



The effect of tectonic plate motion on georeferenced long-term global datasets

Franz-Benjamin Mocnik^{a,b,*}, René Westerholt^{c,d}



^a Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, The Netherlands

^b Institute of Geography, Heidelberg University, Heidelberg, Germany

^c School of Spatial Planning, TU Dortmund University, Dortmund, Germany

^d Centre for Interdisciplinary Methodologies, University of Warwick, Coventry, UK

ARTICLE INFO

Keywords:

Plate tectonics

Spatial data quality

Positional accuracy

OpenStreetMap (OSM)

Volunteered Geographic Information (VGI)

World Geodetic System 1984 (WGS84)

ABSTRACT

Tectonic plate motion affects coordinates resulting from GPS measurements and the referencing of aerial and satellite imagery. It therefore impacts the long-term use of global coordinate systems. Over time, the tectonic plates move relative to each other and coordinates become outdated. Most geographic datasets including OpenStreetMap are no exception, as these are affected in terms of a growing location-coordinate mismatch. Current research is aware of this issue but solution strategies have not been fully explored yet. In this manuscript, we discuss how regular systematic updates of coordinate values can be used to address this issue. We explore the recommended frequency to perform such updates for guaranteeing a minimal loss of accuracy after long periods of time. It is further determined how rounding errors impair such systematic updates and in which ways singular and irregular manual updates impede systematic solution strategies. The solution strategies proposed lead to minimal overall errors and thus guarantee to retain high positional accuracy of coordinate pairs within global datasets, even after years of existence.

1. Introduction

Georeferenced datasets can suffer from a range of data quality issues, such as incompleteness, vague semantic descriptions, and measurement errors potentially leading to positional inaccuracies. Numerous methods to measure such deficiencies have been developed and are discussed in the literature (for an overview of such methods in case of OpenStreetMap, a long-term global dataset exemplarily discussed later, see Mocnik et al. 2018). The importance of such methods becomes evident for applications that require high positional accuracies. For instance, while it is sufficient for pedestrians to navigate in urban environments using data accurate in the metre range, the same accuracy would not be sufficient for autonomous driving applications, often requiring coordinates to be accurate up to a few decimetres. Besides errors actively induced in the processes of creating and maintaining the data, positional and other errors can also emerge without any active involvement of contributors and maintainers. For instance, this happens whenever the frame of reference underlying a dataset changes, which then also causes its meaning to change in possibly unintended ways.

The effect of plate motion causes a change in the geospatial frame of reference, which renders the meaning of geodetic coordinates in the data. This becomes an issue for georeferenced datasets created through a range of pathways, including GPS measurements and by

digitization using aerial and satellite imagery. Before imposing any geodetic coordinate system, a Geodetic Reference System (GRS) needs to be introduced. Such a GRS consists of a reference ellipsoid and a reference frame, which relates the reference ellipsoid to the surface of the Earth by providing a list of reference points on the surface of the Earth and their corresponding positions on the reference ellipsoid. One of the most widely used and known GRSs is the World Geodetic System 1984 (WGS84; National Imagery and Mapping Agency 2000). It aligns with the International Terrestrial Reference Frame (ITRF; Boucher and Altamimi 2001, National Imagery and Mapping Agency 2000), the latter of which presumes, by definition, the reference ellipsoid and the Earth (including its atmospheric mass) to share a common centre of gravity.

The ITRF assumes a no-net-rotation condition (NNRC; Altamimi et al. 2003, Kreemer and Holt 2001, Argus and Gordon 1991), i.e., it assumes the total angular momentum of the tectonic plates to vanish in respect to the ITRF. Such an NNRC is advantageous for global datasets because it minimizes the global effect of tectonic plate motion on coordinates. As the ITRF depends on tectonic plate motion and further changes of the Earth, the ITRF changes over time. Accordingly, realizations for specific points in time exist, such as the ITRF92 and the ITRF2000. Even the methods to determine ITRF realizations have changed over time. The realization ITRF2000 is, e.g., the first one

* Corresponding author at: Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, The Netherlands.

E-mail address: franz-benjamin.mocnik@utwente.nl (F.-B. Mocnik).

URL: <http://www.mocnik-science.net> (F.-B. Mocnik).

that is free from any tectonic plate motion model but instead refers to unconstrained space geodesy solutions only (Altamimi et al., 2002, 2003). The WGS84 is accordingly realized for specific points in time as well, such as the realizations WGS84 (G730) and WGS84 (G1150) for the aforementioned realizations of the ITRF. Despite the intention of the NNRC to minimize the global effect of tectonic plate motion on coordinates, the resulting realizations cannot address the fact that tectonic plates move at different speeds and into different directions relative to each other and thus also relative to the ITRF. The Australian plate, e.g., moves with a speed of around 6.8 cm/a, or about 1 m in 15 years, relative to the ITRF (Kreemer et al., 2014). In addition to (tectonic) plate motion,¹ which is the most relevant factor for such motion of locations on the Earth's surface relative to the ITRF, also intraplate deformation contributes to such motion. As has been argued by Mocnik and Raifer (2018) in a previous publication, the effects outlined should become measurable in the next years or decades in the OpenStreetMap dataset, which is long-term and global per se, creating the eventual need of addressing this issue.

Issues arising from coordinate shifts can be addressed in two ways: either by utilizing several local coordinate systems such as the North American Datum of 1983 (NAD83; U.S. Department of Commerce and U.S. Census Bureau 2006) or the Geocentric Datum of Australia 1994 (GDA94; Donnelly et al. 2014, Intergovernmental Committee on Surveying and Mapping and Permanent Committee on Geodesy 2018); or by systematic updates of coordinate values. The latter has the advantage of only one GRS being used and there is, accordingly, no need to perform transformations between local coordinate systems when using the data. This article addresses the following two research questions:

- RQ1: How can we minimize the positional loss of accuracy caused by recurrent, systematic coordinate updates performed to correct for shifts caused by tectonic plate motion?**
- RQ2: What are the effects additional manual updates have on the systematic updating strategy?**

The article is structured as follows. Section 2 addresses RQ1 by exploring how frequent systematic updates need to be performed in order to minimize the overall decrease in positional accuracy. Such systematic updates, however, often interfere with additional manual updates of the data, e.g., as found in many Volunteered Geographic Information (VGI) datasets. We discuss how to account for the effect of such manual updates on the mentioned systematic updating strategy, which answers RQ2 (Section 3). The consequences of this adaptation of the systematic updating strategy to account for manual updates are discussed in Section 4. Subsequently, we explore how severe the issue of tectonic motion is for the example of OpenStreetMap data, and what the systematic updating strategy would mean (Section 5). Further datasets expose very different characteristics, thereby posing additional challenges to systematic updating strategies (Section 6). Future research opportunities for addressing such challenges as well as limitations of the work presented are discussed in Section 7.

2. Systematic updating strategies

In this section, we discuss how to systematically counteract the effect of plate motion in a long-term global dataset. Consider the following scenario: Due to plate motion, coordinate values contained in a long-term global dataset shift with an average speed of σ , whereby this speed depends on the respective tectonic plate. For counteracting this

¹ The arguments provided in this article apply, for the most part, to both tectonic plate motion and intraplate deformation. For the remainder of this article, however, we refer to tectonic plate motion only for the sake of efficient communication. The interested reader can easily translate the arguments to intraplate deformation by considering the average speed σ for much smaller areas than an entire tectonic plate.

shift, the coordinate values shall be regularly updated in a systematic manner to (more or less) match their new actual positions. In the following, we assume the coordinates to be represented with a technical precision of p , i.e., the level of detail up to which numbers can be expressed in the data. In addition, we make the following assumptions:

- A1:** no phenomena other than plate motion affect the accuracy of the coordinates (e.g., changing the spatial resolution used in the database is not considered),
- A2:** plate motion is not accounted for in the maintenance of the data beyond the systematic updates,
- A3:** the data sources used to improve and maintain the data are up to date when the data is updated, and
- A4:** the data are not updated between two systematic updates.

These assumptions clearly impose simplifying conditions deviating from what is found in most real-world datasets. However, the simplifications made allow for an adequate estimation of how to systematically update real datasets sufficiently well. In order to answer RQ1, we determine the interval δ at which the coordinates shall be updated in order to minimize the average error after Δ years. To be more precise, we aim to minimize the average error on the time interval from $\Delta - \delta$ to $\Delta + \delta$ (Fig. 1).

There are two major issues constraining the accuracy in the considered scenario: rounding errors and plate motion between two updates. First, rounding errors can occur at each update. The maximal rounding error at each update is half the precision, i.e., $p/2$, yielding the expected average of the rounding error to be $p/4$. After Δ years, there will have been Δ/δ updates. We thus expect the rounding errors to sum up to $\Delta/\delta \cdot p/4$ after Δ years.

Despite correcting for the effect of plate motion, the coordinate values are, secondly, inaccurate due to the ongoing plate motion between two updates. When introducing a pair of coordinates in the data, we assume the coordinates to be correct and up to date according to Assumption A3. After a time span of δ , the accuracy will have decreased by $\delta\sigma$, as follows from the speed of plate motion and the Assumptions A1 and A2 (Fig. 1a). Also, the coordinates are not updated other than by the systematic updates according to Assumption A4, meaning that no additional error is introduced. The average of the accuracy can accordingly be maximized by updating the coordinates to the values that the corresponding location will have $\delta/2$ later, instead of updating the coordinates to the values that the corresponding location has at the point in time when the update is performed (Fig. 1b). In this case, the accuracy will be decreased by $\delta\sigma/2$ in maximum, which means an average error of $\delta\sigma/4$. The overall average error \mathcal{E} is thus given as

$$\mathcal{E} = \frac{\sigma}{4} \cdot \delta + \frac{p\Delta}{4} \cdot \delta^{-1}. \quad (1)$$

The overall error depends on the update interval δ . If this interval is chosen in a suitable way, the error can be minimized. For the optimal choice δ_m , the first derivative vanishes²:

$$\left. \frac{\partial \mathcal{E}}{\partial \delta} \right|_{\delta_m} = \frac{\sigma}{4} - \frac{p\Delta}{4} \cdot \delta_m^{-2} \stackrel{!}{=} 0.$$

Further observe that the second derivative $\partial^2 \mathcal{E} / \partial \delta^2$ is positive at δ_m . The overall error is thus minimized when the coordinate values are updated at the interval δ_m where

$$\delta_m = \sqrt{\frac{p\Delta}{\sigma}}. \quad (2)$$

Accordingly, the minimal overall error is

$$\mathcal{E}_m = \frac{\sigma}{4} \cdot \delta_m + \frac{p\Delta}{4} \cdot \delta_m^{-1} = \frac{1}{2} \sqrt{\sigma p \Delta}. \quad (3)$$

Both the optimal update interval δ_m and the optimal overall error \mathcal{E}_m depend on the technical precision of the coordinate values p , the average speed of plate motion σ , and the time interval Δ optimized for.

² A continuous function is extremal only if the first derivative vanishes, and it is minimal if and only if the second derivative is positive in addition.

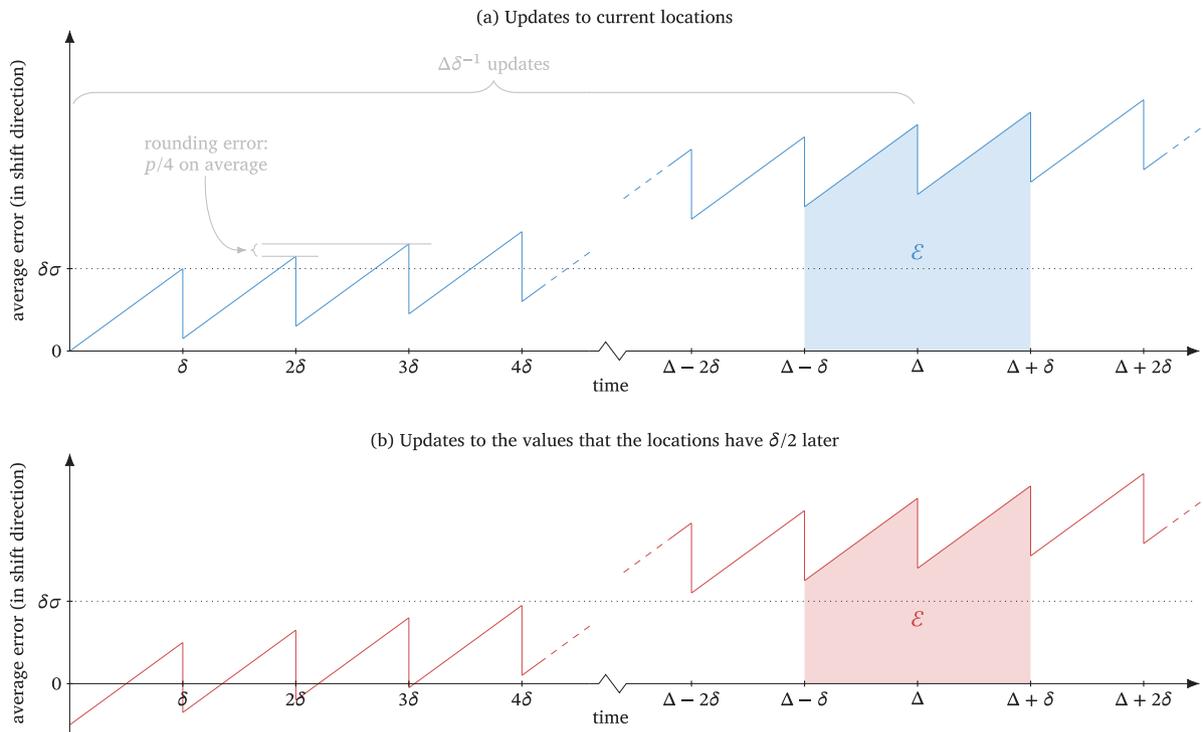


Fig. 1. Systematic updating strategy. The schematic visualization illustrates the regular updates, which are performed at an interval of δ and aim to minimize the positional error in the long run. (a) The positional error is both caused by plate motion (error of $\delta\sigma$ after δ years) and rounding errors (average error of $p/4$ at each update). The update strategy explored aims to minimize the average error \mathcal{E} in the interval $[\Delta - \delta, \Delta + \delta]$, which is depicted by the blue area. (b) When updating coordinates to the values that the locations have $\delta/2$ later instead of to current locations, the average error \mathcal{E} , depicted by the red area, can be minimized even more.

3. Accounting for unsystematic manual updates

In the last section, we have discussed how to systematically update coordinate values in long-term global datasets for minimizing the effect of plate motion. This discussion assumed that plate motion has no impact on the maintenance of the data in between two updates (Assumption A4). For real datasets this is, however, often not the case. As an example, updated geometries reflect plate motion whenever they rely on up-to-date aerial imagery or GPS measurements. In this section, we discuss how such manual updates of geometries impede the systematic updates discussed previously.

In principle, manual updates can easily be reflected in systematic updates. If a geometry has been updated manually at some point in time t_0 , a systematic update at t_1 needs to reflect the plate motion in the timespan $[t_0, t_1]$ only. Systematic updates would, accordingly, be performed individually for each geometry of the dataset. This approach makes two presumptions. First, the point in time of a manual update needs to be known, which is often the case. Secondly, the information the manual update is based on, e.g., aerial imagery, needs to be up to date, or it needs to be known at which point in time this information was up to date. This second presumption is in many cases not met, which is why manual updates need to be reflected in a more general way in the systematic updates. In order to answer RQ2, we thus determine the interval δ' at which the coordinate values shall be updated in order to maximize the accuracy after Δ years, assuming that some $100 \cdot u$ per cent of the coordinate values (u being in the unit interval) are in addition manually updated at an average interval of θ years.

Manual updates of coordinate values increase their accuracy but impede systematic updates. As assumed before, $100 \cdot (1 - u)$ per cent of the coordinate values are not updated manually. For them, Eq. (1) still applies. For the other $100 \cdot u$ per cent, the manual updates need to be reflected. For the remainder, we still assume A1–A3. However, we replace A4 with the assumptions asserting that

A4a: the shift of coordinates is small enough to retain a mental link between coordinates and shifted locations, allowing for meaningful manual updates, and

A4b: the manual updates are made on the basis of up-to-date data sources.

Based on these two Assumptions A4a and A4b, the manual updates correct coordinate values to their current position. They rely on external information and do thus not incorporate the accumulated errors from before. Despite this, manual updates lead to an overcorrection increasing error. They shift the coordinate values to the current position according to Assumption A4b, but the subsequently performed systematic update assumes that such manual update has not been performed. As discussed before, the average error in between two systematic updates is $\delta\sigma/4$, and the overcorrection is likewise. After Δ years, Δ/θ updates have been performed on average, yielding an overall overcorrection of $\Delta/\theta \cdot \delta\sigma/4$.

Rounding errors are, by and large, eliminated by the manual updates. Only rounding errors introduced by systematic updates made after the last manual update need to be considered. The average number of such systematic updates can be estimated by θ/δ , which results in the overall rounding error of $\theta/\delta \cdot p/4$. The resulting overall error needs to incorporate Eq. (1) for the parts of the data not manually updated, as well as the overcorrection and the rounding errors of the part manually updated, yielding

$$\begin{aligned} \mathcal{E}' &= (1 - u) \left(\frac{\sigma}{4} \delta + \frac{p\Delta}{4} \delta^{-1} \right) + u \left(\frac{\Delta}{\theta} \frac{\sigma}{4} \delta + \frac{p\theta}{4} \delta^{-1} \right) \\ &= \left(1 + u \left(\frac{\Delta}{\theta} - 1 \right) \right) \frac{\sigma}{4} \delta + \left(1 + u \left(\frac{\theta}{\Delta} - 1 \right) \right) \frac{p\Delta}{4} \delta^{-1}. \end{aligned}$$

This can be written as

$$\mathcal{E}' = \mu_{\Delta/\theta}(u) \frac{\sigma}{4} \cdot \delta + \mu_{\theta/\Delta}(u) \frac{p\Delta}{4} \cdot \delta^{-1} \tag{4}$$

when defining $\mu_{\cdot}(u)$ accordingly. Similar to the previous section, the error can be minimized by choosing a suitable interval δ_m' at which

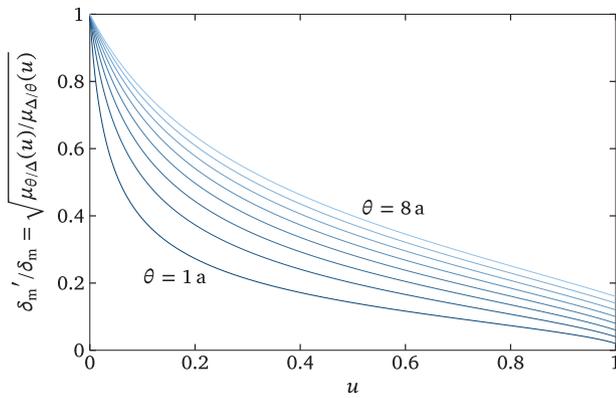


Fig. 2. Coefficient describing the influence of manual updates on the optimal update interval δ_m' . Timespan Δ is equal to 50 years and θ ranges from 1 to 8 years.

the first derivative vanishes:

$$\frac{\partial \delta_m'}{\partial \delta} \Big|_{\delta_m'} = \mu_{\Delta/\theta}(u) \frac{\sigma}{4} - \mu_{\theta/\Delta}(u) \frac{p\Delta}{4} \cdot \delta_m'^{-2} \stackrel{!}{=} 0.$$

The overall error is thus minimized when the coordinate values are updated at the interval δ_m' where

$$\delta_m' = \sqrt{\frac{\mu_{\theta/\Delta}(u)}{\mu_{\Delta/\theta}(u)}} \cdot \sqrt{\frac{p\Delta}{\sigma}} = \sqrt{\frac{\mu_{\theta/\Delta}(u)}{\mu_{\Delta/\theta}(u)}} \cdot \delta_m. \quad (5)$$

Accordingly, the minimal overall error is

$$\begin{aligned} \mathcal{E}'_m &= \mu_{\Delta/\theta}(u) \frac{\sigma}{4} \cdot \delta_m' + \mu_{\theta/\Delta}(u) \frac{p\Delta}{4} \cdot \delta_m'^{-1} \\ &= \sqrt{\mu_{\theta/\Delta}(u) \cdot \mu_{\Delta/\theta}(u)} \cdot \mathcal{E}_m. \end{aligned} \quad (6)$$

Both the optimal update interval δ_m' and the optimal overall error \mathcal{E}'_m can be expressed as scaled versions of their corresponding counterparts that do not consider manual updates. The respective coefficients depend on the fraction u of the coordinate values being updated manually and on the ratio of the average interval θ of the manual updates to the overall number of years Δ optimized for.

4. Discussion of the theoretical findings

The decrease in accuracy caused by tectonic plate motion can be counteracted systematically, as has been shown in previous sections. When updating the coordinate values contained in a dataset at an optimal interval, the overall average error can be kept much lower than without any systematic updates, which addresses RQ1. It is interesting to note that the optimal interval depends on the square root of the overall period of time optimized for. For longer periods of time, different choices of intervals thus lead to only marginal differences.

Manual updates of coordinate values interfere with systematic updates, calling for an adaptation of the systematic update interval. In the previous sections, we have examined this adaption in more detail, by determining a suitable systematic update interval for minimizing the overall average error. As can be seen in Fig. 2, the optimal update interval reacts very sensitive to small fractions of the data being updated manually, in particular, if the manual update interval is much shorter than the systematic update interval. It thus makes a practical difference for the update interval whether no manual updates are performed at all or manual updates are performed for 10 per cent of the data. When larger parts of the data are being updated manually, the optimal update interval reacts, however, less sensitive.

The optimal overall error reacts sensitive to small fractions of the data being manually updated as well (Fig. 3). Again, it makes a practical difference whether no manual updates are performed at all or manual updates are performed for 10 per cent of the data. The same

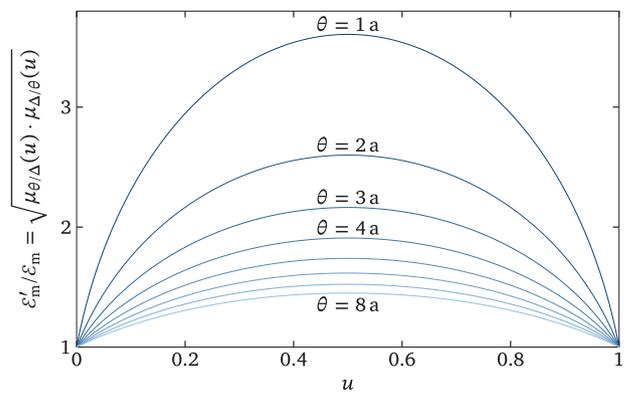


Fig. 3. Coefficient describing the influence of manual updates on the optimal error \mathcal{E}'_m . Timespan Δ is equal to 50 years and θ ranges from 1 to 8 years.

difference applies to the cases of 90 or 100 per cent of the data being updated manually. In fact, the overall error is symmetric in respect to the fraction of the data being updated manually. That is, the overall error is (when adapting the systematic update interval in a suitable way) identical if 10 or 90 per cent are updated manually. The average overall error is maximal when half of the data is manually updated, and it is particularly large if the manual update interval is much shorter than the systematic update interval. The discussed adaption of the systematic updating strategy does thus provide answers to RQ2.

5. The example of OpenStreetMap

5.1. Can the effect be traced in OpenStreetMap?

While tectonic plate motion certainly has an impact on coordinates contained in a global dataset, it is unclear how visible the effect is for a real dataset. Many factors influence how coordinate values are created, changed, and updated, thereby eventually obscuring the effect of plate motion. If these additional effects are larger than the effect of plate motion, there might not be a need to correct for the latter one. In this section, we explore whether the effect of plate motion can practically be traced in OpenStreetMap (OSM) data, an example of a long-term global dataset. The considerations given below extend results published before by Mocnik and Raifer (2018) in a conference contribution.

The OSM dataset is meant to exist long term and to represent geographical features globally. Coordinates are thus represented in WGS84 as GRS, which minimizes the effect of plate motion globally – WGS84 meets a no-net-rotation condition (NNRC) – and maximizes compatibility with many data sources that already utilize WGS84, such as GPS measurements. The data consist of contributions made by volunteers (geometries and attributes) and are, as an ongoing effort, steadily being updated. Changes of the environment are thus potentially incorporated into existing data and the quality of the data potentially improves – initial mapping activities often focus on introducing new features while subsequent updates focus on improving existing information, among them, geometric aspects of mapped features (Neis et al., 2013, 2012).

Manual updates of the coordinate values contained in OSM data are subject to the effect of plate motion. Whenever coordinate values are updated by referring to GPS measurements or aerial imagery – two of the most important data sources for extending and improving the OSM dataset – the updates reflect plate motion to some extent. While GPS measurements are always up to date in the sense of them (or rather: their collection) being independent of plate motion, aerial imagery can become outdated: it may have been created some months or even years before being utilized for mapping purposes. As this affects the data at both the creation as well as at the manual updating stages, manual updates of coordinate values using these data sources should reflect

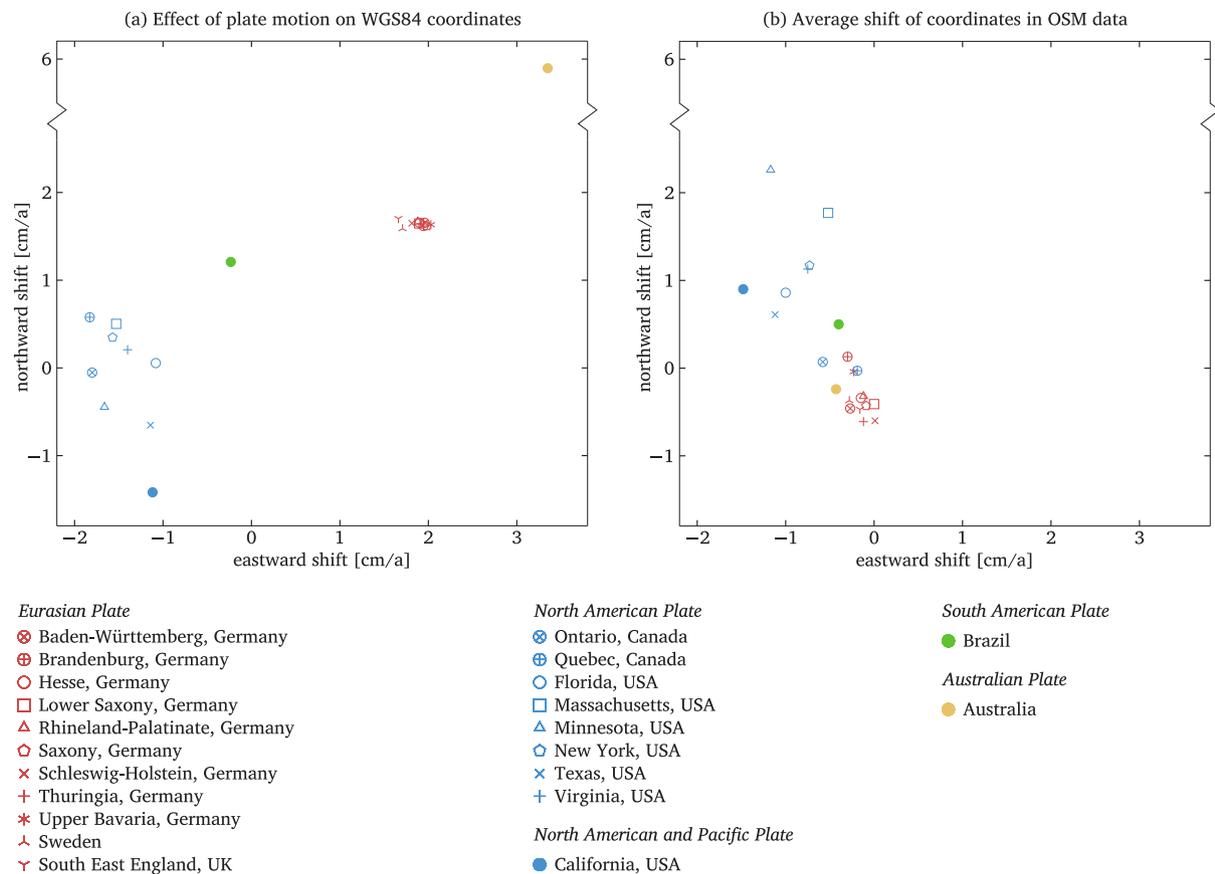


Fig. 4. Comparison of coordinate shifts in OSM data to the effect of tectonic plate motion, for selected locations. (a) Shift of coordinates referring to WGS84 caused by plate motion, according to data found in Kreemer et al. (2014). (b) Average shift of nodes forming part either of a road or street geometry. Only shifts that have occurred after six or more years and the magnitude of which is less than 15 cm/a have been considered. Compare also the figure published by Mocnik and Raifer (2018).

plate tectonic shift, assuming that all other effects show no systematic trends. For the purpose of analysing this shift, all nodes being part of a road or street ('ways' tagged `highway`³) have been analysed. For each such node, the coordinate values assigned at the date of creation have been compared to those assigned during the last update that happened before 24 January 2018. To minimize effects of frequent changes and to focus on long-term effects, nodes were excluded if they existed since less than six years from the date of the last update. Also, nodes have been included only if their updates indicate a shift of less than 15 cm/a, with the aim of excluding major displacements of nodes, which are unlikely to be related to tectonic plate motion (nodes are sometimes reused and attributed with a new meaning to then represent another geographic feature). As a result, one gains an average shift over time for nodes in some particular geographic area (Fig. 4b).

A comparison of the shifts due to plate motion (Fig. 4a) to the shifts that are caused by manual updates (Fig. 4b) reveals similarities as well as differences. First, the magnitudes of the shifts are all in the same range, which renders possible a meaningful comparison. Secondly, areas located on the same tectonic plate expose similar shifts (both in Fig. 4a and in Fig. 4b). In other words: areas that are near on the Earth's surface are also in the same vicinity in the diagrams. Further, the variance of shifts due to plate motion among areas of the Eurasian Plate is very small, which is also true for the average shifts in OSM data. This is in contrast to US states and Canadian provinces, which are scattered across a larger region of the North American Plate and accordingly show a larger degree of variability in both diagrams. Most of the relative differences found in the shifts of nodes that are

located on the same plate are in the millimetre or low centimetre range, especially those found on the Eurasian plate. This shows that the decision to restrict the shifts considered to a range of 15 cm/a maximum has not affected our approach in adverse and unexpected ways.

The structural similarities between the diagrams found in Fig. 4a and in Fig. 4b are, however, in contrast to strong differences. When comparing the average shift for each tectonic plate, there are almost no similarities between the two diagrams. For instance, the Australian Plate is strongly subject to plate tectonic shift (large magnitude when assuming an NNRC) but exposes only very little shift in OSM data. In addition, there is no discernible pattern in how the US states and Canadian provinces are arranged in Fig. 4b. In summary, there are currently little to no *quantitative* similarities between the shift due to plate motion and the average shift in OSM data, apart from both sharing the same order of magnitude.

There are several reasons why the effect of plate motion in OSM data typically goes unnoticed. One reason is the fact that the aerial imagery used as a source of information for adding and updating the data exposes systematic biases (Barron et al., 2014). Such a bias is, e.g., created by the increasing precision of the georeferencing of the imagery and by the imagery being more up-to-date in average. Another reason for the difficult traceability are systematic batch imports. Data imported from other datasets often refer to local GRSs with the aim of minimizing the effect of tectonic shift for those particular local datasets. For instance, large parts of the Tiger/Line dataset have been imported (Zielstra et al., 2013). This dataset uses the North American Datum of 1983 (U. S. Department of Commerce and U. S. Census Bureau, 2006). As part of the import process, the coordinates need to be converted from the original GRS to WGS84, which can lead to systematic shifts. This shift is, however, in parts being retracted during subsequent community-driven

³ For the OSM tagging scheme, see https://wiki.openstreetmap.org/wiki/Map_Features.

data maintenance activities. The same issue applies for datasets about Australia, which tend to prefer the Geocentric Datum of Australia 1994 (GDA94) due to the large plate motion of the Australian Plate (Donnelly et al., 2014; Intergovernmental Committee on Surveying and Mapping and Permanent Committee on Geodesy, 2018). Further, a clear specification of which realization of WGS84 shall be used in OSM data is missing. This is why coordinate values referring to different realizations are used simultaneously, e.g., when using GPS measurements. Also, data quality varies among different parts of the world (Barron et al., 2014). Finally, the fairly short timespan of the existence of the OSM dataset imposes a natural limit for the traceability of the effect of plate motion in the data.

The low variance found in the average shifts of different yet close-by regions located on the same tectonic plate suggests that the effect of plate tectonic shift will become visible in the data in the long run. The average shifts of the examined areas in the OSM data differ by less than 3.1 cm/a, much less than the shifts due to plate motion do. Also, the low variance indicates that the effect of plate motion will eventually become more important than other systematic effects in the long run. This demonstrates the eventual need to update coordinate values systematically in the example of the OSM dataset. Based on the results from Sections 2 and 3, the next section discusses how to perform systematic updates for OSM data.

5.2. Updating OpenStreetMap data

In case of OSM data, the pairs of coordinates come in the form of latitude and longitude values with seven decimal places (OSM, accessed on February 4, 2019). The absolute precision in longitudinal direction thus depends on the latitude of the represented location. For most locations, the precision is around $p \approx 1$ cm. Further, the coordinate shift of the Eurasian plate is about $\sigma \approx 2.5$ cm/a in magnitude. In this section, we assume that no manual updates are performed and focus on systematic updates of data about areas on the Eurasian plate. In this case, the overall error is minimized for the next $\Delta = 50$ a if updates are performed at the interval of

$$\delta_m \approx \sqrt{\frac{1 \text{ cm} \cdot 50 \text{ a}}{2.5 \text{ cm/a}}} \approx 4.5 \text{ a.}$$

With every update, the coordinate values of geometries located on the Eurasian plate are thus shifted by about $\delta_m \cdot \sigma \approx 4.5 \text{ a} \cdot 2.5 \text{ cm/a} = 11.25$ cm. The optimal overall error for locations on the Eurasian plate can accordingly be determined as

$$\mathcal{E}_m = \frac{1}{2} \sqrt{2.5 \text{ cm/a} \cdot 1 \text{ cm} \cdot 50 \text{ a}} \approx 5.6 \text{ cm.}$$

This error is relatively small in comparison to the precision and to the shift of the coordinates. If the dataset is, however, not updated at all, the loss of spatial accuracy would be about $2.5 \text{ cm/a} \cdot 50 \text{ a} = 1.25$ m, i.e., a multiple of the optimal overall error.

OSM data is far from being static. In fact, the data is maintained by the community, and many geometries get updated manually over time. On 5 February 2019, the OSM dataset contained about 5711 000 000 nodes, of which about 787 000 000, i.e., about 13.8 per cent, have been updated with respect to their locations. Considering only those nodes that have been updated, these updates have occurred at an average interval of about 861 days, i.e., 2.36 years. Taking these updates into account, the overall error is minimized for the next $\Delta = 50$ a if updates are performed at the interval of

$$\delta_m' \approx \sqrt{\frac{\mu_{2.36/50}(0.138)}{\mu_{50/2.36}(0.138)}} \cdot 4.5 \text{ a} \approx 2.1 \text{ a.}$$

Systematic updates thus need to be performed about twice as often as when ignoring manual updates. In every update of the data, the coordinate values of locations on the Eurasian plate are shifted by about

$\delta_m' \cdot \sigma \approx 2.1 \text{ a} \cdot 2.5 \text{ cm/a} = 5.25$ cm. The optimal overall error for locations on the Eurasian plate can finally be computed as

$$\mathcal{E}_m' \approx \sqrt{\mu_{2.36/50}(0.138) \cdot \mu_{50/2.36}(0.138)} \cdot 5.6 \text{ cm} \approx 10.1 \text{ cm.}$$

The optimal overall error can thus be expected to be about two times as large as in the case of no manual updates.

6. Discussion in the light of further datasets

We have discussed our findings mostly in the light of OSM so far, but OSM is by no means the only type of georeferenced, long-term dataset with global coverage. Most authoritative mapping agencies have traditionally been curating topographic datasets originating from their professional surveying activities. These datasets represent similar thematic domains as OSM, which is why a range of studies have compared OSM to authoritative data often considering the latter as gold standard (e.g., Dorn et al. 2015, Mahabir et al. 2017, Haklay 2010, Girres and Touya 2010, Forghani and Delavar 2014). One difference with relevance to the findings obtained in this article is that authoritative datasets are usually obtained at a national scale and do thus in the majority of cases not extend to different tectonic plates. It is, however, still worth discussing authoritative geographic data in the context of this article, as exceptions exist in tectonically complex regions, e.g., the USA with respect to California, New Zealand, the Philippines, and some countries located in Central America. Indeed, since professional mapping agencies do not usually perform singular manual updates for the sake of data consistency, it is mostly the obtained results with respect to systematic updating strategies that are of interest for authoritative geographical datasets.

A range of further categories of long-term and global data beyond those pertaining to topographic mapping exist. Many datasets are collected using crowdsourcing principles (See et al., 2016; Mocnik et al., 2019) with examples including those extracted from Twitter, Foursquare (now Swarm), and Flickr. Such datasets are subject to challenges related to their long-term archiving, which is, for instance, useful to allow their spatiotemporal analysis. However, they are often curated and archived in various forms (Hogan and Quan-Haase, 2010) and across different archives. The heterogeneity attached to distributed archiving implies that, when consolidating data from various sources, data integrity and coherence shall be ensured. With respect to the updating strategies proposed in this article, there is a considerable risk that different archives that are to be merged at some point may previously have been subject to different forms of updating. It is therefore required to find means to communicate both the fact that updates have been performed and the way this was done. Clearly, a considerable metadata challenge is implied by the strategies proposed in this article when it comes to more heterogeneous data ecosystems like the ones outlined.

Another difference with crowdsourced data, especially with those coming from social media feeds, is that sometimes attached geotags are given by the platform rather than determined in-situ through GPS receivers. For example, Facebook provides predetermined coordinates for points of interest, and Twitter allows adding geocoded places to individual microblog items. When it comes to correcting for tectonic movement, it remains unclear whether those coordinates coming from unverified sources have been updated already, and whether they are aligned in terms of their date of creation. In the light of the research presented above, this means that Assumption A3 introduced for systematic updating strategies in particular can no longer be upheld in this scenario, as one often cannot know about the reference material used either for creating or updating third-party contributed coordinates. This is a major caveat with respect to the results obtained and should be addressed in future research.

The outlined issue of mixing up coordinates coming from different sources is somewhat less severe with location-based social networks. One example is Swarm, which is a service collecting check-ins made to

venues reflecting the whereabouts of people at certain points in time (Li et al., 2018). The venues are thereby given by the crowdsourced hierarchy of categories from Foursquare. Since users cannot provide arbitrary GPS locations when using these kinds of services, the issue that remains here is the diversity attached to the different venues present in the Foursquare database. This issue, however, is mitigated by the fact that the Foursquare database is still being actively extended as new venues are created by Swarm users (this is akin the folksonomy of OSM; see Mocnik et al. 2017). Beyond the fact that newly created venues behave different with respect to their locational and temporal check-in characteristics (D'Silva et al., 2018), Foursquare as an institution also acts like a mediator in the process of enriching its existing database. The company merges eventual duplicate venues and collates categories that appear very close semantically. The consequence is that any coordinate updates made may be lost, or that sudden locational changes caused by locational adjustments made by Foursquare may obscure tectonic plate motion effects. Clearly, these effects are different from the others outlined above and call for separate treatment in coordinate updating strategies.

The discussion of our theoretical results in the light of a variety of datasets makes clear that it is difficult to find a single solution to correcting coordinates for plate tectonic motion. Beyond OSM and the datasets discussed above, there are further datasets available that are both global and long-term in nature. For instance, Wikipedia articles are often georeferenced too, but their geography might well differ from the characteristics we have outlined so far. Similarly, mapping services other than but still akin to OSM, such as Google Maps, Apple Maps, and Wikimapia, may require different treatment than the one we have proposed in this article.

7. Conclusions and future research

The discussed systematic updating strategy for coordinate values in long-term global datasets corrects for the decrease in accuracy caused by tectonic plate motion. Several factors can obscure the effect of plate motion, thereby interfering with the proposed systematic updating strategy. Among these factors are manual updates the effect of which has been previously discussed in the context of systematic updates. This discussion considers only average values rather than distributions, which is to bear in mind if the manual update frequency varies across the data. Taking full distributions into consideration may render possible tailoring the update interval better to a specific dataset, but this article reveals some important principles outlining how the systematic updating strategy needs to be adapted. The strategy proposed has further been discussed in the light of OpenStreetMap and other datasets.

Future research should address a couple of challenges that are beyond the scope of this article. One such challenge is to more fully address the complexity of plate motion. Several tectonic factors impact the ways in which plates move. Some of these factors lead to shearing and rotation, resulting in heterogeneous plate velocity fields and thus leading to spatially varying plate movement directions and speeds (Conrad and Lithgow-Bertelloni, 2002). These effects, which might be of interest particularly when looking at long time periods, are not accounted for in the strategy presented. In addition, the effect of plate movement is overlaid with annual and semi-annual fluctuations in GPS measurements posed by geographic conditions (Prawirodirdjo and Bock, 2004). Future research might integrate these in follow-up adaptations.

Our discussion presented in Section 6 demonstrates that different kinds of datasets call for different updating strategies. Those strategies need to take account of the geospatial but also the usage and maintenance characteristics of specific datasets. While OSM data is typically characterized by regular data contributions interspersed with some irregularly occurring postings, this is not necessarily the case with datasets extracted from social media or authoritative sources. Adapted

updating strategies thus need to be determined to properly account for these differences and to broaden the scope of the results presented in the given article towards further kinds of data.

While we were able to infer evidence for the future need of corrections in the context of OSM, the evidence found in the data was relatively marginal. This may be largely due to the short duration of the existence of OSM data. Future research should investigate in detail more aspects of the coordinate displacement trajectories. For example, it might be beneficial to investigate the magnitudes and directions in a spatial-statistical manner in order to disclose regions of systematic shift. This could also lead to a better differentiation of different superimposed effects in the data, which could in turn make the effect of plate tectonics traceable not only in the long run but also in the foreseeable future, at least in some regions.

Further factors directly related to the datasets may overlay and thus hide the effect of tectonic plate motion, and their consideration should be subject to future research. Among these are systematic factors, e.g., extensive imports from other, external datasets and the corresponding transformation of coordinate values that are provided in reference to a local coordinate system. Further, some parts of the data in question may have been created at later points in time, which would shorten the overall timespan for which the systematic updating strategy would need to be optimized. Other factors are more unsystematic. Among these are the influence of generally poor data quality, which increases the variety and variance of manual updates in complex ways, potentially leading to bias, and the unsystematic yet simultaneous use of a variety of different data sources. The systematic updating strategy needs to be adapted to these and further factors interfering with the effect of plate motion. To properly address the unsystematic nature of many of these factors, it might be necessary to introduce individual updating strategies. Such strategies would allow to determine for each pair of coordinate values individually which updating strategy is most suitable.

CRedit authorship contribution statement

Franz-Benjamin Mocnik: Developed the main ideas, Drafted the article, Primarily responsible for Sections 1–5 and 7. **René Westerholt:** Improved the draft by providing helpful comments and critical discussions, Primarily responsible for Section 6 and 7.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We express our gratitude to M. Raifer and R. Troilo who provided their help for performing computations related to OpenStreetMap.

Funding

This work was partly supported by the DFG project *A framework for measuring the fitness for purpose of OpenStreetMap data based on intrinsic quality indicators* (FA 1189/3-1) as well as by the Heidelberg Academy of Sciences and Humanities project *Heterogeneity and convergence in shared data sources – the importance of cognitive coherence in collective decision making*.

References

- Altamimi, Z., Sillard, P., Boucher, C., 2002. ITRF2000: A new release of the International Terrestrial Reference Frame for earth science applications. *J. Geophys. Res.* 107 (B10), <http://dx.doi.org/10.1029/2001JB000561>.
- Altamimi, Z., Sillard, P., Boucher, C., 2003. The impact of a no-net-rotation condition on ITRF2000. *Geophys. Res. Lett.* 30 (2), <http://dx.doi.org/10.1029/2002GL016279>.
- Argus, D.F., Gordon, R.G., 1991. No-net-rotation model of current plate velocities incorporating plate motion model NUVEL-1. *Geophys. Res. Lett.* 18 (11), 2039–2042. <http://dx.doi.org/10.1029/91GL01532>.
- Barron, C., Neis, P., Zipf, A., 2014. A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Trans. GIS* 18 (6), 877–895. <http://dx.doi.org/10.1111/tgis.12073>.
- Boucher, C., Altamimi, Z., 2001. ITRS, PZ-90 and WGS 84: current realizations and the related transformation parameters. *J. Geod.* 75 (11), 613–619. <http://dx.doi.org/10.1007/s001900100208>.
- Conrad, C.P., Lithgow-Bertelloni, C., 2002. How mantle slabs drive plate tectonics. *Science* 298 (5591), 207–209. <http://dx.doi.org/10.1126/science.1074161>.
- Donnelly, N., Crook, C., Haasdyk, J., Harrison, C., Rizos, C., Roberts, C., Stanaway, R., 2014. Dynamic datum transformations in Australia and New Zealand. In: *Proceedings of the 1st Conference Research@Locate*.
- Dorn, H., Törnros, T., Zipf, A., 2015. Quality evaluation of VGI using authoritative data – a comparison with land use data in Southern Germany. *ISPRS Int. J. Geo-Inf.* 4 (3), 1657–1671. <http://dx.doi.org/10.3390/ijgi4031657>.
- D'Silva, K., Noulas, A., Musolesi, M., Mascolo, C., Sklar, M., 2018. Predicting the temporal activity patterns of new venues. *EPJ Data Sci.* 7 (13), <http://dx.doi.org/10.1140/epjds/s13688-018-0142-z>.
- Forghani, M., Delavar, M.R., 2014. A quality study of the OpenStreetMap dataset for Tehran. *ISPRS Int. J. Geo-Inf.* 3 (2), 750–763. <http://dx.doi.org/10.3390/ijgi3020750>.
- Girres, J.-F.c., Touya, G., 2010. Quality assessment of the French OpenStreetMap dataset. *Trans. GIS* 14 (4), 435–459. <http://dx.doi.org/10.1111/j.1467-9671.2010.01203.x>.
- Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environ. Plan. B* 37 (4), 682–703. <http://dx.doi.org/10.1068/b35097>.
- Hogan, B., Quan-Haase, A., 2010. Persistence and change in social media. *Bull. Sci. Technol. Soc.* 30 (5), 309–315. <http://dx.doi.org/10.1177/0270467610380012>.
- Intergovernmental Committee on Surveying and Mapping, Permanent Committee on Geodesy, 2018. Geocentric Datum of Australia 2020. Technical manual.
- Kreemer, C., Blewitt, G., Klein, E.C., 2014. A geodetic plate motion and global strain rate model. *Geochem. Geophys. Geosyst.* 15 (10), 3849–3889. <http://dx.doi.org/10.1002/2014GC005407>.
- Kreemer, C., Holt, W.E., 2001. A no-net-rotation model of present-day surface motions. *Geophys. Res. Lett.* 28 (23), 4407–4410. <http://dx.doi.org/10.1029/2001GL013232>.
- Li, M., Westerholt, R., Zipf, A., 2018. Do people communicate about their whereabouts? Investigating the relation between user-generated text messages and Foursquare check-in places. *Geo-spatial Inf. Sci.* 21 (3), 159–172. <http://dx.doi.org/10.1080/10095020.2018.1498669>.
- Mahabir, R., Stefanidis, A., Croitoru, A., Crooks, A., Agouris, P., 2017. Authoritative and volunteered geographical information in a developing country: a comparative case study of road datasets in Nairobi, Kenya. *ISPRS Int. J. Geo-Inf.* 6 (1), 24. <http://dx.doi.org/10.3390/ijgi6010024>.
- Mocnik, F.-B., Ludwig, C., Grinberger, A.Y., Jacobs, C., Klöner, C., Raifer, M., 2019. Shared data sources in the geographical domain—a classification schema and corresponding visualization techniques. *ISPRS Int. J. Geo-Inf.* 8 (5), 242. <http://dx.doi.org/10.3390/ijgi8050242>.
- Mocnik, F.-B., Mobasher, A., Griesbaum, L., Eckle, M., Jacobs, C., Klöner, C., 2018. A grounding-based ontology of data quality measures. *J. Spatial Inf. Sci.* 16, 1–25. <http://dx.doi.org/10.5311/JOSIS.2018.16.360>.
- Mocnik, F.-B., Raifer, M., 2018. The effect of tectonic plate motion on OpenStreetMap data. In: *Proceedings of the 21st AGILE Conference on Geographic Information Science*.
- Mocnik, F.-B., Zipf, A., Raifer, M., 2017. The OpenStreetMap folksonomy and its evolution. *Geo-spatial Inf. Sci.* 20 (3), 219–230. <http://dx.doi.org/10.1080/10095020.2017.1368193>.
- National Imagery and Mapping Agency, 2000. Department of Defense World Geodetic System 1984. Its definition and relationships with local geodetic systems.
- Neis, P., Zielstra, D., Zipf, A., 2012. The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. *Future Internet* 4, 1–21. <http://dx.doi.org/10.3390/fi4010001>.
- Neis, P., Zielstra, D., Zipf, A., 2013. Comparison of volunteered geographic information data contributions and community development for selected world regions. *Future Internet* 5 (2), 282–300. <http://dx.doi.org/10.3390/fi5020282>.
- OSM, Node. <https://wiki.openstreetmap.org/wiki/Node>.
- Prawirodirdjo, L., Bock, Y., 2004. Instantaneous global plate motion model from 12 years of continuous GPS observations. *J. Geophys. Res.* 109 (B08405), <http://dx.doi.org/10.1029/2003JB002944>.
- See, L., Mooney, P., Foody, G.M., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H.-Y., Milčinski, G., Nikšič, M., Painho, M., Pódr, A., Olteanu-Raimond, A.-M., Rutzinger, M., 2016. Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. *ISPRS Int. J. Geo-Inf.* 5 (5), 55. <http://dx.doi.org/10.3390/ijgi5050055>.
- U.S. Department of Commerce, U.S. Census Bureau, 2006. TIGER/Line files. Technical documentation.
- Zielstra, D., Hochmair, H., Neis, P., 2013. Assessing the effect of data imports on the completeness of OpenStreetMap. A United States case study. *Trans. GIS* 17 (3), 315–334. <http://dx.doi.org/10.1111/tgis.12037>.