. In this study, we employ the R language and utilize the developed function for Jenks Natural Break Classification[4].

The data produced by users classified in terms of outbound activity levels was visualized to interpret spatial activity. WGS84 - World Geodetic System 1984 (EPSG:4326) was used as the CRS for this visualization. This CRS presents the world as a whole in the most widely used manner as in The Global Positioning System (GPS). Thus, the CRS enables the easy visual interpretation of user groups' spatial footprints.

2.4. Spatial consistency of inbound and outbound tweets

Each inbound and outbound activity level group is further analyzed by studying their footprints through distribution maps and the pace of tweeting activity between successive tweets. Furthermore, users' inbound and outbound tweeting activity is examined to assess spatial consistency. The data reliability of users can be investigated by considering the spatiotemporal pairs between consecutively generated data points. Accordingly, a consistency level is calculated for each user in three steps.

1. Tweets belonging to each user are ordered by timestamp.
2. The time and location differences between consecutive tweets are calculated, and the speed is determined (kilometer/hour).
3. If the speed level in each leg is greater than 1000 km/hour, it is considered inconsistent; if it is smaller, the spatial movement is considered consistent.

A user cannot travel at a speed of 1000 km/hour; this value is roughly determined with consideration to the speed of the fastest mode of transportation, which is an airplane. Additionally, within the seemingly consistent movement of legs, a spatial distance of 300 m is considered stationary. This takes into account the potential deviation in location accuracy of systems like GPS up to this level. This approach allows for a detailed investigation of users' spatial movements, calculating spatial movement consistency rates based on activity levels.

3. Results

The dataset retrieved for this study within the Antarctica bbox comprises 150,829 tweets contributed by 3873 users. During the data inspection, three primary criteria, namely the tweet count per user, spatial footprints, and the tweeting velocity, were considered to gain insights into the tweeting activity surrounding Antarctica. Analysis reveals that the account with the highest number of tweets has contributed 91,750 tweets, accounting for approximately 60% of the entire collected dataset. Conversely, 2970 users have only posted a single tweet each. Due to the highly right-skewed nature of the grouped data based on user tweet count, conventional classification algorithms like k-means, x-means, and kernel density estimation did not yield satisfactory results when applied to this one-dimensional data. Consequently, after conducting several empirical tests and relying on intuition, we defined the intervals for inbound activity levels manually.

The first activity level dominates both the quantity and spatial distribution of the data, as depicted in Figure 2a. In contrast, the second level consists of 13 users, representing approximately 30% of the entire dataset, but does not exhibit any discernible spatial pattern (Figure 2b). The third level, assigned to 18 users, accounts for around 5% of the total data and displays some patterns, although a more detailed examination is required to ascertain its characteristics with certainty (Figure 2c). The remaining three levels encompass the majority of users but contribute less than 5% of the overall data. Interestingly, the spatial plots of these levels (Figures 2d, 2e, 2f) reveal similarities in certain areas, indicating that most of the content from these levels is likely generated by nonbot users.

To gain further insights into users' activity, their spatial activity beyond Antarctica is examined, and the total number of tweets outside Antarctica for each user is listed in Table 1. Analysis of Table 1 reveals that the number of tweets outside Antarctica gradually increases for activity

[4]'PlotJenks': R function for plotting univariate classification using Jenks' natural break method [online]. Website http://cainarchaeology.weebly.com/uploads/1/4/4/7/14477112/plotjenks.r [accessed 01 May 2023].

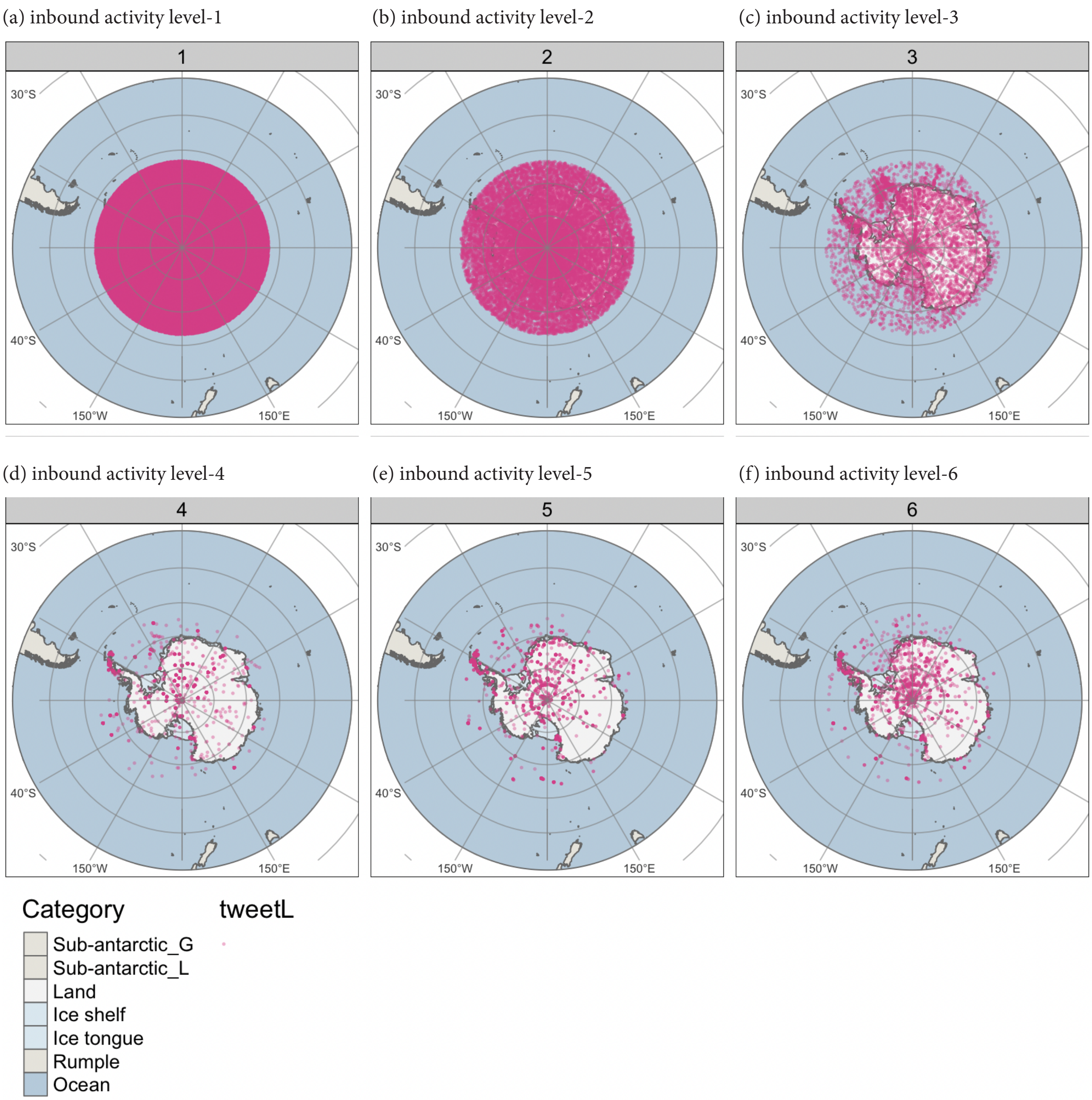


Figure 2. Distribution of tweets in Antarctica bbox in terms of inbound activity level.

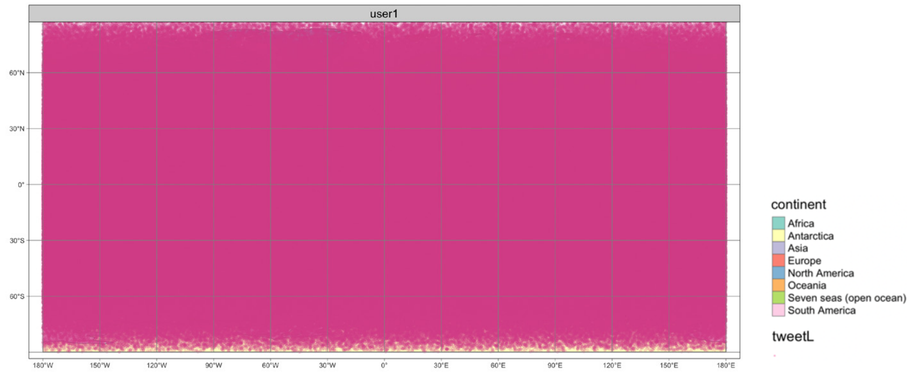
level 5. However, when considering the average number of tweets per user, the activity level classification remains consistent with the previously determined positions. Nonetheless, not all users adhere strictly to their respective activity level ranges. Therefore, a detailed examination of users' inbound activity within Antarctica and outbound is warranted to understand the spatial patterns better.

Based on empirical tests and intuitive analysis, we retrieved overall tweets from our database for each level determined earlier. This dataset includes all tweets from our dataset that cover the global area and belong to users who have tweeted at least once from within the Antarctica bbox. Figure 3 illustrates the global distribution of tweets in terms of inbound and outbound activity levels. In addition to the predefined classes for Antarctic region tweets, we found it necessary to introduce a second classification for the 4th, 5th, and 6th activity levels. For the first three activity classes, where the user count allows for plotting, we directly present the digital footprints of these users in Figure 3 (a), (b), and (c). However, for the higher activity

Table 1. Social media users' data contribution in terms of inbound activity level.

Inbound activity level	Ranges (# of Tweets)	# of users	# of inbound tweets	% of total user count	% of total tweet count	# of outbound Tweets
1	= 91,750	1	91,750	0.02	60.83	1,618,214
2	> 1033 and < 10,304	13	44,048	0.33	29.20	663,638
3	> 101 and < 968	18	7437	0.47	4.93	77,202
4	> 9 and < 102	94	2315	2.43	1.54	293,970
5	> 1 and < 10	777	2309	20.06	1.53	1,389,336
6	= 1	2970	2970	76.69	1.97	736,142

(a) inbound activity level-1



(b) inbound activity level-2

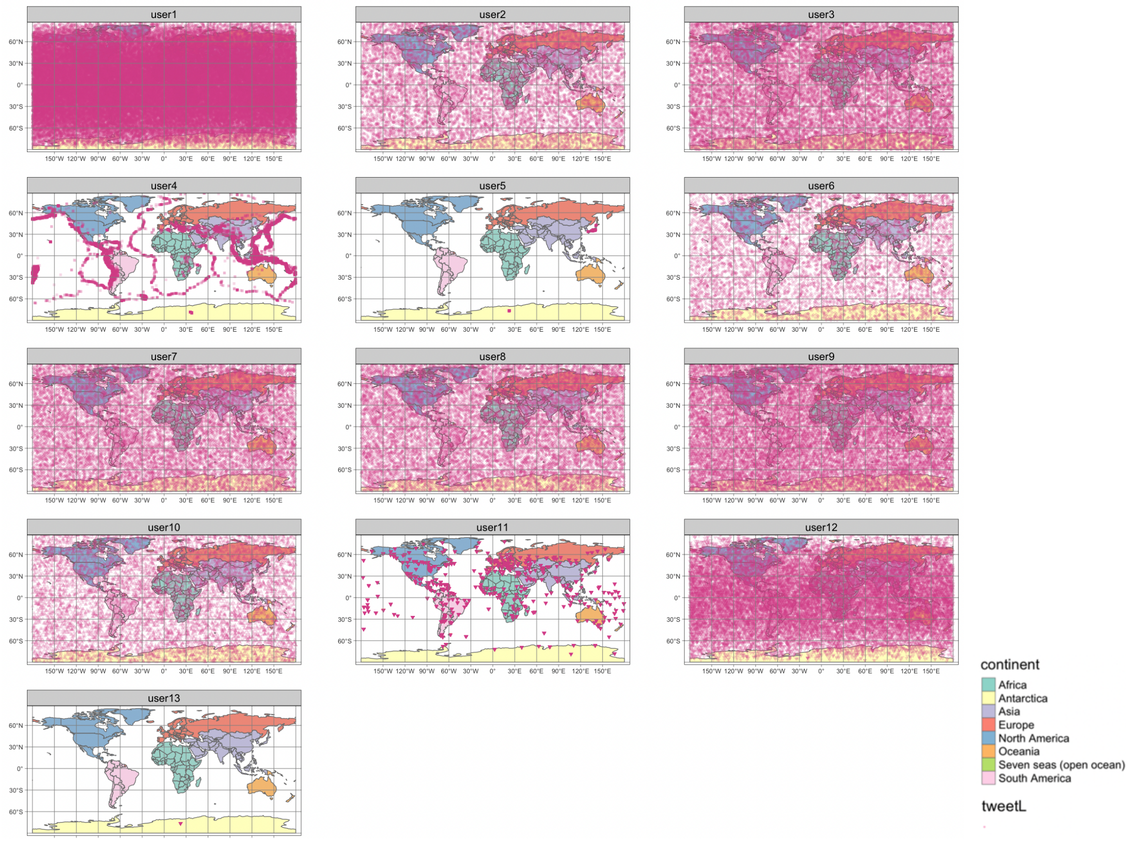
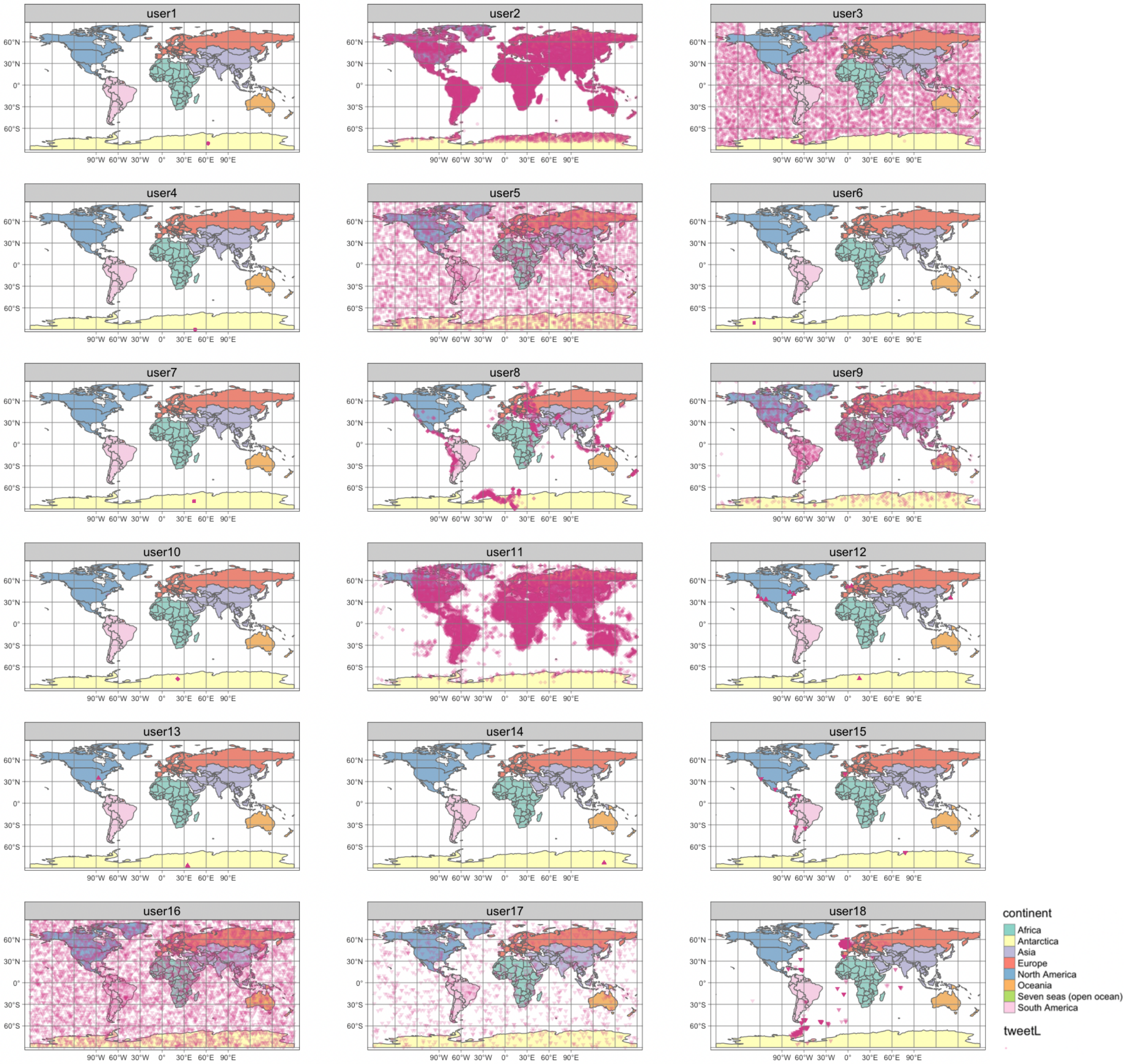


Figure 3. Distribution of tweets in terms of users' inbound and outbound activity levels.

(c) inbound activity level-3



(d) inbound activity level-4

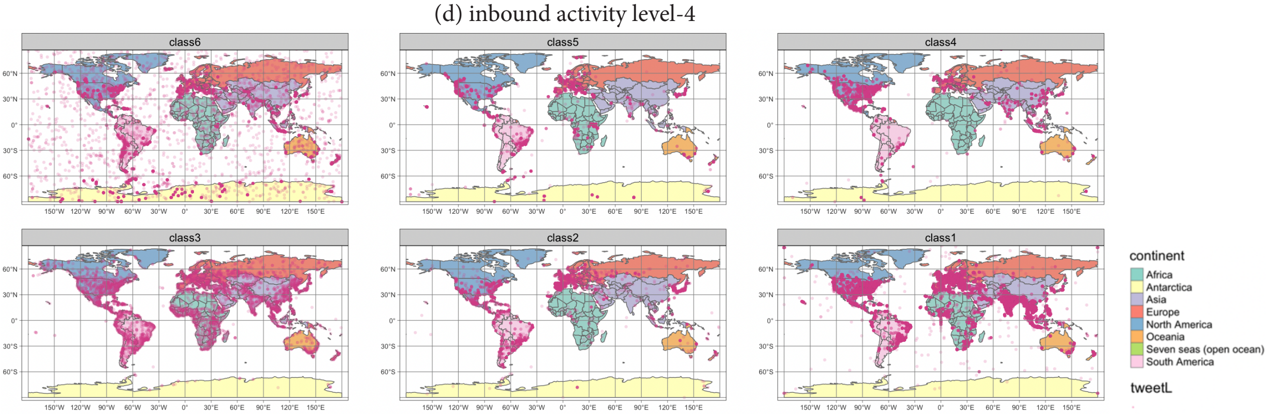


Figure 3. (Continued)

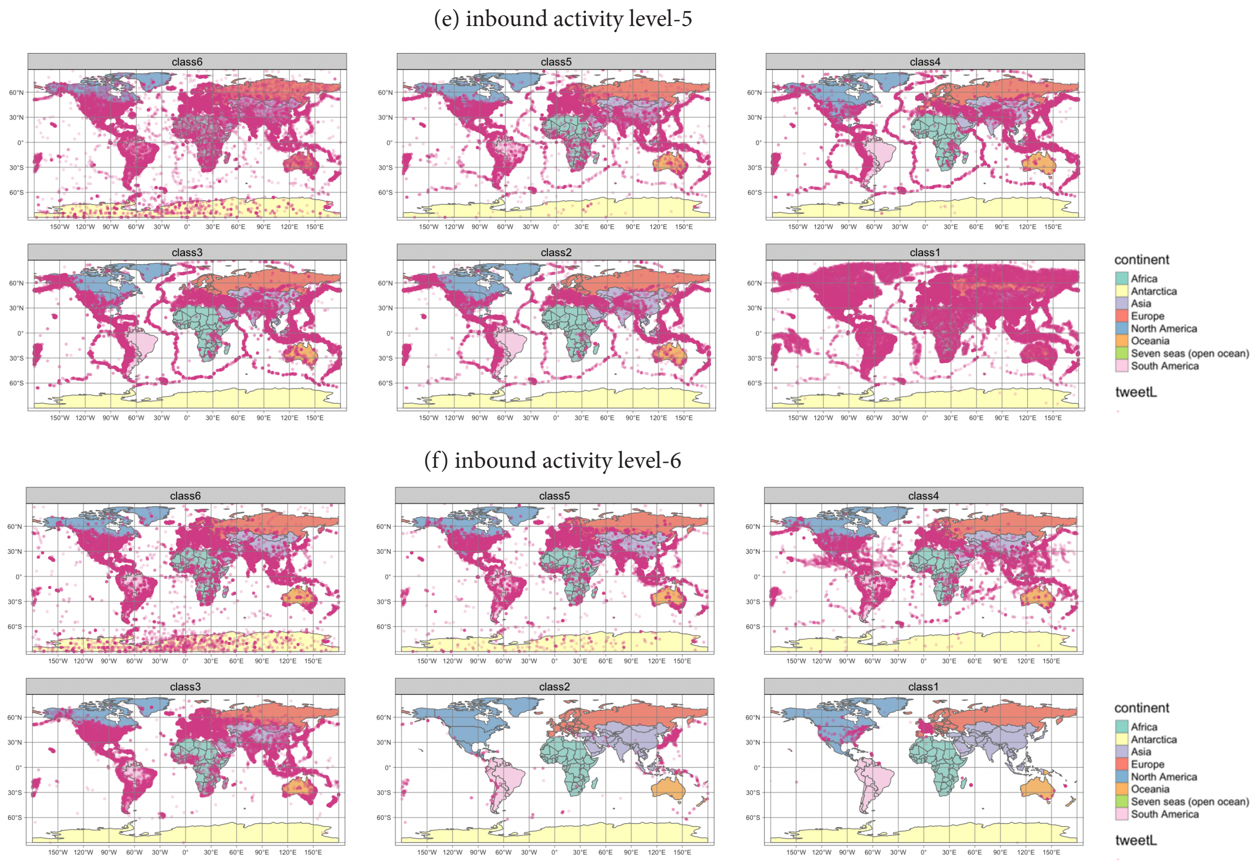


Figure 3. (Continued)

levels (4th, 5th, and 6th), we employ a second classification approach using the Jenks Natural Breaks Classification algorithm. This data-driven algorithm minimizes variance within each class and maximizes variance between classes. After conducting several tests, we determined that six groupings for the subactivity levels adequately represent the general spatial distribution. The resulting levels are displayed in Figures 3d, 3e, and 3f, arranged from the most active to the least active.

The spatial distribution depicted in Figure 3 offers valuable insights into data accountability and provides a deeper understanding of the data. In the first level, the behavior of the most active user remains consistent across the global boundary, reinforcing the initial inference of their activity within the Antarctica bbox. In this case, it is evident that the most active user is a bot disseminating tweets that hold no relevance to Antarctic studies. The second level consists of 13 users with their spatial activity plots, with most of them densely covering various regions worldwide. However, a few users exhibit footprints in specific regions (Figure 3 (b) - users 4, 5, 11, 13).

The analysis reveals interesting findings in the second level, where users with global footprints are likely to be bots providing meaningless spatial information. On the contrary, users with recurring footprints from specific locations (such as user 5 and user 13) could be stationary bots, sensors, or researchers operating within a specific research base area. It is important to consider that users may manually manipulate their location within the Twitter app, potentially adding a location in Antarctica while being situated far away. Similarly, the third level exhibits digital footprints resembling those in the second level. Some users have footprints covering both land and sea across the entire world, while others are concentrated in specific land areas (Figure 3c).

The fourth, fifth, and sixth levels are categorized by the Jenks Natural Breaks Classification algorithm. Although the algorithm was applied to predefined groups of users, the resulting spatial plots within each group exhibit diverse patterns that are challenging to interpret. Therefore, there is a need for further investigation into the spatial footprints of users. To address this requirement, we conducted a series of spatial queries explained accordingly in the 2.4 subsection to examine users' spatial activities. The results of this investigation into the spatial consistency of inbound and outbound tweets are summarized in Table 2.

Users' spatial behavior is evaluated based on their activity level, considering the origin of their tweets (inbound or outbound), overall movement (stationary or moving), and partitioned behavioral tendencies within trajectory legs (stationary or moving). Moreover, each identified moving leg is classified as either consistent or inconsistent based on the velocity (measured in km/h) and the spatial distance/time difference between consecutive tweets from the same user. Stationary action is defined by a threshold of 300 m, while inconsistent velocity is determined as 1000 km/h. Consequently, the most dedicated users in the Antarctic domain are found within the third activity level, with 22% of these users exclusively tweeting within the boundaries of Antarctica. However, it is likely that these users belong to a group of stationary bots, as the stationary users are consistently the same individuals. The behavior of other users, both inbound and outbound, shows similar activity level patterns. Additionally, the proportion of stationary users aligns with the group of users who do not tweet outside of Antarctica. To examine moving behavior in more detail, consecutive tweets from each user are evaluated by considering the time and spatial distance between them. These evaluations result in three categories: stationary leg, inconsistently moving leg, and consistently moving leg. Notably, the user with the least proportion of stationary legs also exhibits the highest proportion of inconsistent movement (as shown in Table 2). Similarly, the user with the least proportion of inconsistently moving legs and the highest proportion of consistently moving legs belongs to activity level class 6. Upon further examination of Table 2, it becomes evident that the percentage of moving legs classified as consistent increases inversely with the activity level. The least active group (class 6) demonstrates the highest proportion of consistently moving legs, whereas the most active group (class 1) exhibits an almost negligible ratio of consistently moving legs.

4. Discussion

This study delves into the findings and implications of the examination of a five-year dataset retrieved from Twitter, with a primary focus on exploring the potential of SMD in mapping Antarctica. The dataset comprises over 150,000 tweets originating from approximately 4000 users and serves as the foundation of this exploration. This study (especially aiming to enhance both geographical and semantic information) investigates a new approach to data reliability in an area with data scarcity. The study disregards the common practice of omitting data from the most active users, which is frequently employed in social media analysis to improve data accuracy. Alternatively, the study explores a new method of data reliability by focusing on the spatial details (distribution and consistency) in terms of user activity levels. Specifically, the hypothesis that users who appear active within a certain bounding box may not be active in another bbox is examined in the context of other subactivity levels (outbound activity level).

This work demonstrates the potential error in preprocessing steps focusing solely on user activity levels, as seen in many studies that follow a user-centric approach to track specific regions. It suggests that evaluating the spatial consistency, whether it is consistent or inconsistent, should be coupled with the user's activity level, rather than relying solely on activity level-based data filtering. In addition, when determining activity levels, the commonly used k-means, x-means, and kernel density estimation algorithms are found to be inappropriate for skewed data. The study emphasizes the necessity for experimental research to define possible activity levels.

The data analysis reveals two main findings. The first finding is that although a positive correlation between inbound activity levels and spatial inconsistency rates is observed, even users with the lowest activity levels display inconsistent movement patterns, with nearly one-third exhibiting such behavior. The second finding is that a significant portion, 70% of the entire dataset, shows inconsistent spatial movement, primarily among the most active user groups. This aligns with the perception that a significant portion of social media data lacks reliability. Nonetheless, data generated with textual and visual content requires further investigation into its individual spatial accuracy.

Table 2. Summary of users' inbound and outbound spatial consistency in terms of inbound activity levels.

Inbound activity level	# of inbound tweets	# of outbound tweets	% of users not having outbound tweets	% of stationary user	% of stationary leg	% of inconsistent moving leg	% of consistent moving leg
1	91,750	1,618,214	0	0	1.2	98.8	0
2	44,048	663,638	8	8	27.6	71.3	1.1
3	7437	77,202	22	22	10.6	80.9	8.5
4	2315	293,970	6	5	43.4	48.3	8.3
5	2309	1,389,336	5	4	10.2	69.8	20.0
6	2970	736,142	5	5	27.7	32.4	39.9

Due to the malleable nature of location data, the study presents a meaningful result in assessing the credibility of user-based reliability in data sources. The evaluation of the spatial accuracy of individual data points calls for additional research, focusing on textual and visual content. Particularly in regions facing data scarcity, social media can contribute significantly to expanding geographical information systems' datasets. Being a continuous and cost-effective data source, it holds immense value, especially when evaluated for data quality and its distinctive contributions.

5. Conclusion

Social media serves as a valuable source of data in various fields, offering significant contributions without the need for financial investment. Nonetheless, it is essential to acknowledge the costs associated with time and accountability. Despite these considerations, the existing content, even without additional effort, holds immeasurable value, as even minor inferences can be drawn from it. Particularly in the context of gathering information from remote regions, social media's contribution can become pivotal in environmental monitoring.

Future studies could focus on examining intrinsic data quality, considering factors such as trends, anomalies, and biases within the data. Although the retrieved data does not lend itself well to plotting anomalies over time, it can be utilized to visualize spatial biases. Additionally, employing normalization techniques could facilitate data analysis on a seasonal basis, despite occasional missing time intervals during data retrieval. Similar steps can be applied to investigate seasonal distribution patterns and concentrations in specific geographic areas.

Our study presents a methodology and a case study to assess the accountability and consistency of SMD based on user activity levels. However, the outcomes are subject to limitations imposed by Twitter API restrictions, system-related issues (e.g., power outages, internet disruptions, software crashes), and reliance on a single social media platform as the data source. Another limitation of this study is the evaluation methodology's application to the overall data concerning activity levels. Real users possess the ability to manipulate their spatial content by manually selecting random locations from the platform's places library for geotagging.

Future research endeavors aim to evaluate user's spatial contribution and accountability by combining spatial consistency and auxiliary data (text and visual content) semantic consistency. This study contributes to understanding the reliability of SMD for mapping Antarctic regions. Although the methodology was applied to a dataset from a single platform, the reproducibility of results across other platforms and data forms is straightforward.

Acknowledgment and/or disclaimers if any

This study was carried out under the auspices of the Presidency of the Republic of Turkey, supported by the Ministry of Industry and Technology (TÜBİTAK, project no. 122Y193), and coordinated by TÜBİTAK MAM Polar Research Institute.

We would like to acknowledge the invaluable assistance provided by ChatGPT, a language model developed by OpenAI, in generating and refining the content presented in this article. ChatGPT played a significant role in enhancing the clarity, elegance, and academic tone of the writing. Its contributions in providing informative and well-structured paragraphs greatly improved the overall quality of the article. We extend our sincere appreciation to OpenAI for developing this powerful language model, which has proved to be an invaluable resource in our research endeavors.

References

Ballatore A, Zipf A (2015). A conceptual quality framework for Volunteered Geographic Information. In Spatial Information Theory: 12th International Conference, COSIT 2015, Santa Fe, NM, USA, October 12-16, 2015, Proceedings 12 (pp. 89-107). Springer International Publishing. https://doi.org/10.1007/978-3-319-23374-1_5

Ballatore A, De Sabbata S (2018). Charting the geographies of crowdsourced information in Greater London. In Geospatial Technologies for All: Selected Papers of the 21st AGILE Conference on Geographic Information Science 21 (pp. 149-168). Springer International Publishing. https://doi.org/10.1007/978-3-319-78208-9_8

Basiri A, Haklay M, Foody G, Mooney P (2019). Crowdsourced geospatial data quality: Challenges and future directions. International Journal of Geographical Information Science, 33 (8): 1588-1593. https://doi.org/10.1080/13658816.2019.1593422

Baumhoer CA, Dietz, AJ, Dech S, Kuenzer C (2018). Remote sensing of Antarctic glacier and ice-shelf front dynamics-A review. Remote Sensing, 10 (9): 1445. https://doi.org/10.3390/rs10091445

Brovelli MA, Ponti M, Schade S, Solís P (2020). Citizen science in support of digital earth. Manual of digital earth, 593-622. https://doi.org/10.1007/978-981-32-9915-3_18

Dong Y, Zhao J, Li C, Liao M (2022). Gapless-REMA100: A gapless 100-m reference elevation model of Antarctica with voids filled by multi-source DEMs. ISPRS Journal of Photogrammetry and Remote Sensing, 186, 70-82. https://doi.org/10.1016/j.isprsjprs.2022.01.024

Frame B, Liggett D, Lindström K, Roura RM, van der Watt LM (2022). Tourism and heritage in Antarctica: exploring cultural, natural and subliminal experiences. Polar Geography, 45 (1): 37-57. https://doi.org/10.1080/1088937x.2021.1918787

Gao S, Janowicz K, Couclelis H (2017). Extracting urban functional regions from points of interest and human activities on location-based social networks. Transactions in GIS, 21 (3): 446-467. https://doi.org/10.1111/tgis.12289

Ghermandi A, Sinclair M (2019). Passive crowdsourcing of social media in environmental research: A systematic map. Global environmental change, 55: 36-47. https://doi.org/10.1016/j.gloenvcha.2019.02.003

Goodchild MF (2007). Citizens as sensors: the world of volunteered geography. GeoJournal, 69: 211-221. https://doi.org/10.1007/s10708-007-9111-y

Gulher E, Alganci U (2023). Satellite-Derived Bathymetry Mapping on Horseshoe Island, Antarctic Peninsula, with Open-Source Satellite Images: Evaluation of Atmospheric Correction Methods and Empirical Models. Remote Sensing, 15 (10): 2568. https://doi.org/10.3390/rs15102568

Gulnerman AG (2021). Changing Pattern of Human Movements in Istanbul During Covid-19. In Computational Science and Its Applications–ICCSA 2021: 21st International Conference, Cagliari, Italy, September 13–16, 2021, Proceedings, Part IX 21 (pp. 220-230). Springer International Publishing. https://doi.org/10.1007/978-3-030-87013-3_17

Gulnerman AG, Karaman H, Basiri A (2021). New age of crisis management with social media. Open Source Geospatial Science for Urban Studies: The Value of Open Geospatial Data, 131-160. https://doi.org/10.1007/978-3-030-58232-6

Gulnerman AG, Karaman H, Pekaslan D, Bilgi S (2020). Citizens' spatial footprint on Twitter—anomaly, trend and bias investigation in Istanbul. ISPRS International Journal of Geo-Information, 9 (4): 222. https://doi.org/10.3390/ijgi9040222

He Y, Liu C (2023). Environmental Impacts and Sustainable Development of Antarctic Tourism: The Chinese Tourists' Perspectives. Towards Sustainable and Resilient Tourism Futures. https://doi.org/10.37307/b.978-3-503-21195-1.04

IAATO. (2022). Overview of Antarctic Tourism: A Historical Review of Growth, the 2021-22 Season, and Preliminary Estimates for 2022-23 (Report No. ATCM17). Antarctic Treaty Consultative Meeting (ATCM). https://iaato.org/wp-content/uploads/2022/06/ATCM44-IAATO-Overview.pdf

Isler DB, Selbesoglu MO, Yavasoglu H, Karabulut MF, Gulal E et al. (2021, December). Sea and Ice level Monitoring on Horseshoe Island by GNSS Observations during the 4th National Antarctic Science Expedition of Turkey. In AGU Fall Meeting Abstracts (Vol. 2021, pp. G35A-0289).

Jenks GF (1967). The data model concept in statistical mapping. International yearbook of cartography, 7: 186-190.

Jurdak R, Zhao K, Liu J, AbouJaoude M, Cameron M, Newth D (2015). Understanding human mobility from Twitter. PloS one, 10 (7): e0131469. https://doi.org/10.1371/journal.pone.0131469

LaRue MA, Ainley DG, Pennycook J, Stamatiou K, Salas L et al. (2020). Engaging 'the crowd'in remote sensing to learn about habitat affinity of the Weddell seal in Antarctica. Remote Sensing in Ecology and Conservation, 6 (1): 70-78. https://doi.org/10.1002/rse2.124

Pina P, Vieira G (2022). UAVs for science in Antarctica. Remote Sensing, 14 (7): 1610. https://doi.org/10.3390/rs14071610

See L, Mooney P, Foody G, Bastin L, Comber A et al. (2016). Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. ISPRS International Journal of Geo-Information, 5 (5): 55. https://doi.org/10.3390/ijgi5050055

Selbesoglu MO, Bakirman T, Vassilev O, Ozsoy B (2023). Mapping of Glaciers on Horseshoe Island, Antarctic Peninsula, with Deep Learning Based on High-Resolution Orthophoto. Drones, 7 (2): 72. https://doi.org/10.3390/drones7020072

Singleton AD, Spielman S, Folch D (2017). Urban analytics. Sage.

Yirmibesoglu S, Oktar O, Ozsoy B. (2022, December). Remote Sensing Research of Sea Ice-free days for Horseshoe Island, Antarctica. In AGU Fall Meeting Abstracts (Vol. 2022, pp. C11C-05).