Assessing completeness of global airport data in OSM

Yijun Chen, Zheng Wei, Qi Zhou

Abstract: Airports are not only important infrastructure for both civil and military use but also have significant impacts on socio-economic development and the built-up environment. OpenStreetMap (OSM) can be an essential data source for acquiring various airport elements, but few studies have investigated data quality. To fill this gap, this study aims to assess the quality (especially completeness) of airport data in OSM by comparing it with locations of airports acquired from the OurAirports platform. More precisely, the three different types (large, medium, and small) and the four different elements (runway, taxiway, apron, and terminal) of airports are assessed for over 40,000 airports worldwide. Results show that completeness varies depending on types, elements, and geographical regions. Specifically, 1) almost all large airports are complete; most medium airports are also complete; but most small airports are not complete. 2) The runway element is much more complete than the terminal element. 3) In most cases, completeness is relatively high in India, China, and Northern Africa but relatively low in Canada, the United States, Russia, and Australia, where the total number of airports is much larger. We conclude that most large and medium airports in OSM have been mapped well. The reasons for incomplete airport data in OSM and potential applications of OSM airport data are also discussed.

Keywords: OpenStreetMap, OurAirports, Data quality, Aerodrome

1. Introduction

An airport, also known as an aerodrome, is a location for the takeoff and landing of aircraft and is considered an important infrastructure for both civil and military use (Fernandes et al. 2014; Senmarougli and Celebi 2018). Many studies have focused on airports because they not only promote socio-economic development (Freestone and Baker 2011; Fernandes et al. 2014; Sun et al. 2020) and the tourism industry (Seetanah et al. 2020), but also affect noise pollution (Fajersztajn et al. 2019; Zheng et al. 2020) and air quality (Unal et al. 2005; Zhu et al. 2011; Hudda et al. 2020) in the built-up environment. Additionally, an airport typically has two major elements: the airfield and the terminal (Rodrique 2020). The airfield can be further divided into runway, taxiway, and apron. Acquiring these elements is essential for the design of new airports and the reconstruction of existing ones (Foster et al. 1995; Alexandre et al. 2002; Kazda and Caves 2015; Ke and Bin 2020). Consequently, obtaining airport data is necessary for various applications.

Most of existing studies have used remote sensing or satellite data to detect/extract an airport and its elements (e.g., runway, taxiway and apron), because remote sensing is a well-known technique to identify objects on the Earth's surface. For instance, Jackson et al. (2015) used very high resolution satellite data (QuickBird, with a resolution of 61cm) to extract a runway's area precisely. Xu et al. (2018) proposed an airport detection method based on convolutional neural networks (CNN). The CNN has also been used by other researchers (Chen et al. 2018; Li et al. 2019; Yin et al. 2020) to automatically detect airports from high-resolution satellite data. However, the use of remote sensing data involves a series of pre-processing steps (including image calibration, feature detection and data classification), most of which are still technical challenges for most planners and designers. On the other hand, high-resolution (e.g., less than 1m) remote sensing data are still not available for the public, thus it is necessary to employ other open data source as supplements for acquiring airport data.

Along with the development of Web2.0 technique, open data edited and provided by global volunteers (also known as volunteered geographical information or VGI, Goodchild 2007) have also been viewed as potential data source for acquiring airport data. OurAirports (https://ourairports.com/) is such a VGI platform, which has provided the location information (including longitude and latitude) of more than 70,000 airports worldwide. OpenStreetMap (OSM) is another well-known and widely-used VGI platform (https://www.openhistoricalmap.org/), which has been edited by almost eight million global volunteers. More important, from the OSM platform, it is possible to acquire various elements (e.g., runway, taxiway, apron and terminal) of an airport, in terms of both geometric and thematic information. As an example, Figure 1 shows a screenshot of the Hannover International Airport in OSM, from which the major elements (including runway, taxiway, apron and terminal) of this airport can be identified. Moreover, the corresponding (geometric) data...
can also be freely acquired with much less technical challenges.

![Figure 1. A screenshot of the Hannover International Airport in OSM.](image)

Despite of the above advantages, the data (especially OSM) may unavoidably have the quality issue, because they are provided by volunteers from different countries and educational backgrounds (Arsanjani and Bakillah 2015). Extensive studies have assessed OSM data quality from various quality measures (Senaratne et al. 2017), e.g., positional accuracy (Brovelli et al. 2018), semantic accuracy (Zhou et al. 2019; Wang et al. 2020), completeness (Zhou et al. 2018; Zhang et al. 2022) and logical consistency (Sehra et al. 2020; Zacharopoulou et al. 2021), which were defined by the International Standard Organization (ISO 2003) and have been widely used. Among these different measures, the completeness, measuring how well an region has been mapped, has received wide attention because the other quality measures should be assessed based on existing data. Extensive research work can also be classified according to different map features that have been assessed. For instance, lots of work has been focused on assessing OSM data quality in terms of roads (Girre and Touya 2010; Zhou and Tian 2018; Lin and Zhou 2020), buildings (Fan et al. 2014; Zhou 2018; Zhang et al. 2022), points of interest (POI, Touya et al. 2017; Yeow et al. 2021), land-use and land-cover (Johnson and Lizuka 2016; Zhou et al. 2019; Wang et al. 2020). Despite of these effects, to the best of our knowledge, there still is a lack of research work on assessing OSM data quality in terms of the feature-airport, which is the main purpose of our study.

Specifically, this study has two main contributions: First, the completeness of more than 4,000 airports in OSM has been investigated at a global scale. This was achieved by comparing with another open dataset (OurAirports).

Second, these airports were assessed at two different scales, i.e., airport-based (each airport was assessed) and national-based (all the airports of a country was assessed). Moreover, such an assessment was carried out by taking not only the different types (Large, Median and Small) of airports, but also the different elements (runway, taxiway, apron and terminal) of them into consideration.

The paper is structured as follows: Section 2 describes the acquired data, the approaches for assessing the completeness of airports worldwide, and also the approaches for analyzing the results; Section 3 reports the results and analyses; Section 4 and 5 are discussion and conclusion, respectively.

2. Data and Approach

2.1 Data

Two categories of datasets (OSM and OurAirports) were involved for the analysis. Specifically, the location data of airports worldwide were acquired from OurAirports, and they were used as references for comparing with OSM data.

- **OSM Data:** The data were freely acquired from the third-party platform on February 2022. This platform has provided OSM datasets of almost all the countries and regions in the world. Moreover, different map features, e.g., roads, railways, buildings, land-use and land-cover, have been included in these datasets. Moreover, an OSM tag, consisting of a key and a value, is used to describe the attribute of an object. As an example, if an OSM object tagged with "aeroway = terminal", it means that this object is the terminal of an airport. Specifically, we extracted OSM objects with the tags related to the four major elements, i.e., runway, taxiway, apron and terminal. The corresponding tags are also listed in Table 1. It should be noted that the two tags "aeroway = runway" and "aeroway = taxiway" may be represented by not only line but also polygon features, both of which were extracted for the analysis.

- **Reference data:** The reference data were acquired from OurAirports. Since February 2022, this platform has provided the locations of 71,611 airports across the global. Moreover, these airports have been classified into seven different types, including (1) Large airport, (2) Medium airport, (3) Small airport, (4) Helicopter, (5) Seaplane, (6) Closed and (7) Balloonport (Table 2). However, the three types of airports (Helicopter, Seaplane and Balloonport) normally do not include runway, taxiway and terminal; and most of the airports classified as the type-Closed were belonged to the above three types. Therefore, this study only involved the other three types (i.e., Large airport, Medium airport and Small airport) into the analysis. Specifically, a total of 43,647 airports worldwide have been found from the reference data. To verify the accuracy of reference data for airports, we randomly selected 10% of the total number of large, medium, and small airports and combined them with Google images for visual interpretation. The accuracy rate was 100% for large and medium airports and 82% for small airports. This data can be used as a reference to assess the quality of OSM airport data.

Furthermore, Figure 2 shows the spatial pattern of airports distributed across the globe. It can be seen from this figure that: there are a total of 247 countries and...
regions that have at least an airport. Six out of these countries and regions have a relatively large number (>1,000) of airports, and the United States has the largest number (>10,000) of airports.

<table>
<thead>
<tr>
<th>Major elements of an airport</th>
<th>OSM Tag</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Runway</td>
<td>Aeroway=runway</td>
<td>Line/Polygon</td>
</tr>
<tr>
<td>Taxiway</td>
<td>Aeroway=taxiway</td>
<td>Line/Polygon</td>
</tr>
<tr>
<td>Apron</td>
<td>Aeroway=apron</td>
<td>Polygon</td>
</tr>
<tr>
<td>Terminal</td>
<td>Aeroway=terminal</td>
<td>Polygon</td>
</tr>
</tbody>
</table>

Table 1 The OSM tags for the major elements of an airport.

<table>
<thead>
<tr>
<th>Type</th>
<th>Total Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large airport*</td>
<td>448</td>
</tr>
<tr>
<td>Medium airport*</td>
<td>4,745</td>
</tr>
<tr>
<td>Small airport*</td>
<td>38,453</td>
</tr>
<tr>
<td>Helicopter</td>
<td>17,846</td>
</tr>
<tr>
<td>Seaplane</td>
<td>1,102</td>
</tr>
<tr>
<td>Closed</td>
<td>8,830</td>
</tr>
<tr>
<td>Balloonport</td>
<td>36</td>
</tr>
<tr>
<td>All</td>
<td>71,461</td>
</tr>
</tbody>
</table>

Table 2 The statistic of airport data in OurAirports according to different types.

*denotes the types of airports that have been involved into the analysis.

![Figure 2](image.jpg) The spatial pattern of airports across the globe

### 2.2 Approaches

To the best of our knowledge, there is a lack of approaches for the quality assessment of airport data in OSM. Thus a simple approach was designed. The tenet of our approach generally includes three steps. First, the correspondence between OSM and reference airport datasets were established. Second, the completeness of OSM dataset was assessed. Third, the assessment results were analyzed.

#### 2.2.1 Establishing correspondence between OSM and reference datasets

The OSM and reference datasets were represented using different features. To be specific, the OSM dataset was represented using lines and/or polygons, but the reference dataset was represented using points. Thus it is necessary to first establish the correspondence between these two datasets, in order to determine which lines and/or polygons in the OSM dataset are corresponded to the corresponding point or location of an airport in the reference dataset. Theoretically, it may be possible to establish such a correspondence according to the name of each airport. However, there are a lack of names for most OSM data. Therefore, a buffer approach was proposed to establishing such a correspondence. The specific steps are listed as follows:

- **Step 1**: Create a buffer around the location of each airport in the reference dataset. According to the International Standards of Aerodrome Design and Operations (International Civil Aviation Organization 2009), the size of an airport depends much on the length of its runway(s), which may vary from 1800 to 2500m. For extreme large airports (e.g., Denver International Airport, https://www.flydenver.com/about/media_center/den_overview), the length of runway may be close to 5000m. Therefore, in our study, the 5000m was empirically determined as a threshold to create the buffer.
- **Step 2**: For each buffer, we searched from the OSM dataset to identify that whether there is an object that located inside the buffer. If this is the case, we identified that there is at least an OSM object corresponded to this buffer and its corresponding airport. Otherwise, there is no OSM object corresponded to this buffer.

#### 2.2.2 Assessment of Completeness

The completeness of 43,647 airports worldwide was assessed at two different scales i.e., called airport-based assessment and national-based assessment.

1. **Airport-based assessment**

   The airport-based assessment denotes that the completeness of each airport was assessed. For each airport, the four different elements (runway, taxiway, apron and terminal) were analyzed individually. Each element of an airport was determined as 'complete', if there is at least an OSM object that not only corresponded to this airport, but also tagged with the corresponding element.

2. **National-based assessment**

   The completeness of all airports in each country was also assessed, in terms of different elements. For each country, the completeness of an element was defined as follows:

   \[
   C(e) = \frac{N_{\text{osm}}(e)}{N_{\text{ref}}} \times 100\% \tag{1}
   \]

   where, \(C(e)\) denotes the completeness of all airports (in a country) for a certain element e; \(N_{\text{osm}}(e)\) denotes the total number of airports that identified as 'complete' in terms of the element e; \(N_{\text{ref}}\) denotes the total number of airports in a country.

#### 2.2.3 Analysis of Results

As the completeness may vary with different airport types (large, medium and small), the 43,647 airports worldwide were analyzed according to the three types, respectively. To be specific:

- First of all, the completeness of OSM airport data was visualized on a map, using both the airport-based and national-based assessments. For each assessment, the three different types (large, medium and small) and the four different elements (runway, taxiway, apron and terminal) of an airport was analyzed, respectively.
Then, the completeness was quantitatively analyzed using the box plot, in terms of different geographical regions (African, Asia, Europe, North America, Oceania and South America). The purpose of this analysis is to investigate whether the completeness varies with different geographical regions.

Lastly, some typical airports were also picked up and overlapped with Google Earth images, in order to explain why they have been mapped well or not. In this analysis, different types and elements were also considered respectively.

3. Results and Analyses

3.1 Results of spatial patterns

First of all, Figures 3-4 visualize the completeness of OSM airport data across the globe, in terms of two scales (airport-based and nation-based assessments), three types (large, medium, and small), and four elements (runway, taxiway, apron, and terminal). For the airport-based assessment, all 43,647 airports were divided into two classes: 'complete' (there is at least one OSM object tagged with the corresponding element) and 'incomplete' (there are no OSM objects tagged with the corresponding element). For the nation-based assessment, the completeness of each country was divided into five classes, ranging from 0% to 100% with an interval of 20%.

Figures 3-4 show that,

1) Airport-based: The four major elements (runway, taxiway, apron, and terminal) of each large airport are almost complete. In particular, the runway element is complete for all large airports. For the other three elements, data is lacking for no more than eight out of the 448 large airports. Most of the four elements have been mapped for medium airports, with completeness rates of 99.3%, 94.3%, 89.1%, and 65.4% for runway, taxiway, apron, and terminal elements, respectively. However, compared to large airports, there are many more medium airports that have not been mapped with some of these elements. In particular, there are many more medium airports that have not been mapped with the terminal element (1,212) compared to the runway element (34). The completeness patterns for small airports are quite different from those of large and medium airports. The completeness of the four elements is much lower for most small airports, with completeness rates of only 74.8%, 28.3%, 22.3%, and 5.7% for runway, taxiway, apron, and terminal elements, respectively. Relatively, the completeness (74.8%) for the runway element is much higher than that (5.7%) for the terminal element. This is probably because almost all airports have a runway, but not all of them have a terminal.

2) National-based: The completeness of each element is almost 100% for the 160 countries and regions with large airport(s). Nevertheless, there is a lack of the terminal element for some large airports in Saudi Arabia, Tanzania, and Papua New Guinea. For medium airports, the completeness of the runway element is close to 100% for all 228 countries and regions with medium airports, except for the South Pole region. However, the completeness of the terminal element is relatively low (e.g., <60%) for some countries and regions such as Canada (42.5%), Germany (54.7%), the United Kingdom (55.2%), Russia (55.7%), the United States (56.3%), and Australia (59.1%). Similar cases can also be found for small airports. That is, for most countries and regions, the completeness is relatively high (e.g., >60%) in terms of the runway element, but much lower (e.g., <40%) for the other three elements. Nevertheless, there is relatively higher completeness in some countries and regions (e.g., Northern Africa, Western Asia, and India). This may be because these countries and regions have a relatively small total number of airports (e.g., <500, as shown in Figure 2), making it easier for OSM volunteers to map them accurately.

Figures 3-4 show that,

1) Airport-based: The four major elements (runway, taxiway, apron, and terminal) of each large airport are almost complete. In particular, the runway element is complete for all large airports. For the other three elements, data is lacking for no more than eight out of the 448 large airports. Most of the four elements have been mapped for medium airports, with completeness rates of 99.3%, 94.3%, 89.1%, and 65.4% for runway, taxiway, apron, and terminal elements, respectively. However, compared to large airports, there are many more medium airports that have not been mapped with some of these elements. In particular, there are many more medium airports that have not been mapped with the terminal element (1,212) compared to the runway element (34). The completeness patterns for small airports are quite different from those of large and medium airports. The completeness of the four elements is much lower for most small airports, with completeness rates of only 74.8%, 28.3%, 22.3%, and 5.7% for runway, taxiway, apron, and terminal elements, respectively. Relatively, the completeness (74.8%) for the runway element is much higher than that (5.7%) for the terminal element. This is probably because almost all airports have a runway, but not all of them have a terminal.

2) National-based: The completeness of each element is almost 100% for the 160 countries and regions with large airport(s). Nevertheless, there is a lack of the terminal element for some large airports in Saudi Arabia, Tanzania, and Papua New Guinea. For medium airports, the completeness of the runway element is close to 100% for all 228 countries and regions with medium airports, except for the South Pole region. However, the completeness of the terminal element is relatively low (e.g., <60%) for some countries and regions such as Canada (42.5%), Germany (54.7%), the United Kingdom (55.2%), Russia (55.7%), the United States (56.3%), and Australia (59.1%). Similar cases can also be found for small airports. That is, for most countries and regions, the completeness is relatively high (e.g., >60%) in terms of the runway element, but much lower (e.g., <40%) for the other three elements. Nevertheless, there is relatively higher completeness in some countries and regions (e.g., Northern Africa, Western Asia, and India). This may be because these countries and regions have a relatively small total number of airports (e.g., <500, as shown in Figure 2), making it easier for OSM volunteers to map them accurately.

Figures 3-4 show that,
Figure 4 The OSM completeness of medium airports across the globe, in terms of the four different elements (runway, taxiway, apron and terminal) in the national-based assessments.

Moreover, Figure 5 displays the box plots of OSM airport completeness for different geographical regions. For each geographical region, the completeness of different countries was statistically analyzed using box plots, in terms of three types and four elements. We can observe from this figure that: First, all four elements are almost complete for large airports. Second, most of these elements have still been mapped for medium airports, although they exhibit relatively lower completeness. Third, in terms of all elements, there is much lower completeness for small airports, although there is a relatively high completeness for the runway element. These results are consistent with those found in Figures 5.

Figure 5 Box plots of OSM airport completeness of six different geographical regions (African, Asia, Europe, North America, Oceania and South America), in terms of the three types (large, medium and small) and the four elements (runway, taxiway, apron and terminal).

Furthermore, some airport data in OSM were selected, and they were overlaid with Google Earth images (Figure 6). This type of analysis may be beneficial for us to understand why some airports have not been mapped. First of all, there is a lack of apron and terminal data for Jeju International Airport (Figure 6a) and Peterborough Municipal Airport (Figure 6b); however, these elements can be clearly interpreted from Google Earth images. In contrast, in Figures 6c and 6d, some elements (e.g., taxiway) have not been mapped well, probably because they cannot be interpreted from the satellite images. In Figures 6e and 6f, we cannot even interpret any element of an airport from the corresponding Google Earth image. Thus, the locations of these airports provided by OurAirports are likely incorrect.

Figure 6 Overlapping the OSM data of six incomplete airports with corresponding Google Earth images.

4. Discussion

This section will discuss not only the potential applications of OSM airport data but also the limitations of this study.

4.1 Application

This purpose of this study is to investigate the data quality (especially the completeness) of airport data in OSM. We found that most of large and medium airports have been mapped well with all the four elements (runway, taxiway, apron and terminal), although there is a lack of OSM data for most small airports. The OSM (airport) data may be used in several potential applications.
First, OSM data can be used for airport planning and design. Although the layouts of various elements (e.g., runway and terminal) may vary among different airports, it is possible to classify them into some common categories (Johnson et al. 2016; Rodrigue 2020). The major elements of an airport can be freely acquired from OSM data, which are also nearly complete, especially for the 448 large airports and the 4,745 medium airports. With such a large number of samples, it may be possible for planners and designers to refer to existing layouts when designing a new airport or reconstructing an existing airport (Sumathi and Selvam 2018).

Secondly, the OSM data may be used as training samples for airport detection and/or identification. Previous studies have reported that machine learning methods, such as convolutional neural networks, are useful for detecting and/or identifying airports from remote sensing images (Xu et al., 2018; Chen et al., 2018). However, using machine learning methods requires a large number of samples for training. The OSM data can provide a vast number of airport elements from around the world. More importantly, these data can be used as training samples for detecting and/or identifying airports that have not been mapped by OSM volunteers.

Thirdly, previous studies have reported on the impact of airports on noise pollution and air quality (Unal et al., 2005; Zhu et al., 2011; Fajersztajn et al., 2019; Hudda et al., 2020). To conduct this type of analysis, detailed data on various airport elements, such as runways, taxiways, and aprons, are needed. As far as we know, OSM data may be the only open-source option for acquiring such information. Additionally, airport data in OSM can be combined with other data sources, such as population and road networks, to study airport ground accessibility (Sun et al., 2020). Thus, airport data in OSM can be used not only to improve the built environment, such as noise pollution and air quality but also to promote socio-economic development, such as airport ground accessibility.

4.2 Limitations

Despite the airport data in OSM have a lot of potential applications, this study also has several limitations. First of all, we acquired the location data of airports from the OurAirports platform as references, which were then compared with OSM data. This platform has freely provided more than 70,000 airports worldwide. However, the data in OurAirports were provided by global volunteers, and thus they may contain errors. For instance, we found in this study that some airports lack OSM data (e.g., Figure 6e), probably because the reference locations of these airports are incorrect. Additionally, open airport data represented by lines and/or polygons are still not available, so we had to use the point data in OurAirports as references for the analysis. Therefore, we could only investigate whether an element had been mapped or not, rather than how many of these elements had been mapped. Thus, even airports identified as 'complete' may still lack OSM data.

Secondly, in addition to completeness, other quality measures such as positional accuracy, attribute accuracy, and logical consistency are also essential for assessing the quality of open airport data. Therefore, in further work, it would be worthwhile to assess the quality of OSM airport data in terms of different quality measures.

Last but not least, in addition to the four elements studied, there are other elements of an airport (e.g., hangars and helipads, https://wiki.openstreetmap.org/wiki/Aeroways) that can be acquired from OSM data. However, these elements are either difficult to interpret from Google Earth images or not major elements of an airport, and thus they were not considered in our study. In future work, it would also be interesting to assess OSM data quality in terms of other elements of an airport.

5. Conclusion

The purpose of this study is to assess the quality, particularly the completeness, of airport data in OSM. This is achieved by comparing it with location data of airports acquired from another platform, OurAirports. First, the four different elements of airports (runway, taxiway, apron, and terminal) were extracted from OSM data. Then, these elements were matched with the location data of airports in OurAirports using buffer analysis. Finally, the airports' various elements in OSM were assessed. More than 40,000 airports across the globe were analyzed, and they were categorized into three different types: large, medium, and small airports. The assessment was carried out at two different scales: airport-based and national-based. Results showed that:

Firstly, the completeness of OSM airport data depends on various types. Specifically, all 448 large airports are almost complete for all four elements. Most of the 4,745 medium airports are complete, especially for the three elements (runway, taxiway, and apron). However, there is a lack of OSM data for most small airports, in terms of various elements. This is probably because large airports have received much attention from OSM volunteers, and thus they have been mapped better than medium and small airports.

Secondly, in terms of the four different elements, the completeness is relatively high for the element "runway," but relatively low for the element "terminal." This is because almost all airports have at least one runway, but not all of them have terminal(s).

Thirdly, the completeness also varies with different geographical regions. The completeness is relatively high in Northern Africa, Eastern Asia, and India, probably because there is a relatively small number of airports in these countries and regions. On the contrary, the completeness is relatively low in Canada, the United...
States, Russia, Australia, and Southern Africa, probably because there is a relatively large number of airports, which may require more effort from OSM volunteers to map well.

In addition, the lack of some or all elements of an airport may be attributed to different reasons, such as the data not being mapped by OSM volunteers, or a lack of elements (e.g., taxiway, apron, and/or terminal) for some airports. Furthermore, the location data of a reference airport may also be incorrect.

We have also discussed the potential applications of OSM airport data and the limitations of this study. Therefore, in future work, some points may be considered to improve our study. Firstly, the quality of airport data in OSM may be assessed using other quality measures (e.g., positional accuracy, attribute accuracy, and logical consistency). Secondly, the data quality of other elements of an airport may also be analyzed. Finally, it is interesting to investigate how to use airport data in OSM for various applications.

6. Acknowledgements
The project was supported by National Natural Science Foundation of China [Grant No. 41771428; 4207040449].

7. References


