University of Tehran

Earth Observation and Geomatics Engineering

Online ISSN: 2588-4360

Homepage: https://eoge.ut.ac.ir/

Estimating the Quality of Areal OSM Data, A Case Study in Zanjan

Fatemeh Fathi [1], **Farhad Hosseinali** [2✉], and **Sogand Khami** [3]

1. Dept. of Surveying Engineering, Faculty of Civil Engineering, Shahid Rajaee Teacher Training University, Lavizan, Tehran, Iran. E-mail: fathi.fatemeh1997@gmail.com
2. Corresponding author, Dept. of Surveying Engineering, Faculty of Civil Engineering, Shahid Rajaee Teacher Training University, Lavizan, Tehran, Iran. E-mail: f.hosseinali@sru.ac.ir
3. Dept. of Surveying Engineering, Faculty of Civil Engineering, Shahid Rajaee Teacher Training University, Lavizan, Tehran, Iran. E-mail: sogandkhm@gmail.com

Article Info

Article type:
Research Article

Article history:
Received 2026-01-15
Received in revised form 2026-02-14
Accepted 2026-02-14
Available online 2026-06-02

Keywords:
OpenStreetMap,
Data Quality,
Volunteered Geographic Information,
Zanjan.

ABSTRACT

The employment of digital maps has become a common tool and a widespread resource and an integral aspect of today's digital lifestyle. These maps are primarily created not through official sources but rather with the involvement of the users themselves, with OpenStreetMap (OSM) being a prominent example. However, the trustworthiness of these maps has always been a significant issue.

This study focuses on how accurate and complete OpenStreetMap (OSM) data is in the city of Zanjan by comparing it to high-resolution 1:2000 reference maps. Using methods like spatial similarity and weighted averages, this research discovered that while OSM data was generally reliable, there were noticeable differences depending on the part of the city.

The results revealed that in historical and deteriorated regions, the accuracy of the data was approximately 10% lower, and their completeness was about 35% less than in other areas. This highlights a key issue: Volunteered Geographic Information (VGI) often depends on where and who contributes to mapping.

Therefore, relying solely on volunteered data can lead to uneven quality across different urban zones. This study highlights the importance of taking local context into account when utilizing VGI and considers targeted efforts to offer users indicators of information quality as crucial.

Cite this article: Fathi, F., Hosseinali, F., & khami, S. (2025). Measuring and Evaluating the Fluctuations in the Quality of Areal OSM Data in Zanjan City, *Earth Observation and Geomatics Engineering*, Volume 9, Issue 2, Pages 72-81. http//doi.org/10.22059/eoge.2026.409787.1202

Publisher: University of Tehran.

DOI: http//doi.org/10.22059/eoge.2026.409787.1202

1. Introduction

Volunteered Geographic Information (VGI) has emerged as one of the most prevalent sources of spatial data in contemporary times (Goodchild, 2007) and represents a new and affordable method for quickly and effectively gathering geospatial data through crowdsourcing (Azari Asgari & Hosseinali, 2023; Khosravi Kazazi et al., 2024; Rajabi et al., 2024). VGI involves a large volume of noisy and unstructured information from numerous participants, effectively merging with big data (Fischer, 2012). Among the several types of voluntary geographic information, OpenStreetMap (OSM) stands out as the most widely utilized source. Initiated in 2004 at University College London, by Steve Coast, OSM gives all Internet users the opportunity to download spatial data without any costs or fees and to use it for their own projects. The goal of the OSM community is to create a map of the world that will contain as much detailed information as possible, and this information is being collected by volunteers (Zielstra and Zipf, 2010). This widespread use has led to growing interest in evaluating the quality of OSM data. The primary concern regarding VGI in the literature is data quality (Khosravi Kazazi et al., 2024). In the OSM project, numerous doubts have arisen regarding the quality and reliability of its data (Goodchild, 2007). The OSM geometric and descriptive attributes consistently are evolving. This process of data refinement exhibits spatial heterogeneity, leading to variations in data quality across different areas (Raifer, et al., 2019).

Evaluating the quality based on the principles of spatial data quality defined according to ISO/TC 21:19113 standards includes five standard criteria: completeness (The degree of presence of the features in a dataset compared to the real world (Zhou, 2018; Brovelli and Zamboni, 2018)), logical consistency (the degree of compliance and consistency with the logical rules of data structure, descriptive information, and relationships), positional accuracy (the absolute or relative deviation of the coordinates of a feature from its real-world location (Zhang and Malczewski, 2019)), temporal accuracy (the accuracy of temporal relationships and descriptions of features), and thematic accuracy (the quantitative and qualitative description, classification, and relationships of features) (Model, 2006). Since the spatial position of features (positional accuracy) affects all other quality components and plays a crucial role in ensuring the overall quality of spatial data, it holds significant importance in various applications.

When the quality of OSM data is analyzed, many researchers mainly focus on completeness (Tian et al., 2019; Wang et al., 2020; Zhang et al., 2022) and positional accuracy (Brovelli and Zamboni, 2018; Zhou and Jing, 2022) of features such as roads (Haklay, 2010) or buildings (Hecht et al., 2013). The completeness, a measure of how well a region has been mapped, is viewed as the most important quality measure because the other measures are assessed based on existing OSM data, which most studies have assessed OSM data quality by comparing it with a reference dataset (e.g. acquired from either a mapping agency or a commercial company) (Zhou et al., 2022). For instance, Brovelli et al. proposed a map matching method to examine both the completeness and spatial accuracy of OSM building data, which was based on a comparison with the regional topographic database of Italy (Brovelli and Zamboni, 2018). Similarly, Törnros et al. estimated the completeness of buildings in OSM for a medium-sized German city and its surroundings by comparing the OSM data with data from an official cadastral building. They compared two measures of completeness: the count ratio (the ratio of OSM buildings to reference buildings) and an area ratio (the ratio of total OSM building’s area to total reference building’s area). Their results indicated that the count ratio tends to underestimate completeness in the study area, Whereas the area ratio tends to overestimate it (Törnros et al., 2015).

Previous studies addressing the quality, specifically the positional accuracy of VGI in relation to OSM, have pointed to non-homogeneity (Zhang et al., 2022; Tian et al., 2019). To understand these issues, researchers have investigated the relationship between volunteered geographic data and its quality using various geographical and demographic indicators. For example, the research on evaluating the quality of OSM roads and buildings in the Québec province of Canada. The results indicate that the quality of OSM data varies across different regions of Québec, and factors such as population density, income level, and distance from the city center can influence the quality of this data (Moradi, 2020). Studies that have addressed the relationship between population density and VGI have shown that in densely populated urban areas, more contributions can be expected, which may ultimately lead to higher data quality (Tian et al., 2019). Zhang et al. in 2022 examined the quality parameter of OSM data completeness. They used various approaches (type-based and regression-based) to estimate building completeness from global population data at varying spatial resolutions. They concluded that the use of population data as referenced building data is an effective method for assessing OSM building completeness (Zhang et al., 2022). However, certain demographic groups are more likely to engage with VGI. Based on research by Azari Asgari and Hosseinali, age and familiarity with region have significant effects on the accuracy of participant’s responses (Azari Asgari and Hosseinali, 2023). For instance, older individuals are less likely to engage in social media activities (Haworth et al., 2015). Jokar Arsanjani and Bakillah conducted a study aimed at identifying potential socio-economic characteristics of contributors in areas with high participation levels to gain deeper insights into the behavioral patterns of users in Germany—the most active OSM country. They showed, areas with high population density, middle level of education, high income, high rate of overnight stays, high number of foreigners, and residents aged 18 to 69 are more likely to participate in OSM.

Furthermore, they noted that the degree of dynamism in OSM is a function of proximity to built-up areas (Jokar Arsanjani and Bakillah, 2015). Moreover, factors such as ethnicity, race, religion, and gender have been found to affect participation levels (Nicolosi et al., 2020; Gardner and Mooney, 2018). In this context, a recent study by Shin et al. indicated that younger and male participants tend to be more active spatially and temporally on platforms such as OSM, potentially introducing biases in spatial coverage and data quality across regions (Shin et al., 2025). Similarly, Scholz et al. evaluated the potential of OSM data for analyzing the vulnerability and exposure of at-risk populations in Sudan, particularly Internally Displaced Persons (IDPs) and refugees exposed to climate-related hazards. With the escalation of conflict in 2023, OSM mapping activities increased, but challenges such as slow updates of satellite imagery and a lack of spatial data still persist. Their findings suggest that OSM can serve as a valuable tool for identifying high-risk areas and addressing data gaps, if connections between data producers and users are improved and local community engagement (especially among women) is enhanced (Scholz et al., 2024). Another study by Moradi et al. aimed to assess the quality of OSM roads in the Québec province. This study examined three quality elements namely: completeness, positional accuracy, and attribute accuracy. Then, it analyzed their relationship with five quality indicators: population, average income, OSM road density, OSM building density, and the number of POIs (Points of Interest). The results showed that population and road density were the best indicators for completeness, while population and income were the best indicators for positional and attribute accuracy. Except for two indicators, OSM building density and number of POIs, all indicators showed significant correlations with the quality elements (Moradi et al., 2022). However, most of these studies primarily focused on demographic, social, and economic indicators, while fewer analyses have been conducted regarding geographic indicators and urban regions.

Overall, the prevailing consensus in existing research indicates that the quality of OSM data displays significantly different patterns across various regions, influenced by demographic and socio-economic factors (Herfort, et al., 2023) as well as indicators pertaining to participant behavior (Solomon, et al., 2024). In general, previous studies have mainly focused on evaluating the total quality of OSM data at the city or national level, analyzing it as a single unit. However, few studies have examined variations in data quality across different urban regions. This is despite the fact that cities exhibit significant spatial heterogeneity, and factors such as physical structure, building density, land-use type, and status of urban renewal can directly influence the accuracy and completeness of OSM data.

This study aims to examine the relationships between OSM data quality (positional accuracy and completeness) within different urban regions. We conduct this analysis through a case study of Zanjan, one of the historical cities of Iran. To achieve this, official data at a scale of 1:2000 is employed to evaluate the quality of OSM, and various quality elements are examined. The central question posed is whether the quality of OSM data is consistent throughout the city. To address this, the methodology is applied across different city regions to uncover potential variations in the quality of OSM data.

The remainder of this article is organized as follows: Methodology section presents the rationale and approach used to assess data quality and completeness across different urban regions; Results section reports the findings of the analysis; and Conclusion section provides a comprehensive discussion and summary of the study.

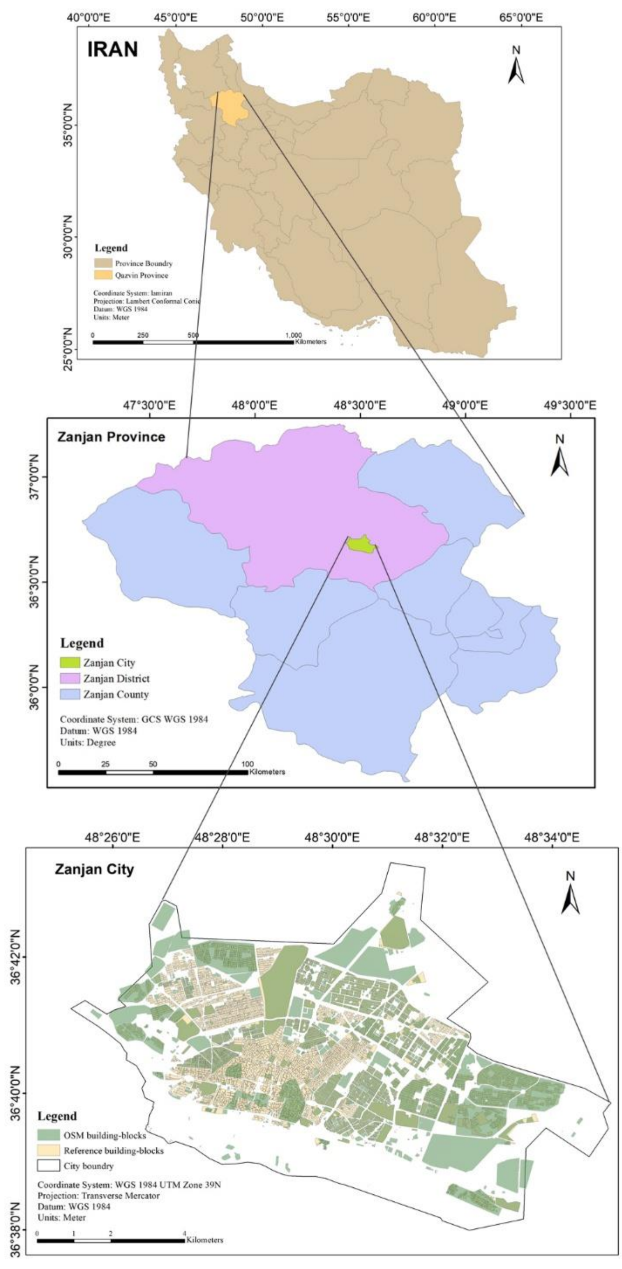


Figure 1. Zanjan city position with its reference and OSM city blocks.

2. Methodology

2.1. Data Collection

The study area of this research is the city of Zanjan (Figure 1), which has a population of approximately 500,000. Zanjan is a city located in northwestern Iran and serves as the capital of Zanjan Province. Zanjan's urban structure consists of a combination of densely built-up central areas and developed peripheral zones. This city was selected due to the accessibility of essential data along with existence of historical areas in the city that separates it from other parts of the city. Furthermore, being a medium-sized city allows for the easier and faster implementation of algorithms across the entire urban region. The city's diverse spatial structure, in terms of urban development patterns, access networks, and differences in building densities, has provided the opportunity to analyze the quality of OSM data under various spatial conditions.

In this study, both OSM data and reference (official) data are required. Reference data at a scale of 1:2000 was sourced from the National Cartography Centre of Iran. The reference maps used in this study were up to date at the time of data collection. This dataset includes urban blocks and land uses. Additionally, the historical and deteriorated areas of the city were acquired from the Zanjan Municipality which have been separately used for quality analysis in specific areas. OSM data of 2025 was retrieved from the OSM website and include building polygon information within the urban region of Zanjan. Given the primary objective of this study—to evaluate the quality of polygonal data—point and line layers have been removed, and only surface geometries (polygons) have been analyzed.

2.2. Methods and Implementation

Zanjan, with its diverse physical structure (including historical and deteriorated areas) provides a suitable context for analyzing OSM data quality across different urban regions (Figure 2).

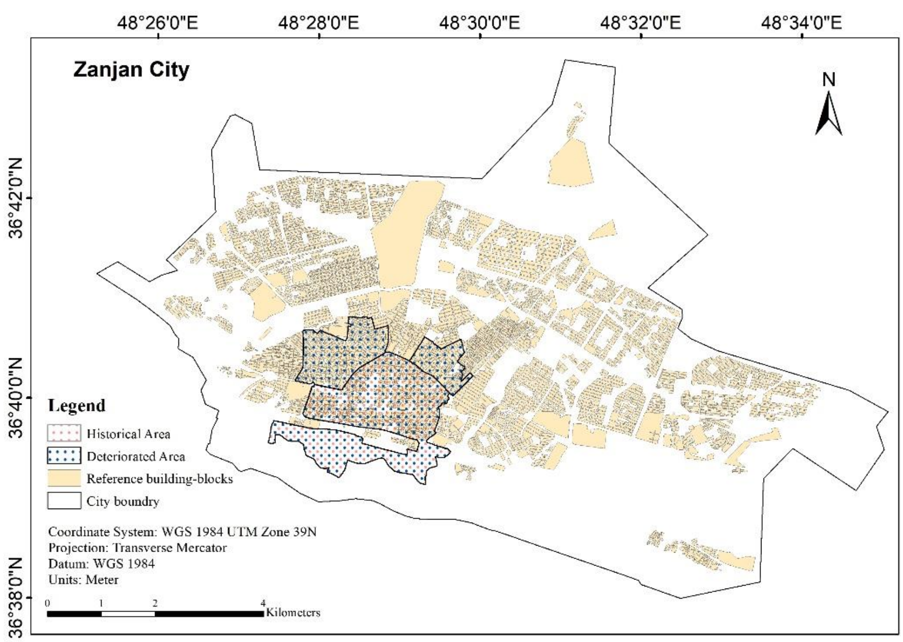


Figure 2. Zanjan city reference blocks and its historical and deteriorated areas.

Prior to any analysis, both datasets preprocessed. Data preprocessing involved the correction of topological errors, extraction of reference and OSM blocks within the designated boundary of each city, and harmonization of their coordinate systems. This step is a critical step to ensure the accuracy and integrity of the analysis. The first step for evaluating OSM data quality is correspondence between the features of the OSM database and the reference data. Researchers commonly use two main polygon matching algorithms: 1) The buffer intersection method and 2) The centroid comparison method. While these methods are effective for the majority of OSM building footprints, they may not achieve high accuracy in complex situations. (Moradi et al., 2023). Therefore, in this research, in addition to the main matching algorithms mentioned, geometric criteria were also employed to examine the correspondence between polygons in the OSM and reference datasets. Accordingly, a comprehensive algorithm, as shown in Figure 3, developed to determine the accuracy of the corresponding data. The importance of this algorithm lies in the fact that by performing the necessary preprocessing steps and having access to reference data, this algorithm can be applied to any city.

In this research, features derived from OSM data that fall within specified limits are identified as candidate features.

This limit serves as a coefficient for the positional accuracy of the reference. To find the matched candidate for each feature from the reference dataset, OSM data that exhibit less than 90% overlap with the reference are evaluated using similarity indicators, including: Overlap Percentage, Area Differences, Shape Orientation, Shape Complexity and Centroids Distance. In other words, in this stage, the degree of spatial similarity between polygon features is calculated based on specific similarity measurement criteria. During this stage, the weight of each polygon is computed, and if its value exceeds the threshold, it is designated as almost identical feature. The overlap and the threshold value which were 90% and 70%, respectively, were determined based on previous studies and further validated through performance evaluation of the algorithm in additional cities.

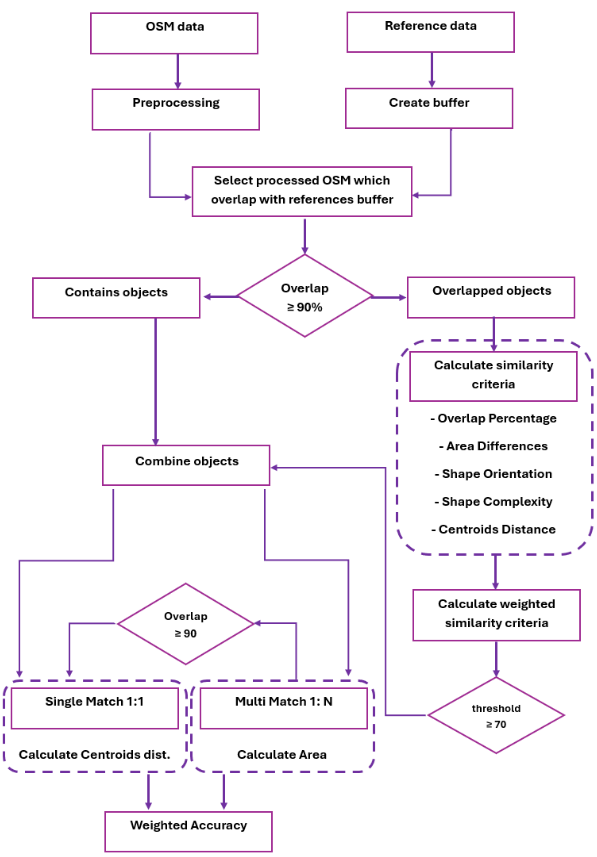


Figure 3. The methodology of the research.

Similarity indices are employed to evaluate the degree of likeness between two spatial features. The combined use of these indices has previously been employed by Küçük and Anbaroğlu (2020) and Naumann et al. (2025) for building polygon matching and evaluating the spatial quality of VGI data and representing a common approach in the analysis of OSM data quality. In fact, each index describes one aspect of the differences or similarities between spatial shapes, allowing for a more precise analysis of VGI data quality. The first index, overlap percentage, is a basic measure that represents the ratio of the intersecting area between two polygons to the total area of the smaller polygon and serves as one of the primary indices for identifying corresponding polygon pairs.

$$s_{overlap} = \frac{Area(ref \cap osm)}{\min(Area(ref), Area(osm))} \tag{1}$$

In Equation 1, $S_{overlap}$ is the normalized similarity degree based on the overlap area, $Area(ref \cap osm)$ is the intersection area between the reference and OSM polygons, and $min(\mathrm{Area(ref)}, \mathrm{Area(osm)})$ is the minimum value between the areas of the two polygons (reference and OSM) (Chamani et al., 2018).

The second index is Area Difference, which measures the degree of disparity in area between the two features and is effective for detecting non-correspondence in the size of the polygons. In Equation 2, $S_{area\ difference}$ is the normalized similarity degree based on the area difference, $Area(abs(ref, osm))$ is the absolute difference in area between the reference and OSM polygons, and $\max(Area(abs(ref, osm)))$ is the maximum absolute area difference between the reference and OSM polygons in the entire dataset.

$$s_{area difference} = 1 - \frac{Area(abs(ref, osm))}{\max\left(Area(abs(ref, osm))\right)} \tag{2}$$

The third index is Shape Orientation. The directional measure of a linear feature can be expressed by the angle of a line segment joining start and end points. The areal feature's directional measure is calculated using a minimal boundary rectangle (MBR). Directional similarity of two linear features is usually measured by comparing the angles of the diagonal lines of the two MBRs (Tong et al., 2009). according to Figure 4, it is defined as the angle formed between the diagonal of the MBR and its horizontal axis.

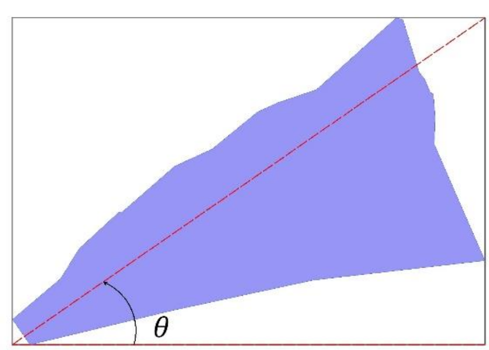


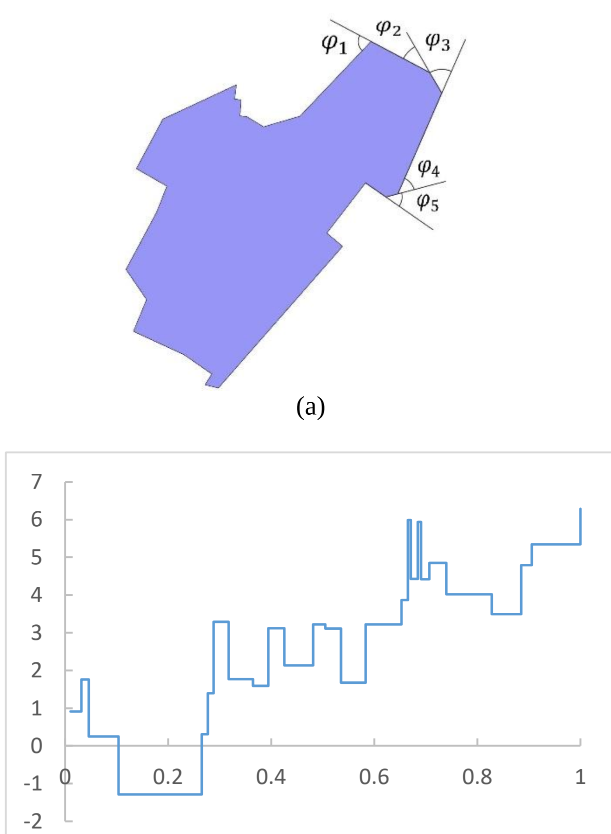
Figure 4. MBR and Polygon Orientation.

In Equation 3, $S_{shape\ orientation}$ is the normalized similarity degree based on the directional difference. Also, θ_{ref} and θ_{osm} are the orientation of the reference polygon and the orientation of the OSM polygon, respectively (Chamani et al., 2018).

$$s_{shape\ orientation} = 1 - \frac{\left|\theta_{ref} - \theta_{osm}\right|}{\frac{\pi}{2}} \tag{3}$$

The fourth index is the Shape Complexity Function, which reflects the difference in the overall geometry of each feature. Figure 5 illustrates the shape complexity function of a polygon.

Arkin et al. (1991) introduced turning function (or Shape Complexity in this paper) to represent polygons. As shown in Figure 5(a), the tangent angle at the starting vertex is $\theta_1 = \varphi_1$. Then θ_i can be calculated as $\theta_i = \theta_{i-1} + \varphi_i$. Therefore, in a clockwise direction, the cumulative angle increases when turning to the right and decreases when turning to the left. Figure 5(b) shows the change of tangent angles (y-axis) along the normalized accumulated length of the polygon sides (x-axis). From this perspective, the tangent angle can be treated as a function of the normalized accumulated length $sc(l)$ (Arkin et al., 1991; Fan et al., 2014).



(a)

(b)

Figure 5. The shape Complexity Function of the Polygon. (a) Polygon (b) Graph of the shape Complexity Function for Polygon a.

$$d_{sc}(ref, osm) = \left\| SC_{ref} - SC_{osm} \right\| = \left(\int_0^1 \left| sc_{ref}(l) - sc_{osm}(l) \right|^p dl \right)^{\frac{1}{p}} \tag{4}$$

$$s_{shape\ complexity} = 1 - \frac{d_{SC}(ref, osm)}{U} \tag{5}$$

In Equations 4 and 5, $d_{SC}(ref, osm)$ is the value of the complexity function for the two polygons, $S_{shape\ Complexity}$ is the normalized similarity degree based on the complexity function, SC_{ref} and SC_{osm} are the complexity function of the reference polygon and the complexity function of the OSM polygon, respectively, and U is the maximum difference between the two complexity functions, which is used as the normalization factor for the equation. The operator $\|.\|_p$ indicates the L_p-norm, which is typically calculated for p=2 (Chamani et al., 2018).

Finally, the fifth index, Centroids Distance, is a measure for assessing the relative position of features, which is commonly used as a complementary criterion for assessing spatial pairs. In this Equation, $S_{centroids\ distance}$ is the normalized similarity degree based on the distance between the centroids, and $p_1{:}(x_1, y_1)$ and $p_2{:}(x_2, y_2)$ are the centroid of the reference polygon and the centroid of the OSM polygon, respectively.

$$S_{centroids\ distance} = 1 - \frac{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}{U} \tag{6}$$

In Equation 6, U serves as the normalization factor, ensuring that the similarity based on Euclidean distance falls within the range of zero to one (Chamani et al., 2018). In other words, U is equal to the maximum possible distance that two polygons in a dataset can have.

Up to this point, for the features with less than 90% overlap, the similarity degree of each OSM polygon is calculated using the defined similarity indices. In fact, the similarity degree of each polygon represents a weighted combination of the employed similarity indices. According to the research methodology diagram in Figure 3, any polygon with a degree of similarity higher than 70% is recognized as corresponding features at this step. These corresponding features, which were selected from features with less than 90% overlap, are then merged with the features that had more than 90% overlap for the final analysis. After the merging process, the corresponding relationships between the reference and OSM polygons are examined. At this step, as shown in Figure 3, the polygons are classified into two categories: 1) one-to-one Relationship: Where each OSM polygon corresponds to exactly one reference polygon. 2) one-to-many

Relationship: Where multiple OSM polygons correspond to a single reference polygon.

In this step, prior to calculating accuracy, one-to-many relationships must be examined. If the total overlap area of OSM polygons exceeds 90% of the area of their corresponding reference polygon, those OSM polygons are merged and added to the one-to-one category. Subsequently, to assess regional accuracy, centroid-to-centroid distances are computed for polygons exhibiting a one-to-one correspondence, and the resulting value is considered as the similarity measure (s_i) for this case. For polygons involved in one-to-many relationships, the total overlap area is determined; in this case, the overlap area is likewise regarded as the similarity measure (s_i). The obtained similarity values are normalized according to Equation 7. Finally, the regional accuracy is evaluated based on the weighted average of these values, as presented in Equation 8. In these equations, i is OSM polygon's ID number, n is the number of polygons in region, w_i is the area of each OSM polygon, A is the OSM polygon, B is Candidate polygon in reference dataset matched with OSM polygon.

$$S_i(A,B) = \begin{cases} \dfrac{Overlap\ Area}{reference\ Area}, & For\ Multimatch\ category \\ \dfrac{Centroids\ Distance}{Biggest\ Centroids\ Distance} & For\ Single\ match\ category \end{cases} \tag{7}$$

$$Accuracy = \frac{\sum_{i=1}^{n} w_i . s_i(A,B)}{\sum_{i=1}^{n} w_i} \tag{8}$$

This study also examines the completeness of OSM data. Accordingly, the completeness criterion was evaluated in terms of both the number of features (the ratio of the number of OSM data to the reference data) and their area (the ratio of the OSM–reference overlap area to the reference area). Based on the defined relationship, it is expected that completeness, whether measured by count or by area, yields values ranging between 0 and 1. Completeness based on the number and area is calculated based on following Equations:

$$c_{count} = \frac{N_{osm}}{N_{ref}} \tag{9}$$

$$c_{area} = \frac{Area_{osm}}{Area_{ref}} \tag{10}$$

In Equation 9, C_{count} is the ratio of the total number of OSM data to the total number of reference data, and In Equation 10, C_{area} is the ratio of the overlapping area between OSM and reference data to the total area of the reference data (Chamani et al., 2018). The closer the value is to one, the higher the completeness of the OSM data. Values greater than one indicate that the number of OSM polygons is higher and they cover a larger area compared to the reference dataset.

3. Results

In this research, the quality of OSM data was evaluated in the city of Zanjan and urban regions including the historical, deteriorated districts, and areas outside these boundaries. The analyses were performed based on two main criteria: "Completeness" (in terms of area and number) and "Positional Accuracy". The results clearly indicated that the quality of OSM data is not homogeneous across the city and shows varying spatial patterns in different urban districts. The results are shown in Figure 6.

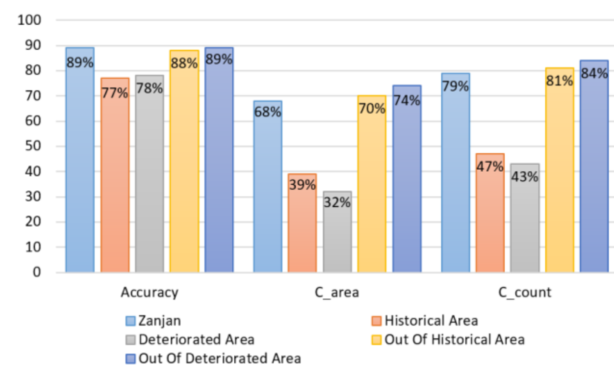


Figure 6. The differences in accuracy and completeness criterion of OSM across the regions of the study area.

The positional accuracy of OSM features for the entire city of Zanjan was estimated to be 89%, while this value decreased to 77% and 78% in the historical area and deteriorated area, respectively. This decrease in accuracy indicates challenges associated with feature matching process and the extraction of precise boundaries in older areas characterized by irregular urban structures. These findings align with previous studies. As noted by Chen et al. (2025), OSM datasets tend to be more complete and of higher quality in developed areas (Chen et al., 2025). This point suggests that the lower quality of OSM data in the historical and deteriorated areas of Zanjan is not a local exception, but rather part of a general pattern in VGI data.

In terms of completeness (both in feature count and total area) significant variations were observed across urban regions. The area-based completeness for the entire city was 68%, while this value dropped to 39% and 32% in the historical and deteriorated areas, respectively. In terms of count, city-wide completeness was 79%, while this index was calculated as 47% in the historical area and 43% in the deteriorated area. These figures indicate that the OSM coverage in these areas is significantly less complete compared with the entire city.

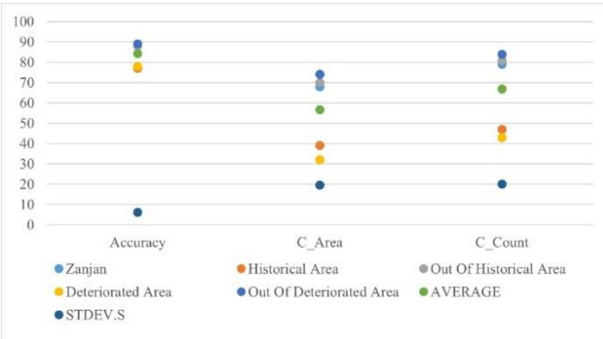


Figure 7. Positional accuracy and completeness assessment results for Zanjan and its urban regions.

Statistical analyses, as shown in Figure 7, indicated that the Standard Deviation (SD) for the accuracy is approximately 6%, which signifies a relatively low dispersion of data in the accuracy index. In contrast, the completeness index had a higher standard deviation, approximately 19% to 20%, which suggests a noticeable difference in data coverage across urban regions. This finding suggests that the primary issue in OSM data quality lies more in the level of completeness than in positional accuracy.

Furthermore, the ratio of accuracy to the number of OSM features was analyzed in Table 1 to gain a more precise understanding of the relationship between positional accuracy and completeness. The results for the historical area showed that despite maintaining a level of accuracy, the data are deficient in terms of completeness. This issue indicates that the presence of desirable accuracy does not necessarily mean the sufficiency or complete coverage of information within that context, and thus, quality indicators must be evaluated in a multiple and compound manner.

Table 1. Ratio of Positional Accuracy to the Number of OSM Features in each Area.

	(Accuracy/Count OSM) *100
Zanjan	0.033
Historical Area	0.74
Out Of Historical Area	0.033
Deteriorated Area	0.397
Out Of Deteriorated Area	0.035

The analysis confirms that achieving a satisfactory level of positional accuracy alone cannot ensure the adequacy of spatial information, and that quality indicators must be examined in an integrated manner. Furthermore, this study demonstrates that the quality of OSM data across different urban regions can be influenced by the spatial, historical, and physical characteristics of urban regions. Ignoring these differences when utilizing VGI may lead to analytical errors. These findings can effectively assist researchers and urban planners in the correct selection and application of volunteered data.

4. Conclusion

This research assessed the quality of OSM data in the city of Zanjan by comparing it with reference data at a scale of 1:2000. The findings indicated that the employed methodology effectively captured geometric discrepancies through the use of spatial similarity indices and weighted average calculations. The results revealed that in historical and deteriorated regions, the accuracy of the data was approximately 10% lower, and their completeness was about 35% less than in other areas.

Overall, the results show that the reliability of OSM data is not consistent across different urban regions. The main challenge lies in data completeness rather than positional accuracy, indicating the unequal participation of contributors in different areas. Collectively, the findings emphasize that the use of VGI such as OSM, without considering the unique spatial and social characteristics of each locality, may lead to incomplete or incorrect interpretations. Therefore, future studies need to conduct more localized analyses in assessing the quality of this type of data and utilize the integration of multiple indicators for more accurate decision making.

References

Arkin, E. M., Chew, L. P., Huttenlocher, D. P., Kedem, K., & Mitchell, J. S. (1991). An Efficiently Computable Metric for Comparing Polygonal Shapes. https://doi.org/10.1109/34.75509

Azari Asgari, E. & Hosseinali, F. (2023). Evaluating the VGI Users' Level of Expertise: An Application of Statistical and Artificial Neural Network Approaches. International Journal of Applied Geospatial Research (IJAGR), 14, 1–16. https://doi.org/10.4018/IJAGR.316770

Brovelli, M. A. & Zamboni, G. (2018). A New Method for the Assessment of Spatial Accuracy and Completeness of OpenStreetMap building footprints. ISPRS International Journal of Geo-Information, 7, 289. https://doi.org/10.3390/ijgi7080289

Chamani, M., Abbaspour, R. A., & Chehreghan, A. R. (2018). Matching of Polygon Objects Based on Geometric Measures in a Multi-scale Dataset. Journal of Geomatics Science and Technology, 7(4), 73–87. http://jgst.issgeac.ir/article-1-646-en.html

Chen, S., Liu, L., Li, K., Ding, X. & Jiang, W. (2025). Simulation and Prediction of the Expansion of OpenStreetMap Building Data Based on the Markov-FLUS Model in Shenzhen, China. Geocarto International, 40, 2459109. https://doi.org/10.1080/10106049.2025.2459109

Fan, H., Zipf, A., & Fu, Q. (2014). Estimation of building types on OpenStreetMap based on urban morphology analysis. In Connecting a digital Europe through location and place. Springer. 19–35. https://doi.org/10.1007/978-3-319-03611-3_2

Fischer, F. (2012). VGI as Big Data. GeoInformatics, 15, 46.

Gardner, Z. & Mooney, P. (2018). Investigating gender differences in OpenStreetMap Activities in Malawi: a small case study. AGILE, Lund, Sweden, 17–20.

Goodchild, M. F. (2007). Citizens as Sensors: The World of Volunteered Geography. GeoJournal, 69, 211–221. https://doi.org/10.1007/s10708-007-9111-y

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, 682–703. https://doi.org/10.1068/b35097

Haworth, B., Bruce, E. & Middleton, P. (2015). Emerging technologies for risk reduction: assessing the potential use of social media and VGI for increasing community engagement. The Australian Journal of Emergency Management, The, 30, 36–41. https://search.informit.org/doi/abs/10.3316/informit.365583026501615

Hecht, R., Kunze, C. & Hahmann, S. (2013). Measuring completeness of building footprints in OpenStreetMap over space and time. ISPRS International Journal of Geo-Information, 2, 1066–1091. https://doi.org/10.3390/ijgi2041066

Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J. & Zipf, A. (2023). A spatio-temporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap. Nature Communications, 14, 3985. https://doi.org/10.1038/s41467-023-39698-6

Jokar ArsanJani, J. & Bakillah, M. (2015). Understanding the potential relationship between the socioeconomic variables and contributions to OpenStreetMap. International Journal of Digital Earth, 8, 861–876. https://doi.org/10.1080/17538947.2014.951081

Khosravi Kazazi, A., Rahmani, Y., Amiri, F. & Hosseinali, F. (2024). A School of Thought on VGI Challenges: A Literature Review. Papers in Applied Geography, 10, 53–68. https://doi.org/10.1080/23754931.2023.2256344

Model, C. D. (2006). ISO/TC 211 Geographic Information/Geomatics. Cite seer.

Moradi, M. (2020). Evaluating the quality of OSM roads and buildings in the Quebec Province. https://hdl.handle.net/20.500.11794/67232

Moradi, M., Roche, S. & Mostafavi, M. A. (2022). Exploring five indicators for the quality of OpenStreetMap road networks: A case study of Quebec, Canada. Geomatica, 75, 178–208. https://doi.org/10.1139/geomat-2021-0012

Moradi, M., Roche, S. & Mostafavi, M. A. (2023). A Novel Feature Matching Method for Matching OpenStreetMap Buildings with Those of Reference Dataset. International Symposium on Web and Wireless Geographical Information Systems, Springer, 139–152. https://dl.acm.org/doi/abs/10.1007/978-3-031-34612-5_10

Nicolosi, E., Medina, R., Riley, C. & McNeally, P. (2020). Crowdsourcing sensitive VGI: Constructing the hate incident reporting system. Digital Geography and Society, 1, 100003. https://doi.org/10.1016/j.diggeo.2020.100003

Raifer, M., Troilo, R., Kowatsch, F., Auer, M., Loos, L., Marx, S., Przybill, K., Fendrich, S., Mocnik, F.-B. & Zipf, A. (2019). OSHDB: A Framework for Spatio-Temporal Analysis of OpenStreetMap History Data. Open Geospatial Data, Software and Standards, 4, 1–12. https://doi.org/10.1186/s40965-019-0061-3

Rajabi, F., Hosseinali, F. & Rabiei-Dastjerdi, H. (2024). An Examination and Analysis of the Clustering of Healthcare Centers and Their Spatial Accessibility in Tehran Metropolis: Insights from Google POI data. Sustainable Cities and Society, 117, 105845. https://doi.org/10.1016/j.scs.2024.105845

Scholz, C., Jaime, C., Raju, E., Coughlan de Perez, E. & Van Aalst, M. (2024). Off the grid: Utilizing OpenStreetMap for Early Warning and Early Action in Conflict Settings in Sudan. Frontiers in Climate, 6, 1439940. https://doi.org/10.3389/fclim.2024.1439940

Shin, H., Gardner, Z., Solomon, G. & Basiri, A. (2025). Diagnosing Spatial and Temporal Biases of OSM Contributors: Identifying Differences Between Gender and Age from an Online Survey. Annals of the American Association of Geographers, 115, 782–802. https://doi.org/10.1080/24694452.2024.2447507

Solomon, G., Sutton, D., Polat Kayali, M., Yuan, X., Gardner, Z. & Basiri, A. (2024). Evaluating geotemporal behaviors of OpenStreetMap contributors. AGILE: GIScience Series, 5, 14. https://doi.org/10.5194/agile-giss-5-14-2024

Tian, Y., Zhou, Q. & Fu, X. (2019). An analysis of the evolution, completeness, and spatial patterns of OpenStreetMap building data in China. ISPRS International Journal of Geo-Information, 8, 35. https://doi.org/10.3390/ijgi8010035

Tong, X., Shi, W., & Deng, S. (2009). A Probability-based Multi-Measure Feature Matching Method in Map Conflation. International Journal of Remote Sensing, 30(20), 5453–5472. https://doi.org/10.1080/01431160903130986

Törnros, T., Dorn, H., Hahmann, S. & Zipf, A. (2015). Uncertainties of completeness measures in OpenStreetMap–A case study for buildings in a medium-sized German city. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2, 353–357. https://doi.org/10.5194/isprsannals-II-3-W5-353-2015

Wang, S., Zhou, Q. & Tian, Y. (2020). Understanding completeness and diversity patterns of OSM-based land-use and land-cover datasets in China. ISPRS International Journal of Geo-Information, 9, 531. https://doi.org/10.3390/ijgi9090531

Zhang, H. & Malczewski, J. (2019). Quality evaluation of volunteered geographic information: The case of OpenStreetMap. Crowdsourcing: Concepts, Methodologies, Tools, and Applications, 1173–1201. https://doi.org/10.4018/978-1-5225-8362-2.ch058

Zhang, Y., Zhou, Q., Brovelli, M. A. & Li, W. (2022). Assessing OSM building completeness using population data. International Journal of Geographical Information Science, 36, 1443–1466. https://doi.org/10.1080/13658816.2021.2023158

Zhou, Q. (2018). Exploring the relationship between density and completeness of urban building data in OpenStreetMap for quality estimation. International Journal of Geographical Information Science, 32, 257–281. https://doi.org/10.1080/13658816.2017.1395883

Zhou, Q. & Jing, X. (2022). Evaluation and comparison of open and high-resolution LULC datasets for urban blue space mapping. Remote Sensing, 14, 5764. https://doi.org/10.3390/rs14225764

Zhou, Q., Zhang, Y., Chang, K. & Brovelli, M. A. (2022). Assessing OSM building completeness for almost 13,000 cities globally. International Journal of Digital Earth, 15, 2400–2421. https://doi.org/10.1080/17538947.2022.2159550

Zielstra, D. & Zipf, A. (2010). A comparative study of proprietary geodata and volunteered geographic information for Germany. 13th AGILE International Conference on Geographic Information Science, Guimarães, Portugal, 1–15.