Received: 23 February 2023 | Revised: 8 June 2023 | Accepted: 19 June 2023

DOI: 10.1111/tgis.13089

RESEARCH ARTICLE

Transactions in GIS WILEY

Effects of reverse geocoding on OpenStreetMap tag quality assessment

Batuhan Kilic[1] | Müslüm Hacar[1,2] | Fatih Gülgen[1]

[1]Department of Geomatic Engineering, Yildiz Technical University, Istanbul, Türkiye

[2]Visiting fellow at DESTeC, Architettura e Urbanistica, University of Pisa, Pisa, Italy

Correspondence
Batuhan Kilic, Department of Geomatic Engineering, Yildiz Technical University, Istanbul 34220, Türkiye.
Email: batuhank@yildiz.edu.tr

Abstract

OpenStreetMap (OSM), a widely-used open-source geographic information system platform, provides a vast geographic dataset in which users contribute both geometric information (nodes, ways, and relations) and semantic information (tags). This method of voluntary contributions is governed by the collective effort of the users. It is widely acknowledged that the quantity of tag information is substantial, but its quality is often poor. Researchers are therefore trying to assess the quality of the tags and enhance the data through various integration experiments. This article investigates the validity of the tags for geographical objects in metropolitan areas using municipal data and a reverse geocoding technique. The proposed method evaluates the data quality and the matching process carried out by reverse geocoding, using municipal points of interest as a reference. The accuracy of the tag and address information and road network centrality metrics were assessed for the OSM objects that were matched to the locations of interest. The tags were found to match the points of interest with an accuracy of 88%. Furthermore, the tag values were categorized and analyzed based on their similarity. It is concluded that in metropolitan settings where centers of interest are closely located, the accuracy of tags and addresses tends to decrease.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2023 The Authors. *Transactions in GIS* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

In all domains of Geography and Geographic Information Science (GISc), georeferenced location plays a crucial role. With advancements in Earth observation technologies, it has become easier to establish a connection between general data and their geospatial points on digital maps (Kilic & Gülgen, 2020a; Ningsheng et al., 2015). The process of transforming between the spatial and place-based representations of a location, both forward and backward, is an important task that forms the basis of various research, applications, and system architectures (McKenzie & Janowicz, 2015).

Geocoding is the process of assigning geographic coordinates on the Earth's surface to associated georeferenced textual information, such as postal addresses and place names (Goldberg et al., 2007; Kilic & Gülgen, 2020b; Zandbergen, 2008). The inverse process, known as reverse geocoding, involves assigning the postal addresses of a place to its geographic coordinates (Kilic & Gülgen, 2020a; Kounadi et al., 2013; Li, 2018). The concept behind reverse geocoding is that a user can search for features or attributes near their location or in any area by providing specific geographical coordinates. This approach links the location of interest, such as points of interest (POIs), to the nearest geographic position based on Euclidean distance. Reverse geocoding can be performed in either a conventional (classical) or online manner. Online reverse geocoding operates on a network-accessible system and is usually accessible through the internet using a web application programming interface (API), while conventional reverse geocoding uses the core packages included in GIS software (Kilic & Gülgen, 2020a). Online reverse geocoding is mainly provided by commercial map providers such as Google Maps, Bing Maps, Esri, and HERE Maps to the public and end-users.

Depending on the research objective and the type of data collected, reverse geocoding is extensively used across various fields. Some studies focusing on the reverse geocoding process address crime and privacy issues (Kounadi et al., 2013; Krumm, 2007). For example, Krumm (2007) attempted to identify individuals' home geographical coordinates (with a median error below 60 m) and identities (with over 5% success) based on 2 weeks of anonymized GPS tracks gathered from 172 drivers. On the contrary, as additional data such as human activities and POIs are integrated with locations, reverse geocoding has been applied to retrieve information for enhancing evacuation warnings, site selection, location-based social networks, and human trajectory recovery (Li et al., 2017; Lv et al., 2016; McKenzie & Janowicz, 2015; Yameqani & Alesheikh, 2019). These applications not only increase the content of reverse geocoding output but also generate a significant need for improved reverse geocoding output quality to assist end-users and researchers. Despite the broad use of reverse geocoding in professional fields, relatively few studies focus on the quality of existing reverse geocoding methods (Yin et al., 2020). For example, Kounadi et al. (2013) investigated the accuracy of seven different online reverse geocoding services, including ArcGIS, Bing Maps, Geonames, Google Maps, Nominatim, TravelGIS, and Yahoo. They evaluated the precision level by comparing the retrieved addresses with the reference addresses. Kilic and Gulgen (2017) examined the accuracy of online reverse geocoding using Google and Bing Maps for accommodation providers in a selected study region in Turkey and the USA. Similarly, Kilic and Gülgen (2020a) conducted a comparative evaluation of four different reverse geocoding services (ArcGIS Online, Bing Maps, Google Maps, and Yandex Maps) specifically focusing on POIs in urban areas. Yin et al. (2020) incorporated metadata descriptions into the reverse geocoding process to improve the quality of reverse geocoding outputs. They proposed a probabilistic framework that provides multiple suggestions closely related to the input coordinates, corrected typologically, and ordered according to an organized standard.

OpenStreetMap (OSM) is a volunteered geographic information (VGI) platform that continues to evolve, serving as a valuable resource for geospatial information. Through crowdsourcing, geographical contributions are created, structured, and shared. Unlike commercial mapping platforms like Google Maps and Bing Maps, OSM allows both volunteers and non-experts to contribute to mapping through online interfaces (Goodchild, 2007; Heipke, 2010). The number of registered users for the OSM project is steadily increasing, and there is a wealth of research highlighting the participants and their contribution trends (Arsanjani

14679671, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tgis.13089 by Yildiz Technical University, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

et al., 2015; Hacar, 2022; Neis & Zielstra, 2014; Neis & Zipf, 2012). This emphasizes the importance of evaluating the geometric, topological, and semantic quality of the map features added and edited in VGI before using them in research (Haklay, 2010). Two primary strategies are employed to analyze the data quality of OSM: intrinsic measurements (without reference data) and extrinsic measurements (with reference data). The accuracy of the semantic tags and address information in VGI can be evaluated through geocoding and reverse geocoding. OSM, being a prominent VGI project and a global mapping platform, provides a free and open-source search engine called Nominatim. Through reverse geocoding, users can derive different labels or address information for OSM data using existing coordinate values. Conversely, geocoding enables users to search for existing data using descriptive terms such as names and addresses (https://wiki.openstreetmap.org/wiki/Nominatim). Similar to other mapping services, the reliable implementation of OSM reverse geocoding depends on the accuracy of the input data and the service database. Amelunxen (2010) examined the suitability of VGI for geocoding research and found that the geocoding results using OSM data were more accurate than the commercial geocoding service provided by Google Maps, but this was observed only when door numbers were available in the reference database. Additionally, Kounadi et al. (2013) determined that the Nominatim reverse geocoding service ranked third in terms of high precision out of seven different approaches. Clemens (2015) compared the accuracy of the Nominatim service and the Elasticsearch geocoding services (written in Java) using 1000 pharmacy records in Germany and found that Nominatim was unable to geocode 54 addresses, while Elasticsearch geocoded all addresses within 100 m or less. Additionally, it is stated that the Nominatim service is better suited to particular address formats, and the fact that different addressing systems exist around the world is a limitation of the Nominatim service.

In this study, we investigate the performance of the Nominatim reverse geocoding service and its impact on the quality of OSM tags, with a specific focus on two domains: positional accuracy evaluation and tag (metadata) information. The benchmark dataset was obtained from the Directorate of Information Technologies of the Fatih Municipality, which is responsible for generating official POIs. First, the postal addresses and tag information of OSM objects were obtained through reverse geocoding. The matching status was then determined by manually comparing the OSM objects with the reference data (True/False). Unlike other studies, the authors evaluated how centrality and density measures influence the success of the reverse geocoding process on OSM tag metadata. The remainder of this article is structured as follows. The next section provides detailed information on the study region, the data utilized, and the methodology. The third section presents the results and discussion derived from the analysis. The article concludes with a discussion of future work perspectives.

2 | MATERIALS AND METHODS

2.1 | Nominatim

The Nominatim reverse geocoding service was selected and analyzed to improve the quality of OSM tags (https://nominatim.org/release-docs/develop/). The service operates on a network-accessible system and is typically accessed through a Web API. Reverse geocoding transforms geographic coordinate pairs into corresponding textual descriptions, and presents parameters such as a postal address, coordinate pairs, and precision levels as a place name and additional information to the user. However, like any reverse geocoding service, Nominatim has varied constraints, licensing requirements, and limitations on the usage of free parameters, and offers different output formats, including JavaScript Object Notation (JSON) and Extensible Markup Language. To perform the reverse geocoding process, a Python library called "Geocoder" was used in this study (Carriere, 2018). This library supports various reverse geocoding services, such as Google, Bing, and OSM. Using this library, each coordinate pair was automatically reverse geocoded and descriptive information was presented to the users by the system.

14679671, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tgis.13089 by Yildiz Technical University, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.2 | Study area, data set, and methodology

The empirical test site was selected as Istanbul, Turkey, a city with a high concentration of POIs. With a population of 15.9 million, Istanbul is one of the most populous cities in Europe (TUIK, 2021). It is also a popular tourist destination, attracting 13.4 million visitors each year, making it the eighth most visited city in the world (Travelness, 2022). Istanbul is known for its unique blend of European and Asian cultures, with a rich Byzantine and Ottoman cultural heritage. It was designated as the European Capital of Culture in 2010 for its historical sites that draw many visitors (UNESCO, 2006). Most tourists use mobile mapping services to navigate to tourist destinations, hotels, and other POIs. The Fatih district, located on the historical peninsula of Istanbul, is considered the most cultural, authentic, and touristic area. Surrounded by the Byzantine walls, the Golden Horn, and the Sea of Marmara, Fatih contains the heritage of the Ottoman and Byzantine empires and is one of the most densely populated areas in Istanbul. With a total area of 15.6 km^2, Fatih consists of 57 neighborhoods (Figure 1).

Fatih is a district of Istanbul where small-scale businesses are concentrated and shop circulation is low, making it an important destination for both locals and tourists. These businesses typically have a long history of serving their communities, and any changes in ownership for financial reasons are recorded and updated in the information system used by the local municipality as well as in the spatial address registration system. In Fatih, which has a predominantly urban structure, there are few significant changes in the building footprint. However, the local municipality keeps track of any modifications resulting from building renovation activities, undertaken for reasons such as aging or earthquake safety. Meanwhile, Istanbul Metropolitan Municipality is responsible for regularly updating maps and creating street views at approximately 5-year intervals (https://rehber.tucbs.gov.tr/).

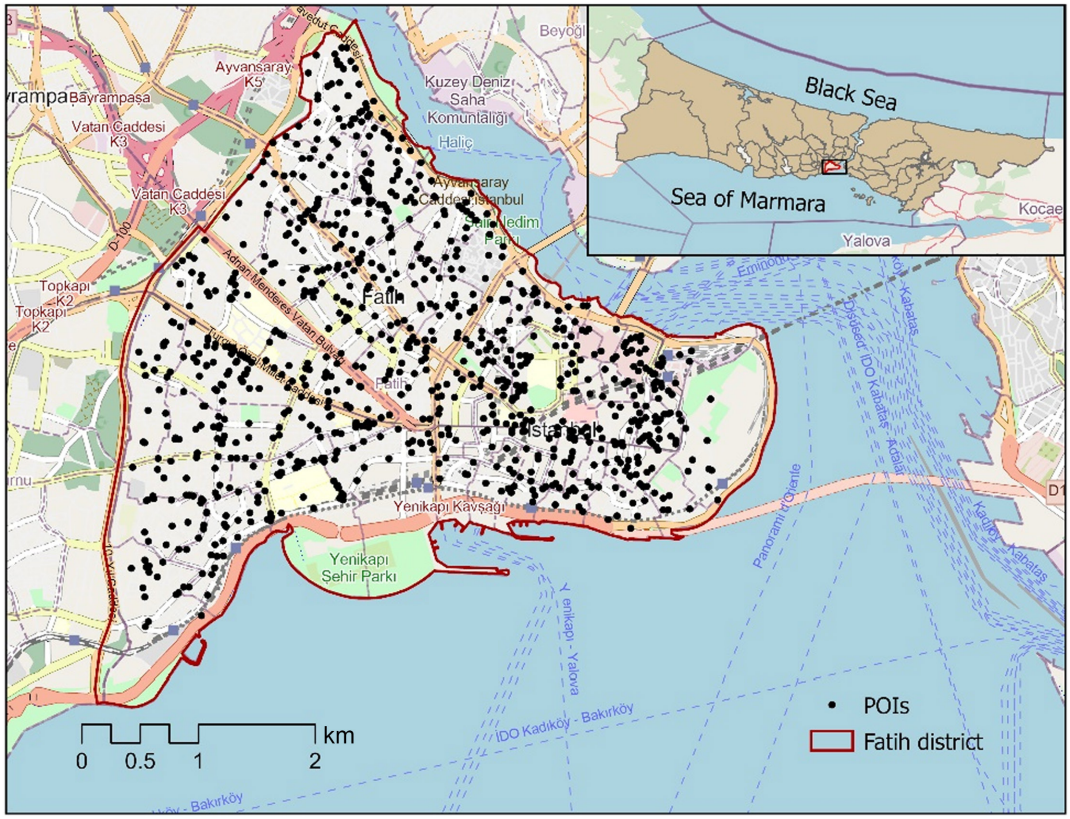

FIGURE 1 The study area.

14679671, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tgis.13089 by Yildiz Technical University, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

POI data and descriptive information, such as geographic coordinates, postal address, and POI type, were obtained from the Directorate of Information Technologies of the Fatih Municipality. This organization is responsible for creating source geographic data and was accepted as the official reference for this study. The total number of POI data, which includes various types such as sales, government organizations, tourism, infrastructure, health, religious facilities, cultural, transportation, and educational institutions, was 921. The relevant employees of the Directorate conduct field observations periodically to collect POI data, which is then entered into digital map databases. This ensures that the POI data are accurate and up-to-date.

The aim of the proposed approach is to evaluate the effectiveness of the Nominatim service for reverse geocoding research by examining the OSM tags that correspond to the POI data from the Fatih Municipality. The process involves utilizing the Nominatim reverse geocoding service to extract various descriptive information, such as a postal address, coordinate pairs, OSM id, OSM type, Place id, OSM Category, and OSM Ctype, of the POIs in GeoJSON format. This information is then transferred to a GIS database for analysis. To verify the quality of the reverse geocoding process, the key-value pairs between the reference geographic data and the VGI data are compared and manually matched. The success of the Nominatim reverse geocoding method is measured by calculating the total accuracy, as determined by the match rate results for each OSM category type. In addition, centrality and density measurements are used to analyze the structure of correctly and incorrectly matched POIs in the street network, and to determine their significance to the network as a whole and their impact on urban sprawl. The OSM road network data are used for the centrality measure, while the OSM polygon data are used for the density measure.

$$\text{Total accuracy} = \frac{\text{True match}}{(\text{True match} + \text{False match})} \tag{1}$$

Centrality measures determine the importance of some nodes in a network compared to others and assess how centrality values are distributed across all nodes (Hacar et al., 2018). Urban planning and analysis use various centrality measures such as degree, closeness, betweenness, PageRank, katz, and straightness (Corcoran et al., 2013; Crucitti et al., 2006; Cutini, 2001, 2016; Freeman, 1977; Strano et al., 2012). In this study, the values of the closeness centrality measure (CentCls-POI) for matched POIs are calculated using the current OSM road nodes. Each POI gets its centrality measure from the OSM node that is closest to it geographically. The density measure, on the contrary, assesses the density of point details surrounding each output raster cell and applies this process to the entire region based on the spatial relationship between the measured value and its position. The proposed approach utilizes the kernel density tool in ArcMap 10 software (https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analysis-toolbox/how-kernal-density-works.htm) based on the quartic formulation of Silverman (1986) and about the geometric center points of OSM buildings. As a result, each POI is assigned a kernel density value (DensPOI) from the nearest pixel. To evaluate the compatibility between municipal POIs and OSM data, the proposed approach enriches the analysis dataset with such measurements (Figure 2).

3 | RESULTS AND DISCUSSIONS

The descriptive statistical results of the correct and incorrect matching indices between the municipality's POIs and their respective objects in OSM are presented in Table 1. The evaluation of the corresponding items in OSM was based on the geometry type (point and polygon) and the correct and incorrect matching rates, as well as the total accuracy.

When the findings in Table 1 are analyzed, it can be seen that the majority of the corresponding objects in OSM are POIs with the “amenity” tag, followed by those with the “building” and “shop” tags. In general, a total accuracy of 88% was found for all tags in the matches with geometry types. In terms of correct matching rates for the polygon geometry type, correct matching values above 90% were achieved for all tag types, except for the

14675671, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tgis.13089 by Yildiz Technical University, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

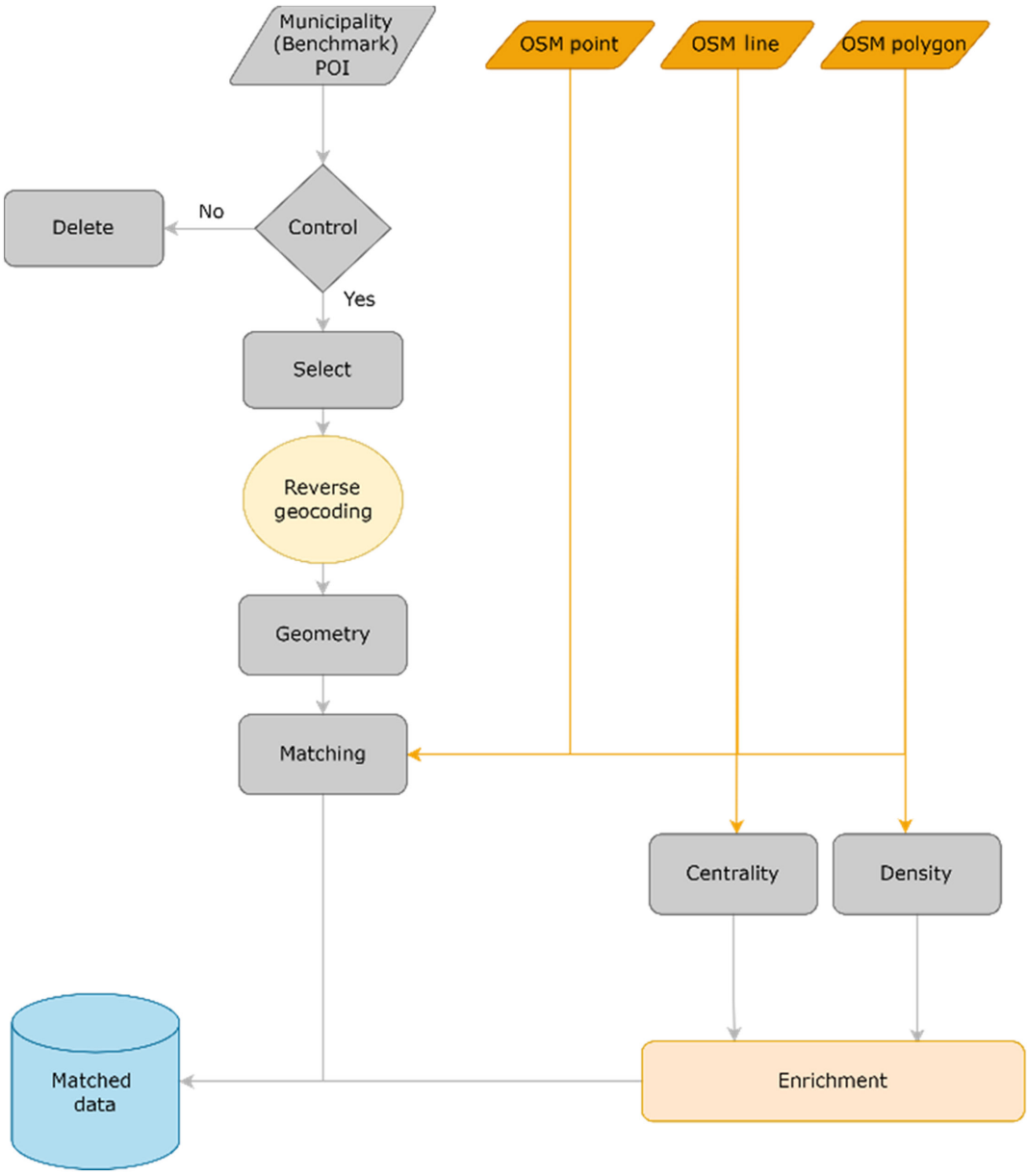

FIGURE 2 The flowchart of the proposed approach.

tags "leisure", "shop", and "tourist". The extremely low number of objects (3, 1, and 16, respectively) associated with these labels renders statistical evaluation insufficient. In contrast, when examining the findings of the point geometry type, it reveals that the correct matching rates for objects with the tag "amenity" and objects with the tags "shop" and "tourist" are lower than 90%. Among these tags, the number of objects with the "tourism" tag was not taken into consideration, as it was not statistically significant to examine (the number of significance test measures was less than 30). The remaining objects with the tags "amenity" and "shop" were subjected to a more in-depth analysis due to their lower correct match rates relative to other tags and statistically significant data quantities.

14675671, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tgis.13089 by Yildiz Technical University, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Statistical results of matching points of interests.

Key	Point			Polygon			Total accuracy (%)
	Count	True (%)	False (%)	Count	True (%)	False (%)	
Amenity	369	85.6	14.4	291	97.3	2.7	90.8
Building	2	100.0	0.0	326	99.4	0.6	99.4
Historic	–	–	–	4	100.0	0.0	100.0
Leisure	4	100.0	0.0	3	33.3	66.7	71.4
Office	38	92.1	7.9	3	100.0	0.0	92.7
Shop	58	10.3	89.7	1	0.0	100.0	10.2
Tourism	22	45.5	54.5	16	81.3	18.7	60.5
Total (%)							88.0

3.1 | Evaluation results for address accuracy

For the qualitative evaluation of objects with the tags "amenity" and "shop," in addition to the overall matching rates (correct, incorrect, and total accuracy), spatial evaluations for address accuracy and various statistical values, such as CentCls-POI and DensPOI, were assessed for tag accuracy. This allowed for the determination of the direction in which the OSM reverse geocoding service results can impact address and tag accuracy.

A simple method for assessing the quality of correctly and incorrectly matched objects belonging to the tags "shop" and "amenity" is positional accuracy, defined by the Euclidean distance between the benchmark address and the address retrieved by the OSM service. This distance value indicates how close the geocoded object is to the "actual" location of the address (Zandbergen, 2009). The overall positional accuracies of the reverse geocoded objects generated by the OSM service for both tags are shown in Figure 3.

Examining the results of positional accuracy reveals a similar distribution of objects tagged with "shop" across the ranges of 0–10 m, 10–20 m, and 20–100 m. However, approximately 80% of objects labeled as "amenity" have a positional error of less than 10 m, while the number is lower in other error ranges (as shown in Figure 3). In addition, it can be observed from Figure 3 that the positional accuracy of all matches (correct and incorrect) for both tags is 100 m or less. In the literature on geocoding (Bonner et al., 2003; Kounadi et al., 2013; Strickland et al., 2007), a geocoded point is considered to be very close to its actual location if its positional accuracy is 100 m or less. However, when analyzing mismatched OSM data in this study, it was found that the error distances were 100 m or less. This confusion arises mainly because the study area is a historical and touristic settlement characterized by high density, meaning numerous POIs per unit area exist. Although the postal addresses of two adjacent buildings with different door numbers are different, their entrances can be close to each other (Kilic & Gülgen, 2020a). On the contrary, current postal address retrieval solutions often only focus on distance, ignoring spatial topology linkages. The closest candidate returned may not necessarily be the correct match, as it could be on the other side of the street, even though the OSM reverse geocoding service detects a low positional error in the matched object (Yin et al., 2020). The results from this study highlight that in areas with high densities of POIs, a low degree of spatial accuracy does not necessarily indicate successful reverse geocoding.

3.2 | Evaluation results for tag accuracy

The similarity between the spatial accuracy values of the OSM reverse geocoding service for objects with the "shop" and "amenity" tags has a direct impact on the accuracy of the tag values. As a result, the tag accuracy of objects with these two tags was evaluated separately (Tables 2 and 3).

14679671, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tgis.13089 by Yildiz Technical University, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

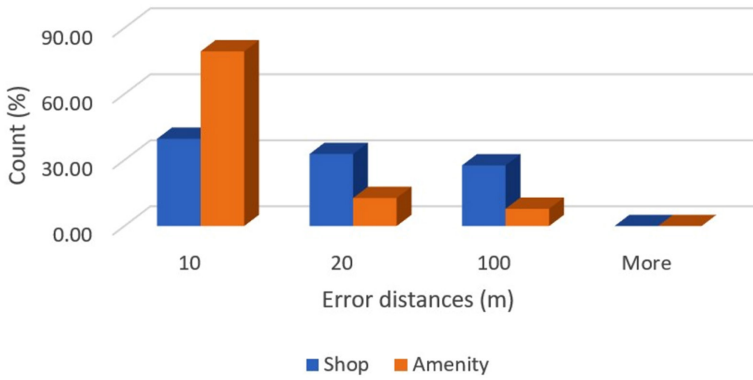

FIGURE 3 Positional accuracy of "shop" and "amenity" tags using OSM reverse geocoding.

The "shop" tag represents places in the study area where various products are sold. An analysis of the 58 POIs with the "shop" tag reveals that there are numerous classes. To conduct a semantic evaluation, the data were grouped into several classes (OSM Wiki, 2022), as shown in Table 2.

When analyzing the results in Table 2, it is evident that the OSM reverse geocoding process has a low success rate for correctly matched objects, with a tag accuracy of less than 70% in almost all groups. Furthermore, it was discovered that correctly matched objects were not grouped, but were dispersed among different groups. Notably, there were completely inaccurate matches found in high-volume data groups, such as "Clothing & Shoes," "Food & Beverages," and "General store" (15, 9, and 7 respectively), although the number of objects in these groups was not significant enough for statistical analysis. The results showed that the closeness values of the incorrect matches were greater than those of the correct matches, particularly in groups like "Art, Music & Hobbies," "Electronics," "Health & Beauty," and "Stationery & Gifts," where there were both correct and incorrect matches. This trend can still be studied despite the limited data. The OSM reverse geocoding service was found to have a higher failure rate for stores located in the center of the study region, compared to other areas. However, no significant difference was identified in terms of the density measure values between correct and incorrect matches.

Regarding objects with the "amenity" tag, which represent important and useful facilities of interest to tourists or residents in the study area, no general pattern or cluster of mismatched objects was found when analyzing the 369 correctly and incorrectly matched objects in the study region (Figure 4). The objects with the "amenity" tag were treated similarly to those with the "shop" tag, and a grouping similar to the OSM Wiki classification (2022) was given in Table 3 to establish if the erroneously matched objects belong to a particular semantic group.

Examining the values in Table 3 shows that for most groups ("Education & Learning", "Health", "Public Service", and "Religion") of objects with the "amenity" tag, a success rate of over 90% was achieved. However, in terms of false matches, it can be said that the data belonging to the "Entertainment & Meal" group has much lower accuracy (0.03%) compared to other groups, and the reverse geocoding process has not been successful for this group. A similar situation was observed for the "Bank" and "Car Service" categories, despite their low number of observations. In terms of centrality and density values, the "Health" category contains significantly more objects than the others. However, the match with the highest accuracy level (98.7%) was still obtained. The centrality value of the three incorrect matches in the study region was higher than that of the correct matches. Similar circumstances can be found within the "Education & Learning" group. It can be argued that the OSM reverse geocoding service is inadequate for such groups as it is for the "shop" tag. When examining correct and incorrect matchings, the density measure does not provide any distinguishable information.

14679671, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tgis.13089 by Yildiz Technical University, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Statistical results for "shop"-tagged objects.

G	V	C	T	TClsC (mean)	TDnst (mean)	F	FClsC (mean)	FDnst (mean)	TA (%)
Art, music & Hobbies	Craft	3	2	0.000190	0.003691	1	0.000381	0.001319	66.7
Clothing & Shoes	bag clothes fabric jewelry tailor	15	0	–	–	15	0.002300	0.002819	0.0
Electronics	mobile_phone	6	2	0.000644	0.002325	4	0.002931	0.002245	33.3
Food & Beverages	butcher convenience seafood wine	9	0	–	–	9	0.001252	0.003374	0.0
Furniture & Interior	furniture interior_decoration	2	0	–	–	2	0.004108	0.002306	0.0
General store	department_store kiosk supermarket	7	0	–	–	7	0.001912	0.002863	0.0
Health & Beauty	hairdresser perfumery	7	1	0.000152	0.004254	6	0.001842	0.003042	14.3
Household	glaziery	1	0	–	–	1	0.000152	0.003340	0.0
Stationery & Gifts	books gift stationery	5	1	0.000152	0.001815	4	0.000296	0.002121	20.0
Others	pet travel_agency	3	0	–	–	3	0.002898	0.002296	0.0

Abbreviations: C, count; F, false; FClsC, false closeness centrality; FDnst, false density; G, group; T, true; TA, total accuracy; TClsC, true closeness centrality; TDnst, true density; V, value.

14679671, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tgis.13089 by Yildiz Technical University, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Statistical results for "amenity"-tagged objects.

G	V	C	T	TCIsC (mean)	TDnst (mean)	F	FCIsC (mean)	FDnst (mean)	TA (%)
Bank	bank	3	0	–	–	3	0.000668	0.003348	0.0
Car Service	car_wash	1	0	–	–	1	0.000152	0.001874	0.0
Education & Learning	kindergarten library school university	52	49	0.001421	0.002591	3	0.002801	0.001998	94.2
Entertainment & Meal	cafe events_venue fast_food restaurant	37	1	0.000203	0.000991	36	0.002315	0.002262	0.03
Health	clinic dentist hospital pharmacy	227	224	0.001581	0.002563	3	0.005606	0.002494	98.7
Public Service	court_house police post_box post_office townhall	15	14	0.002331	0.002068	1	0.000499	0.001301	93.3
Religion	place_of_worship	27	25	0.001940	0.002764	2	0.000152	0.002233	92.6

Abbreviations: C, count; F, false; FCIsC, false closeness centrality; FDnst, false density; G, group; T, true; TA, total accuracy; TCIsC, true closeness centrality; TDnst, true density; V, value.

14679671, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tgis.13089 by Yildiz Technical University, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

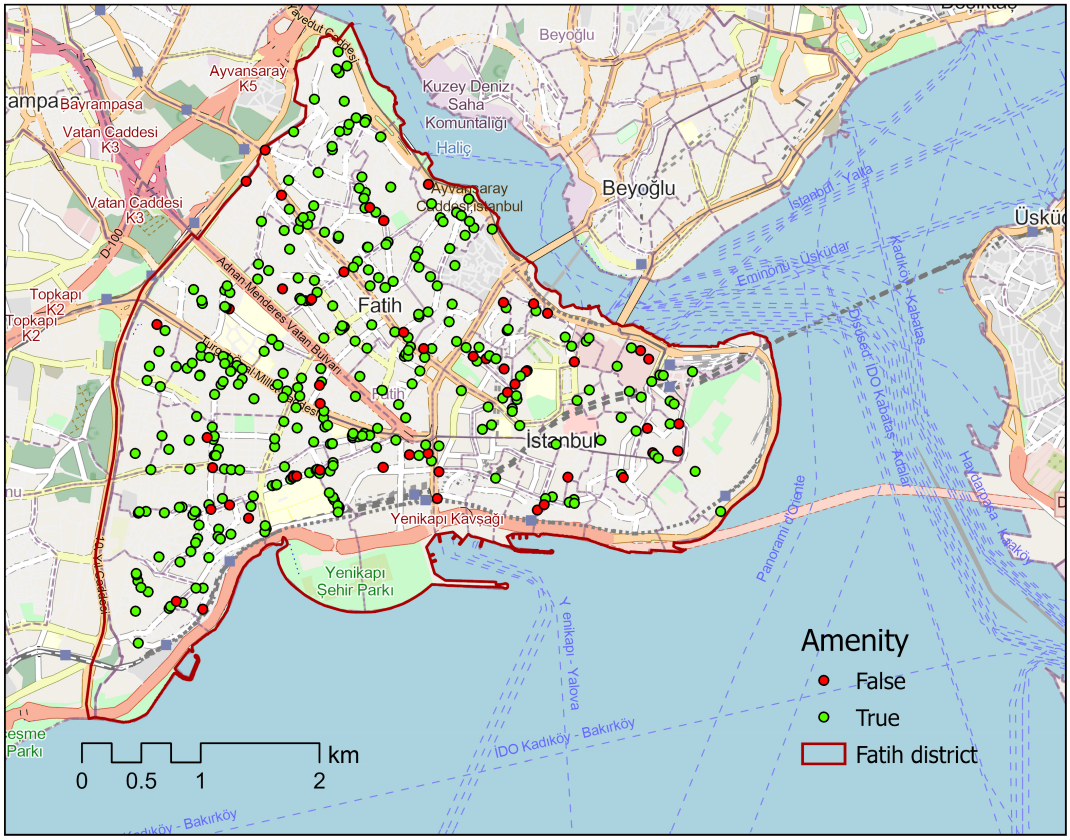

FIGURE 4 General match distribution of objects belonging to the tag "Amenity".

3.3 | Limitations of the study

This study has been conducted in the historical peninsula of Istanbul, an area characterized by its urban density and vibrant tourism activities. Considering the diversity of both OSM data and the real-world geography it represents, it is essential to acknowledge that our analysis is limited to this particular region. Nevertheless, the findings of this study undoubtedly hold relevance for other similar research areas and contexts. While the exact patterns and outcomes may vary, the insights gained from our research can provide valuable guidance for researchers examining the quality and reliability of OSM data in different urban settings.

The reverse geocoding process accesses the closest OSM object in the target data to an entry in the reference database, and the user is presented with the object's appropriate additional information, such as name, address, and tag information (https://nominatim.org/release-docs/latest/api/Reverse/). However, the OSM reverse geocoding API may not be able to accurately compute the relevant tag information for the received coordinate in certain situations. For example, if small and unnamed roads exist in the OSM database, reverse geocoding cannot be performed. Similarly, if a requested site of interest is not available in OSM, the closest incorrect result will be displayed to the user via reverse geocoding. Differences in the reference coordinate system can also limit the application of the study. For instance, if the reverse geocoding service is specified in a global coordinate system and the map service being analyzed is defined in a local coordinate system (or vice versa), almost all of the input data could be matched to incorrect map data. Lastly, the OSM object with the closest match may not always have an address that is sufficiently similar to the desired POI, especially in areas with intensive urbanization, where the user may be presented with completely different street or avenue

14675671, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tgis.13089 by Yildiz Technical University, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

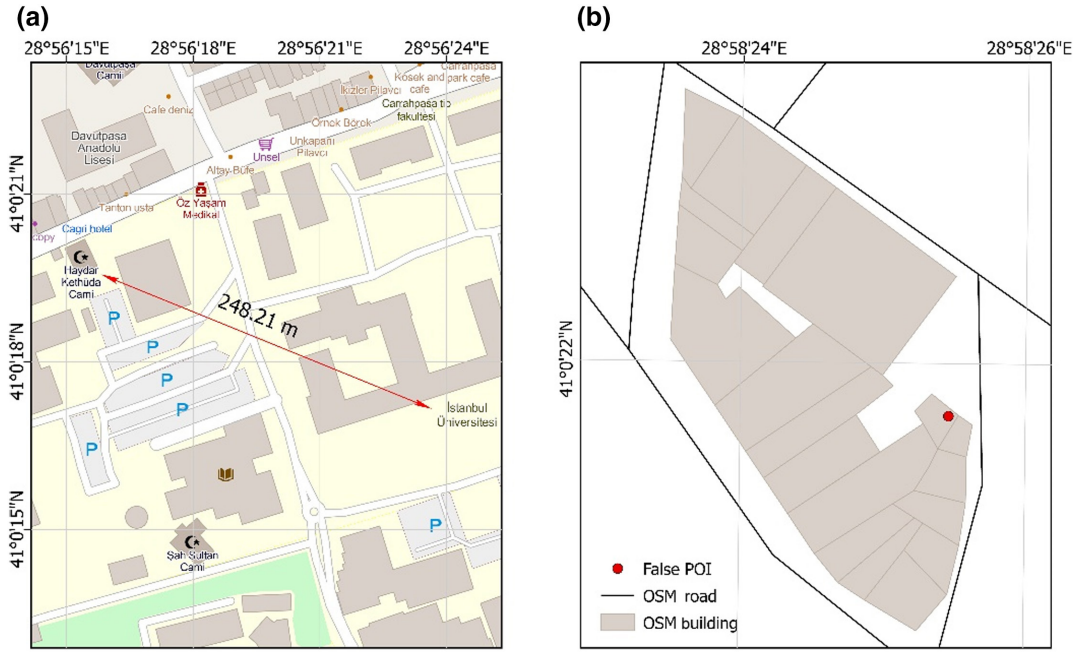

FIGURE 5 The incorrect OSM reverse-geocoding in urban areas: (a) major and (b) intrinsic errors.

information (https://nominatim.org/release-docs/latest/api/Reverse/). These circumstances make it challenging to evaluate the quality of the OSM reverse geocoding procedure. For example, instead of the intended result, Haydar Kethüda Mosque, which has the tag "facility = worship area," Istanbul University, which has the tag "facility = university" and is around 250 m away, was returned as a result of the reverse geocoding procedure (Figure 5a). This demonstration of a different result despite the presence in the reference database highlights the limitations of the used reverse geocoding service. Also, a potential intrinsic error source may stem from the discrepancy between the size of OSM building geometries and the spatial resolution of the reference base map utilized by VGI contributors. For example, certain building facades within the block in Figure 5b possess lengths less than 5 m. Should the spatial resolution of the source employed for mapping exceed this value, it could potentially lead to inaccuracies in the identification and representation of closely situated POIs.

4 | CONCLUSIONS

This study assesses the compatibility between OSM, a popular VGI project and global mapping platform, and municipal geographic data. The study uses the OSM reverse geocoding service to analyze the official coordinates of the municipality as a reference, and compare the accuracy of the address and tag information of the data that matches OSM objects.

The results of our study reveal significant inconsistencies in the OSM reverse-geocoding process. Factors such as intense urbanization and the resolution of source maps used by volunteers may contribute to the quality of reverse geocoding, directly impacting its accuracy. These factors led to discrepancies in the result of different buildings (intrinsic errors) or streets and avenues (major errors) being shown to the user. Consequently, these inconsistencies have prompted researchers to investigate the accuracy of reverse geocoding, particularly in metropolitan areas, to better understand the underlying challenges and potential improvements.

14679671, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tgis.13089 by Yildiz Technical University, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The study finds that address accuracy supports the findings of previous studies on geocoding, but notes that positional deviations of 100 m or less may not be sufficient in certain circumstances, particularly in densely populated areas. When the reference POIs correspond to objects tagged with "shop" and "amenity" in OSM, there is enough comparative data for statistical analysis. However, the study finds that the OSM reverse geocoding procedure is insufficient for almost all "shop" tags, likely due to the high density of stores in tourist areas. The same is found to be true for "amenity" tags in central regions with a high density of health services or educational institutions.

The varying quality of tag classes in OSM data necessitates a thorough quality control process to ensure the reliability of the data for various research and application purposes. Users must carefully evaluate the geometric and semantic quality of their target POIs within a representative sample of OSM data. This evaluation should cover both the accuracy and completeness of the data. Moreover, users should establish a confidence interval for their research or application, which will help them gauge the reliability of the data. They must consider whether the accuracy and completeness of the OSM data meet the requirements of their intended research or application. If the data falls short of the necessary quality standards, users should consider alternative data sources, as well as geocoders or methods (i.e., transferring, updating, and enriching the data) for improving the quality of the data.

The study also explores the relationship between municipal POIs and OSM data specific to the study region and finds that the kernel density value derived from building centers does not reflect the quality trend of the reverse geocoding process. To improve the quality of reverse geocoding in densely populated metropolitan areas, future research should identify measures that are appropriate for the city's morphology.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the Municipality of Fatih Directorate of Information Technologies for their support in providing the datasets of points of interest.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Batuhan Kilic https://orcid.org/0000-0002-0529-8569
Müslüm Hacar https://orcid.org/0000-0002-8737-8262
Fatih Gülgen https://orcid.org/0000-0002-8754-9017

REFERENCES

Amelunxen, C. (2010). An approach to geocoding based on volunteered spatial data. In A. Zipf, K. Behncke, F. Hillen, & J. Schaefermeyer (Eds.), *Geoinformatik 2010* (pp. 7–12). Die Welt im Netz.

Arsanjani, J. J., Mooney, P., Helbich, M., & Zipf, A. (2015). An exploration of future patterns of the contributions to OpenStreetMap and development of a contribution index. *Transactions in GIS*, *19*(6), 896–914. https://doi.org/10.1111/tgis.12139

Bonner, M. R., Han, D., Nie, J., Rogerson, P., Vena, J. E., & Freudenheim, J. L. (2003). Positional accuracy of geocoded addresses in epidemiologic research. *Epidemiology*, *14*(4), 408–412. https://doi.org/10.1097/01.EDE.0000073121.63254.c5

Carriere, D. (2018). *Geocoder documentation—Release 1.38.1*, *Geocoder Python package*. https://geocoder.readthedocs.io/

Clemens, K. (2015). Geocoding with OpenStreetMap data. *Seventh International Conference on Advanced Geographic Information Systems, Applications, and Services*, Lisbon, Portugal (pp. 1–2).

14679671, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tgis.13089 by Yildiz Technical University, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Corcoran, P., Mooney, P., & Bertolotto, M. (2013). Analysing the growth of OpenStreetMap networks. *Spatial Statistics*, *3*, 21–32. https://doi.org/10.1016/j.spasta.2013.01.002

Crucitti, P., Latora, V., & Porta, S. (2006). Centrality in networks of urban streets. *Chaos*, *16*(1), 015113. https://doi.org/10.1063/1.2150162

Cutini, V. (2001). Centrality and land use: Three case studies on the configurational hypothesis. *Cybergeo: European Journal of Geography*, 188. https://doi.org/10.4000/cybergeo.3936

Cutini, V. (2016). Motorways in metropolitan areas: The northwestern growth of Florence and the urban use of Motorway A1. *ISPRS International Journal of Geo-Information*, *5*(6), 77. https://doi.org/10.3390/ijgi5060077

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, *40*, 35–41. https://doi.org/10.2307/3033543

Goldberg, D. W., Wilson, J. P., & Knoblock, C. A. (2007). From text to geographic coordinates: The current state of geocoding. *URISA Journal*, *19*(1), 33–46.

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, *69*(4), 211–221. https://doi.org/10.1007/s10708-007-9111-y

Hacar, M. (2022). Analyzing the behaviors of OpenStreetMap volunteers in mapping building polygons using a machine learning approach. *ISPRS International Journal of Geo-Information*, *11*(1), 70. https://doi.org/10.3390/ijgi11010070

Hacar, M., Kılıç, B., & Şahbaz, K. (2018). Analyzing OpenStreetMap road data and characterizing the behavior of contributors in Ankara, Turkey. *ISPRS International Journal of Geo-Information*, *7*(10), 400. https://doi.org/10.3390/ijgi7100400

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. *Environment and Planning B: Planning and Design*, *37*(4), 682–703. https://doi.org/10.1068/b35097

Heipke, C. (2010). Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(6), 550–557. https://doi.org/10.1016/j.isprsjprs.2010.06.005

Kilic, B., & Gulgen, F. (2017, Kasım 18–21). Comparing the performance of reverse geocoding services: Google and Bing maps. *International Symposium on GIS Applications in Geography and Geosciences*, Çanakkale-Türkiye (pp. 155–161).

Kilic, B., & Gülgen, F. (2020a). Investigating the quality of reverse geocoding services using text similarity techniques and logistic regression analysis. *Cartography and Geographic Information Science*, *47*(4), 336–349. https://doi.org/10.1080/15230406.2020.1746198

Kilic, B., & Gülgen, F. (2020b). Accuracy and similarity aspects in online geocoding services: A comparative evaluation for Google and Bing maps. *International Journal of Engineering and Geosciences*, *5*(2), 109–119. https://doi.org/10.26833/ijeg.629381

Kounadi, O., Lampoltshammer, T. J., Leitner, M., & Heistracher, T. (2013). Accuracy and privacy aspects in free online reverse geocoding services. *Cartography and Geographic Information Science*, *40*(2), 140–153. https://doi.org/10.1080/15230406.2013.777138

Krumm, J. (2007). Inference attacks on location tracks. *5th International Conference on Pervasive Computing*, Toronto, ON, Canada (pp. 127–143).

Li, D. (2018). Geocoding and reverse geocoding. In B. Huang, T. J. Cova, & M.-H. Tsou (Eds.), *Comprehensive geographic information systems—1st edition: GIS methods and techniques* (pp. 95–109). Elsevier. https://doi.org/10.1016/B978-0-12-409548-9.09593-2

Li, D., Cova, T. J., & Dennison, P. E. (2017). Using reverse geocoding to identify prominent wildfire evacuation trigger points. *Applied Geography*, *87*, 14–27. https://doi.org/10.1016/j.apgeog.2017.05.008

Lv, M., Chen, L., Xu, Z., Li, Y., & Chen, G. (2016). The discovery of personally semantic places based on trajectory data mining. *Neurocomputing*, *173*, 1142–1153. https://doi.org/10.1016/j.neucom.2015.08.071

McKenzie, G., & Janowicz, K. (2015). Where is also about time: A location-distortion model to improve reverse geocoding using behavior-driven temporal semantic signatures. *Computers, Environment and Urban Systems*, *54*, 1–13. https://doi.org/10.1016/j.compenvurbsys.2015.05.003

Neis, P., & Zielstra, D. (2014). Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. *Future Internet*, *6*(1), 76–106. https://doi.org/10.3390/fi6010076

Neis, P., & Zipf, A. (2012). Analyzing the contributor activity of a volunteered geographic information project—The case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, *1*(2), 146–165. https://doi.org/10.3390/ijgi1020146

Ningsheng, T., Chongjun, Y., LiuZhong, Y., & Yuan, L. (2015). An address regional tessellation method for spatial subdivision and geocoding in digital earth system. *International Journal of Digital Earth*, *8*(10), 825–839. https://doi.org/10.1080/17538947.2014.915995

OSM Wiki. (2022). *OpenStreetMap Wiki—Main page*. https://wiki.openstreetmap.org/wiki/Main_Page

14679671, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tgis.13089 by Yildiz Technical University, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall, CRC Press.
Strano, E., Nicosia, V., Latora, V., Porta, S., & Barthélemy, M. (2012). Elementary processes governing the evolution of road networks. *Scientific Reports*, *2*(1), 1–8. https://doi.org/10.1038/srep00296
Strickland, M. J., Siffel, C., Gardner, B. R., Berzen, A. K., & Correa, A. (2007). Quantifying geocode location error using GIS methods. *Environmental Health*, *6*(1), 1–8. https://doi.org/10.1186/1476-069X-6-10
Travelness. (2022). What are the 20 most visited cities in the world?. https://travelness.com/most-visited-cities-in-the-world
TUİK. (2021). *Türkiye İstatistik Kurumu websitesi. Adrese Dayalı Nüfus Kayıt Sistemi Sonuçları*. https://data.tuik.gov.tr/Search/Search?text=n%C3%BCfus&dil=1
UNESCO. (2006). *Cities named '2010 European Capital of Culture' include World Heritage sites*. https://whc.unesco.org/en/news/248/
Yameqani, A. S., & Alesheikh, A. A. (2019). Evaluating a location distortion model to improve reverse geocoding through temporal semantic signatures. *Computers, Environment and Urban Systems*, *77*, 101349. https://doi.org/10.1016/j.compenvurbsys.2019.101349
Yin, Z., Goldberg, D. W., Hammond, T. A., Zhang, C., Ma, A., & Li, X. (2020). A probabilistic framework for improving reverse geocoding output. *Transactions in GIS*, *24*(3), 656–680. https://doi.org/10.1111/tgis.12623
Zandbergen, P. A. (2008). A comparison of address point, parcel and street geocoding techniques. *Computers, Environment and Urban Systems*, *32*(3), 214–232. https://doi.org/10.1016/j.compenvurbsys.2007.11.006
Zandbergen, P. A. (2009). Geocoding quality and implications for spatial analysis. *Geography Compass*, *3*(2), 647–680. https://doi.org/10.1111/j.1749-8198.2008.00205.x

How to cite this article: Kilic, B., Hacar, M., & Gülgen, F. (2023). Effects of reverse geocoding on OpenStreetMap tag quality assessment. *Transactions in GIS*, *00*, 1–15. https://doi.org/10.1111/tgis.13089

14675671, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tgis.13089 by Yildiz Technical University, Wiley Online Library on [10/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License