Forensic Science International: Digital Investigation 50 (2024) 301813

Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

journal homepage: www.elsevier.com/locate/fsidi

DFRWS APAC 2024 - Selected Papers from the 4th Annual Digital Forensics Research Conference APAC

Geotagging accuracy in smartphone photography☆

Elénore Ryser [a,c,*], Hannes Spichiger [b], David-Olivier Jaquet-Chiffelle [c]

[a] *Cranfield Forensic Institute, Cranfield University, College Road, Bedford, MK43 0AL, United Kingdom*
[b] *Hochschule Luzern Informatik, HSLU I, Suurstoffi 1, Rotkreuz, CH-6343, Switzerland*
[c] *Université de Lausanne, Ecole des Sciences Criminelles, Batochime, Lausanne, 1015, Switzerland*

A R T I C L E I N F O

Keywords:
Forensic science
Geolocation error
Smartphone
Evaluation
Bayes

A B S T R A C T

After a decade of technological advancements, digital forensic science is under increasing pressure to deliver investigative findings with a high degree of scientific rigor. The judicial community has voiced growing concerns regarding digital traces and their interpretation. This research focuses on assessing the significance of geolocation information embedded within the metadata of photographs captured using a mobile phone. In order to examine the variability in the accuracy of this geolocation metadata and identify potential external influences, images were taken at 29 different locations distributed along three distinct paths. The photographs were captured using two Samsung Galaxy S8 SM-G950F devices running on Android 8.0. Various configurations of GNSS and mobile network connections were tested, and their potential impact on the accuracy of geolocation metadata was investigated. The findings show the dependency of geolocation accuracy on the specific measurement location. This research ultimately highlights the imperative for evaluative approaches to take into account the specific characteristics of each point of interest, as opposed to leaning on broad statements about the reliability of geolocation processes in general.

1. Introduction

In the contemporary landscape, where mobile phones have become an integral part of daily life, the significance of geolocation data cannot be overstated. Its application spans a multitude of domains and notably, criminal investigations. The possibility to infer the whereabouts of a device's user has established it as a source of information frequently relied upon in criminal cases. However, there is a growing emphasis on not automatically presuming inherent reliability with digital and multimedia evidence. Many studies and publications now advocate for the evaluation and expression of uncertainties associated with such types of evidence (Casey, 2019; Bosma et al., 2020; Casey et al., 2020; Spichiger, 2022). The term *evidence* is used here in a broad sense, acknowledging that the information conveyed by a trace only becomes evidence once it has been interpreted within the context of the case to which it is linked.

Turning to the precision of geolocation data linked with mobile phones, extensive attention has been directed toward this topic in various fields (transportation, public health, forestry). While these domains have conducted experiments to assess the precision of mobile phone location (as outlined in Section 2), digital forensic science has predominantly concentrated on constructing models for interpreting geolocation data recovered from phones (Tart et al., 2019; Tart et al., 2021; Bosma et al., 2020; Casey et al., 2020; Spichiger, 2023). Yet, to effectively interpret the information encapsulated in the geolocation data of a phone, understanding how external factors may influence this data is crucial. While previous studies have addressed the evaluation of mobile geolocation data under different investigative hypotheses, little has been said about the consequences of various factors influencing mobile geolocation on the entire forensic process.

The study presented in this paper endeavors to bridge this gap by furnishing empirical data on the precision of geolocation data produced by mobile phones, shedding light on its reliability in forensic investigations. It is important to note that the aim is not to be exhaustive; given the diversity of phones, softwares, and physical environments, testing all potential factors is infeasible. Instead, the focus is on understanding the variability inherent in the geolocation-related information provided by a smartphone, pinpointing critical considerations, and exploring potential approaches to interpreting and conveying this variability. We consider that, despite being produced 4 years ago, the

☆ The dataset is available on github.com/Squirrl0x00.

* Corresponding author. Cranfield Forensic Institute, Cranfield University, College Road, Bedford, MK43 0AL, United Kingdom.
E-mail address: elenore.ryser@cranfield.ac.uk (E. Ryser).

https://doi.org/10.1016/j.fsidi.2024.301813

Available online 18 October 2024

dataset remains relevant. The main purpose of the present study is to underline the presence of uncertainties as well as the importance of an informed use of mobile related geolocation data. Even if more recent phones might provide a stronger accuracy, the present study shows that, in any investigation relying on phone geolocation, accuracy should always be tested. It is critical to document the possibility of error.

The rest of the article is structured as follows: After the introduction in Section 1, existing work is presented in Section 2. The hypotheses and aims of this work are described in Section 3. Section 4 explains the experimental approach followed in this paper and the experimental results are presented in Section 5. The signification of these results is discussed in Section 6 before a conclusion is reached in Section 7.

2. Previous work

This section focuses on evaluating the accuracy of geolocation obtained through mobile phones, excluding investigations related to emergency localization conducted by operators. Additionally, studies employing mobile devices as receivers of raw GNSS data, such as precise point positioning, are not within the scope of this work. They necessitate prior application installation, a feature typically absent in most phones analyzed via forensic process.

It is important to acknowledge that the methodologies, measurements, and data processing techniques employed in these studies exhibit significant variability in terms of quality and approach, limiting the conclusions that may be gained based on them. An overview of mean reported error for studies since 2015 is shown in Table 1.

Table 1
Overview over average accuracies (rounded to the next meter) reported in studies since 2015.

Study	Environment	Device Type	Approx. Reported Mean Error
Garnett and Stewart (2015)	University Campus	iPhone 4S	7m
Schaefer and Woodyer (2015)	Seaside	iPhone 4	2m
		iPhone 5/5c	3m
		Samsung Galaxy S3 mini/S4	2m
		Sony Xperia E/P/Z	2m
Tomaštík Jr et al. (2017)	Forest	ZTE Blade	3m/7m/12m[a]
		LG G2	3m/6m/11m[a]
		Sony M4 Aqua	4m/6m/9m[a]
		Lenovo Yoga 8	3m/7m/11m[a]
Liu et al. (2018)	University Campus	HTC One (M7)	5m–200m
Merry and Bettinger (2019)	University Campus	iPhone 6	7m–13m
Yoo et al. (2020)	Diverse	iPhone (unspecified)	950m
Tomaštík et al. (2021)	Forest	LG G2	2m/8m/10m[a]
		Lenovo A5000	4m/7m/9m[a]
		Lenovo Phab 2 Pro	4m/7m/7m[a]
		Huawei P20 lite	3m/9m/7m[a]
		Xiaomi Mi8	2m/4m/5m[a]
Purfürst (2022)	Forest	Xiaomi Mi8	6m
		Xiaomi Mi8 Pro	7m
		Xiaomi Mi10 light	4m
		Huawei P20	5m
		Huawei P40	6m
		Samsung A7	6m
		Samsung S5 Xcover 4	10m
		Xcover 4s A	7m

[a] Open area/leaf-off season/leaf-on season. Aside from Yoo et al. (2020), all studies were conducted outdoors.

Table 2
Selected position accuracy measures. Adapted from Specht (2020b).

Accuracy measure	Probability	Definition
RMS	68 %	Root mean square error for ϕ, λ or d (or RMSE)
DRMS	63–68 %	Distance root mean square error for ϕ, λ or d
CEP50	50 %	The radius of circle centered at the true position, containing the position estimate with probability of 50 %
R50	50 %	The radius of circle (sphere) centered at the true position, containing the position estimate with probability of 50 %
R95	95 %	The radius of circle (sphere) centered at the true position, containing the position estimate with probability of 95 %

where ϕ – geodetic (geographic) latitude; λ –geodetic (geographic) longitude; d – radial distance.

2.1. Controlled experimental conditions

Several geolocation studies have assessed the accuracy of various mobile devices and methods, obtaining highly varying degrees of results.[1]

Liu et al. (2018) conducted a study using an HTC One (M7) smartphone in both indoor and outdoor settings. For indoor measurements, two different modes were utilized: a battery preservation one and a high accuracy one. The results demonstrated varying levels of accuracy depending on the mode and setting used for geolocation measurements. Specifically, the high accuracy mode exhibited the highest error, with an average of 215 m indoors.

Garnett and Stewart (2015) investigated the iPhone 4S and the iTouch, finding that there was a significant difference in accuracy based on the height of surrounding buildings. In areas with obstructed skies, the accuracy of geolocation tended to decrease. Additionally, the time when data collection took place (between 8 and 13h) did not significantly affect location accuracy, whereas data collected at 16h exhibited different behaviors. Weather conditions did not appear to have a significant impact on geolocation precision.

Merry and Bettinger (2019) investigated the potential impact of the season and of people present on site, potentially having an impact through WiFi usage. The season did not seem to have an impact on the accuracy of the measurements, whether done when GPS-only capability was enabled or when WiFi access was provided. The accuracy seemed to improve during the leaf-off period in the afternoon or in general when WiFi usage appeared to be higher (estimated by the number of persons in the location). Tomaštík Jr et al. (2017) compared GPS performance across different smartphones in both forested and open areas. The accuracy of location estimates varied depending on the environment (leaf-on, leaf-off, or open), with the best accuracy achieved in the open environment. They also observed differences in accuracy between different smartphone models. The study was repeated in Tomaštík et al. (2021) with another selection of smartphones. The results indicated that the accuracy varied depending on factors such as leaf coverage (leaf-on, leaf-off, or open conditions) and the capabilities of the smartphones, with devices equipped with multiple positioning systems demonstrating better accuracy. Purfürst (2022) conducted a study involving various smartphone models from Samsung, Xiaomi, and Huawei. Results concurred with Tomaštík et al. (2021), indicating that phones equipped with multiple frequencies exhibited better accuracy (Distance Root Mean Square (DRMS): 6.99 m) compared to those with a single frequency (DRMS: 9.13 m). The maximum average distance was 10.44 m, with the highest standard deviation reaching 10.19 m.

Schaefer and Woodyer (2015) analyzed a range of devices (Sony

[1] Where corresponding recent work exists, articles prior to 2015 were disregarded.

Experia E, P and Z, Samsung Galaxy S3 mini and S4 as well as iPhones 4, 5 and 5c), with measurements taken at intervals of 1 min and 3 min. The maximum absolute error observed was approximately 44 m, with a mean error of 3.49 m and a standard deviation of 3.67 m. The maximum relative error was around 45.23 m, with a mean error of 2.44 m and a standard deviation of 3.65 m. The study identified a significant difference between the Sony model with GLONASS activated and the same model without GLONASS, indicating that the latter was more precise and less variable. However, no significant difference was found based on the time elapsed during data collection.

Miluzzo et al. (2008) conducted a comparative analysis between the iPhone and iPhone 3G, highlighting discrepancies in error estimations. Zandbergen and Barbeau (2011) investigated Sanyo SCP 7050 and Motorola i580 accuracy, finding no correlation between the error estimations offered by the API and the observed error. They also observed that the locations appeared in a grid form, which they suggested might be due to rounding of measurements to the 5th decimal position by the devices. This result was confirmed by Jones et al. (2015), who observed an effect of grid sensitivity on accuracy with multiple devices.

2.2. Non-controlled conditions

Yoo et al. (2020) conducted a study to address the limitations of previous research, which were characterized by short durations and lacked proper controls. Yoo et al. (2020) collected geo-temporal data from 1464 iPhone users over a span of 3–5 months, utilizing an application that recorded the phone's position when it detected movement exceeding 500 m for a period of 5 min or less. Yoo et al. (2020) categorized the data into three groups based on reported horizontal accuracy: greater than 1000 m, between 65 m and 1000 m, and less than 65 m. The study revealed that the data exhibited a scattered and right-skewed distribution, with an average horizontal accuracy of 950 m. Furthermore, the study found that accuracy was lower in rural areas but higher in recreational areas compared to agricultural regions.

Yoo et al. (2020) offer insights into the generation of geolocation data during regular mobile device use, diverging from studies conducted in controlled experimental settings. However, it is important to note that the specific method used by iOS for calculating horizontal accuracy (as employed by Yoo et al. (2020)) remains unknown, and several studies (Zandbergen and Barbeau, 2011; Tomaštík Jr et al., 2017) have shown no correlation between this value and independently measured error. Therefore, the results of Yoo et al. (2020)'s study should be interpreted with caution.

3. Objectives and hypotheses

As pointed out in the literature review, several factors can affect the accuracy of mobile phone location data (e.g. the type of phone, its settings, and the physical environment).

In the case of geolocation metadata generated by a mobile phone, the present study's main goal is to illustrate the significant variation in the quality of information provided by such a trace. It also seeks to investigate various specific variables to evaluate their impact on the accuracy of geolocation data obtained from a mobile phone. To achieve this, the following pair of opposite hypotheses have been formulated.

H1. The radial and ϕ, λ errors[2] from the GNSS, WiFi + 4G access data have the same distribution as the errors from the 4G-only data.

H2. The radial and ϕ, λ errors from the GNSS, WiFi + 4G access data do not have the same distribution as the errors from the 4G-only data.

H3. The radial and ϕ, λ errors from the data collected in a rural area (GNSS, WiFi + 4G) have the same distribution as the errors from an urban area (GNSS, WiFi + 4G).

H4. The radial and ϕ, λ errors from the data collected in a rural area (GNSS, WiFi + 4G) do not have the same distribution as the errors from an urban area (GNSS, WiFi + 4G).[3]

H5. The radial and ϕ, λ errors from the 2G-only data, 3G-only data and 4G-only data have the same distribution.

H6. The radial and ϕ, λ errors from the 2G-only data, 3G-only data and 4G-only data do not have the same distribution.

The third pair of hypotheses (5 & 6) actually encompasses three pairs of comparisons, as each group (2G, 3G, 4G) distributions should be assessed separately against each other. However, as explained in Section 4, the employed method allows the assessment of three distributions initially. If the outcome of the first assessment contradicts the null hypothesis (H5), an additional test is then employed. This rationale is why three pairs of hypotheses are presented as one set in this context.

In the next section, an exploratory study of geolocation error across 29 diverse locations is conducted to test these pairs of hypotheses.

4. Methodology

In the following, the choice of locations, the measurement process, as well as the conducted statistical analysis is presented.

4.1. Location overview and selection

The canton of Neuchâtel is located in the central part of the Jura Mountains in Switzerland, with its northwest border adjoining France. It can be divided into three distinct geographical regions: the lakeside area along Lake Neuchâtel, the valley region consisting of two valleys at an altitude of approximately 700 m, and the mountainous region, ranging from 900 to 1065 m, characterized by a long valley. The inhabitants are distributed throughout the canton, with numerous small villages and two main settlements. Neuchâtel serves as the canton's capital and is situated near Lake Neuchâtel, while La Chaux-de-Fonds, the second largest community, is located in the mountainous region. For this experiment, twenty-nine location points are chosen, distributed within the lakeside region and the mountainous region (see Table 4 in Appendix A). The choice of the points focuses on covering a high diversity of physical environment as well as quality of network coverage. The coordinates of these points are derived by referencing a topographical map supplied by the cantonal topological service of Neuchâtel. The precision of these maps is designated as 1:250 in urban areas and 1:500 in rural areas.

4.2. Sampling design

The data collection takes place between September 2019 and February 2020. Each route is walked on foot nine times, each time on a different day, at roughly the same time (with a 1-h variation), except for the NE path where the collection time varies from morning to evening.

The data is collected using two Samsung Galaxy S8-G950F(GSM/HSPA/LTE) devices running on Android 8.0.0 (API 26). Both phones are connected to the Swiss mobile phone network through Swisscom SIM cards. The first phone, referred to as the *Standard* phone, is configured to detect WiFi networks, with GNSS services and LTE (4G) connection enabled. Android 8.0.0 allows users to manually select the preferred network generation (2G, 3G, or 4G). In cases of poor coverage,

[2] Latitude (ϕ), Longitude (λ).

[3] A location is considered as rural as soon as it is located outside a city, a village or away from a complex infrastructure (train station). Because of the locations selected for the experiment, this definition allows for a clear distinction. The locations include urban and rural canyons (places where the GNSS satellite coverage is limited).

the phone defaults to the older generation. The second phone, labeled as the *Experimental* one, is set up to prohibit WiFi connections and GNSS services. The phone network connection is adjusted based on the specific network generation being evaluated on the day of the experiment. Both phones are equipped with the Network Cell Info and Physics Toolbox Sensor applications. The Network Cell Info tracking feature is activated at the start of each collection day and deactivated at the end. During the walk, both phones are placed in the vertical interior pockets of a vest, with tracking on. When a measurement point is reached, the experimentator removes the Standard phone, waits for 30s,[4] and takes a photograph. This process is then repeated with the Experimental phone. Care is taken to ensure that the phones are held perpendicular to the ground, as previous research (Weaver et al., 2015) demonstrated the significance of the GNSS receiver's position on measurements.

After capturing the photos, the GNSS coordinates displayed by the Physics Toolbox Sensor are recorded, along with connection conditions (cellular, GNSS, WiFi), sky coverage, and weather conditions. The equipment is then stowed, and the journey continues to the next point. Each point necessitates roughly 10 min for documentation and photography. Upon completing the route, the data (pictures & Network Cell Info tracking output file) is saved on an external drive for subsequent analysis. The pictures are USB transferred without an extraction tool as only the geolocation metadata are of relevance.

A total of 29 trips are completed, with two instances requiring repetition due to adverse weather conditions.[5] Out of the originally planned 522 photographs, 497 are successfully captured. This deviation can be attributed primarily to road closures and challenges faced by the experimentator in reaching certain locations. Among these 497 photographs, a total of 348 sets of geospatial data points are gathered. For four locations (BELV, ESCA, CHOU, MAIL), fewer than five data points are obtained. As a result, these locations are omitted from the analysis concerning the influence of the physical environment on geolocation accuracy. The locations CHOU and MAIL are situated near the Swiss-French border, which limits access to the cellular network. BELV was inaccessible throughout the study period. ESCA is located in an urban setting, with the experimentator positioned atop stairs surrounded by substantial stone walls.

The geolocation metadata is extracted from each photograph. Subsequently, the coordinates are projected on a WGS84 UTM 32 system. The data are then converted from longitude/latitude to Cartesian coordinates. This conversion process is executed using the Geographiclib Python library (version 1.50), which employs a Mercator transverse projection method to convert geodesic coordinates into Cartesian coordinates. The approach for this transformation is based on Krüger's method, extended to the 4th and 6th order, as elucidated by Karney (2011). This projection method ensures that any introduced error in distance calculations remains under a centimeter. This level of accuracy is maintained as long as the coordinates are situated within 35° of the central meridian (9°E in this case). The library is subsequently employed to compute the radial distance between two sets of coordinates, as well as the North-South and East-West distances. Projections and distances are calculated by the library up to the eightieth decimal place, providing a level of precision greater than that required for this study. These results are rounded to the meter, as are the ensuing statistical calculations (standard deviation, mean, median).

4.3. Statistical analysis

4.3.1. Normality

The described sampling design enables the collection of four location-related variables in meters: radial error (d error), north error (ϕ error), west error (λ error), and angular position. The use of a normal distribution is commonly advocated and assumed in many publications for these variables. However, several publications have raised concerns regarding this assumption and suggest that the radial, ϕ and λ errors may not necessarily follow a normal distribution. For instance, Zandbergen (2008) proposes that only the ϕ error adheres to a normal distribution, and advocates for the application of a Rayleigh distribution for the radial error instead. Specht (2020b) demonstrates the normality of ϕ and λ error distributions in the context of very large samples. However, it is specified that for small sample sizes ($n < 1000$), a criteria only very rarely fulfilled in forensic applications, the distributions of ϕ and λ errors are not normal.

Moreover, Specht (2020a) illustrates that assuming a χ^2 distribution for the radial error based on the normality of ϕ and λ error distributions is not justified. Instead, Specht (2020a) shows that the radial error distribution exhibits a closer resemblance to the beta, gamma, logistic, lognormal, or Weibull distributions. In geolocation studies, two specific statistical tests appear to be frequently utilized: the Shapiro–Wilk test (Merry and Bettinger, 2019) and the Kolmogorov–Smirnov test (Specht, 2020a).

For those tests, a study demonstrated that the Shapiro–Wilk test is notably more sensitive to non-normality in the data compared to the Kolmogorov–Smirnov test. As a result, the Shapiro–Wilk test is recommended for samples with a small size (fewer than 30 measurement points) (Ahad et al., 2011). However, it is important to note that the sample size does reduce the power of the Shapiro–Wilk test, thus necessitating cautious interpretation of the results for $n < 30$ (Razali, Wah et al., 2011). Given that in the present study, certain subsets of data points intended for testing different hypotheses may have a size smaller than 30, the Shapiro–Wilk test was chosen over the Kolmogorov–Smirnov test.

4.3.2. Hypothesis testing

Statistical tests are employed to assess the significance of differences among various subsets of data, created for the purpose of testing pairs of opposite hypotheses (see Table 3). These subsets are created by isolating data according to specific variables, thereby limiting variation to only those variables being tested. For all subsets, the radial error must be less than 1847 m. For the subset used to test the first set of hypotheses (H1, H2), only data collected with a 4G connection is considered. For the subset used to test the second set of hypotheses (H3,H4), only data collected with the standard phone and with a 4G connection are considered. For the subset used to test the third set of hypotheses (H5, H6), only data collected with the Experimental phone with 2G, 3G, 4G connection are considered.

The normality tests consistently indicate that the distributions of these subsets deviate from a normal distribution (see Table 5 in Appendix A). Consequently, non-parametric tests are deemed necessary for the subsequent statistical analysis.

Table 3
Positional radial error of Samsung Galaxy S8 SM-G950F by using different network configurations.

	n	Min [m]	Max [m]	RMSE
All data	315	2	27259	599.17
Experimental	174	7	27259	716.03
Standard 4G	141	2	11533	411.50
Stand. 4G Urban	54	2	300	87.00
Stand. 4G Rural	87	3	11533	519.36
Experimental 2G	52	40	21935	827.43
Experimental 3G	58	7	27259	550.13
Experimental 4G	64	43	27024	750.58

Standard is always connected to 4G,GNSS and Experimental varies between 2G, 3G, 4G, no GNSS.
Errors >1847m removed for n, RMSE.

[4] This wait is added in the methodology because of the time it takes sometimes for the phones to fix an antenna. The phones are always on and operative.
[5] Heavy snowfalls. The temperatures prevented the procedure to be applied correctly.

The Mann–Whitney test is a commonly employed statistical method in numerous geolocation-related studies (Zandbergen and Barbeau, 2011; Abdi et al., 2012; Moriarty and Epps, 2015; Merry and Bettinger, 2019). In this study, it is systematically implemented at a 95 % confidence level. This test is used to test the first two pairs of hypotheses (H1, H2) (H3,H4).

The Mann–Whitney test is not suitable for simultaneous comparisons of multiple subsets as required by hypotheses H5 and H6. An analogous situation is encountered in the study by Tomaštík Jr et al. (2017), where a Kruskal–Wallis test is employed. The same is done in this study. In cases where the null hypothesis (indicating no population differences) is rejected by the Kruskal–Wallis test, the Conover–Inman test is selected for further analysis.

Visual observations of the results suggest that antenna proximity and the quality of cellular connection might impact the dispersion of measurements at a particular location. To verify this, the Kendall correlation test is used. This test is chosen due to its minimal assumptions (requiring ordinal data and a monotonic relationship) and its previous use in Merry and Bettinger (2019). Antenna locations were logged during geolocation measurements and cross-referenced with the publicly accessible OpenCell and Mozilla Location Service (MLS) databases. In instances where cross-referencing with OFCOM (Federal Office of Communications - Switzerland) data is feasible, it is assumed that the antennas are the same if the locations appeared to correspond. The identification number of the antenna and the connection quality are recorded using the Network Cell Info application.

4.3.3. Error measuring

In this study, the quantity and nature of available measurement points constrain the range of applicable analytical tools (see Table 2). Consequently, metrics such as 2DRMS and CRLB, along with many CEP calculation methods, are excluded.

A variety of methods are available for calculating Circular Error Probability (CEP). The selection of the most appropriate method depends on the experimental context and the available data. Various studies have been conducted to compare different equations and provide recommendations for their applicability in specific contexts (Williams, 1997; Yakimenko, 2013; Wang et al., 2014; Carlson and Beer, 2021).

Given the experimental conditions, it was decided to customize the selection of the calculation method based on the available data. To compute the 50 % CEP, the Ethridge method is preferred due to its independence from the type of distribution, provided there are more than three data points available and the bias is less than 0.75 times the standard deviation of the radial error. If the number of data points is less than three, the Rand method is favored. The computation of the CEPs are made using the R library shotGroups.

The Root Mean Square Error (RMSE) is a commonly used metric in GNSS studies, though it is frequently employed with an assumption of a normally distributed error. Zandbergen (2008) advises a cautious approach, suggesting the removal of 5–10 % of the most significant outliers to account for potential non-normality in the distribution. Following John Tukey's method (1) (Appendix B), 10 % of the most extreme data points are excluded prior to conducting all statistical tests and accuracy assessments.

To evaluate the Circular Error Probability at 50 % (CEP50) and RMSE measurements, we compare them to the 50th and 68th percentiles (R50, R68) of each data subset. The percentiles are computed using formula (2) (Appendix B), as discussed in Hyndman and Fan (1996).

5. Results

5.1. Radial error

Taking into account all data points, the maximum radial error for the Standard and Experimental phones was 11′533 m and 27′259 m, respectively (Table 3). The visual inspection of the data distribution indicated a deviation from normality, which was corroborated by the Shapiro–Wilk normality test result (Table 5 in Appendix A).

The data analysis uncovers groups of photos sharing identical geo-metadata. 172 data points can be distributed across 65 groups. Out of these, 28 groups consist entirely or partially of duplicates, where the data points occur consecutively in the measurement time sequence, possibly due to a delayed geo-location update. The remaining 37 groups involve measurements taken at the same location on different days or at different locations and days, deviating from the delayed update hypothesis. This pattern is observed in most locations (26 out of 29), each affected to varying degrees, with some exhibiting important loss of variability (ARVR, EGAD, VREV, CRET, FERM, GACH, MOCH and PORT). Notably, the Experimental phone and 3G connections show a higher frequency of identical data points compared to the Standard phone, 2G, and 4G connections. Additionally, four specific days (out of the 27 field days) account for one third of the duplicated measurements, and 33 outliers with unusually large radial distances have been identified in this dataset.

5.1.1. Comparing the impact of satellite access (H1; H2)

Subsequently, to test the previously established hypotheses, specific subgroups of data are curated to minimize data variability, as described in the methodology. Focusing exclusively on data collected with a 4G connection, a comparison using a Mann–Whithney test is made between the error distributions of the Experimental and Standard phones. This analysis reveals a significant difference ($p < 0.05$), underscoring the impact of WiFi or satellite availability on geolocation accuracy (Table 6 in Appendix A). Additionally, the RMSE of the Standard Phone (4G) is lower than that of the Experimental phone (4G) (Table 3).

5.1.2. Comparing urban and rural locations (H3; H4)

A comparison is conducted between error distributions from rural and urban locations, using data coming from the Standard phone with a 4G connection. This comparison also yields a significant difference ($p < 0.05$) (Table 6 in Appendix A), suggesting that geolocation measurements in urban settings substantially vary from those in rural environments. When focusing solely on data from the Standard phone, a noticeable improvement in RMSE is observed for urban locations in contrast to rural ones (Table 3). This tendency is illustrated in Fig. 1, where the locations are arranged based on the median of the radial errors observed across all recorded data points.

5.1.3. Comparing network generation (H5; H6)

Employing the Kruskal–Wallis test in tandem with the Conover–Inman method on data points from the Experimental phone, a difference is identified between data collected with 2G and 3G connections ($p < 0.025$) (Table 6 in Appendix A). However, this result is very sensitive to outlier identification, potentially impacting its reliability. In terms of accuracy, it appears that the RMSE of data collected with a 3G connection is lower compared to data collected with 2G and 4G connections.

Using the Experimental dataset, a Kendall correlation test examines the relationship between radial error and network reception strength. Reception strength seems to show an inverse correlation with radial error, implying lower error with better reception. However, the

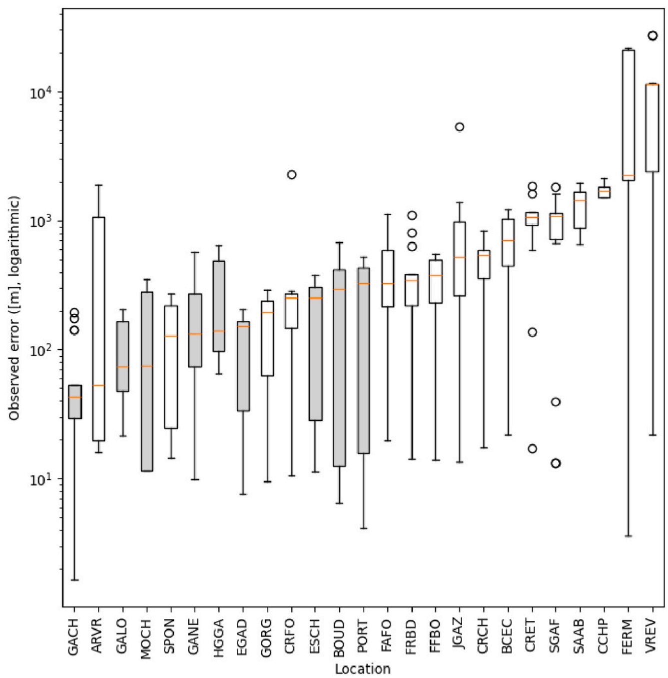

Fig. 1. Boxplot of observed distances [m] from the location the device was actually at per location (Logarithmic scale, ordered by median). Urban locations are shown in grey, rural locations in white.

correlation is weak (R = −0.24, p = 0.0061). The same conclusion is reached when comparing phone subgroups by type of network generation.

5.1.4. Additionnal observations

Despite the limited number of data points, analyzing the location-specific data without any restrictions (related to phone, networks, or outliers) reveals distinct patterns based on where the photographs were taken. Specifically, when examining the RMSE values for each location, significant variations emerge, with some locations exhibiting an RMSE between 10 and 50 m, while others demonstrate an RMSE exceeding 500 m. The CEP50 calculations exhibit a similar diversity from one location to another (Table 7 in Appendix A). When conducted after removing outliers, the results reveal projected radii ranging from 25 m to 555 m, depending on the location.

The comparison between CEP50 and Q50 metrics reveals varying levels of data dispersion across different locations. Some locations demonstrate a wider spread of data, while others have a more concentrated distribution around the median. Locations like SGAF, FFBO, CRFO and PORT have notably high relative error percentages, indicating substantial differences between CEP and Q50. Upon closer examination, it becomes evident that certain locations, like CRFO, SGAF, and FFBO, contain two distinct groups of data points: one with minimal radial error and another with substantial radial error. This considerable variation in data dispersion may contribute to the observed behavior of CEP50 in these locations. Concerning PORT, the presence of identical measurement (4 out of 6) might be the explanation of the highly skewed result. The CEP50 metric seems to be as sensible to outliers as the RMSE meaning that it might also not be an appropriate metric when the measurements are in small numbers or the location tend to present a high variation.

5.2. Angular orientation

When analyzing the data from both the Standard and Experimental phones, a relatively uniform distribution of angular orientation is observed. This suggests that the measuring devices (i.e., the phones) do not significantly impact the orientation of the recorded positions. However, a distinct trend emerges when examining data from individual locations: in slightly over two-thirds of cases, geolocations exhibit a clear orientation towards one side of the compass. This phenomenon is particularly noticeable in some locations, indicating a potential influence from both the physical environment and the positions of available antennas (Fig. 2).

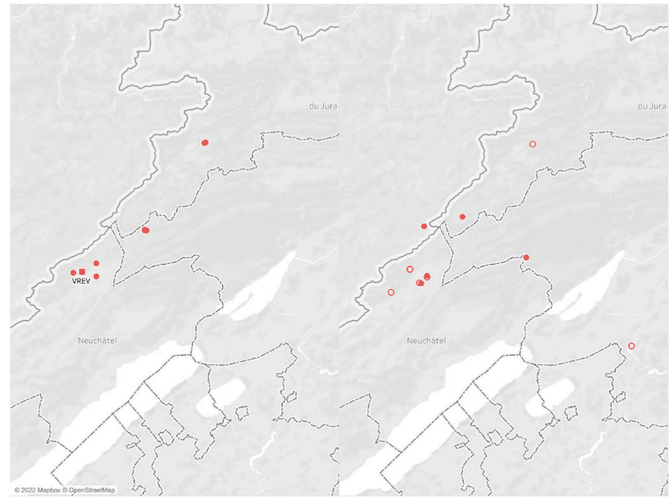

Fig. 2. Positions of the measurements (on the left) and the antennas the phone was connected to (on the right). The square represents the real position where the pictures were taken. The accuracy of the antennas position depend on the data collected within the Opencell and MLS databases. Location: VREV.

Using the Experimental dataset, a Kendall correlation test examines the relationship between radial error and antenna distances. Antenna distances refer to the separation between the indicated photo metadata location and the corresponding antenna location in established databases. Notably, a correlation is observed between the 3G connection subgroup and the antenna reported in the MLS database, with radial error increasing as antenna distance decreases (R = 0.39, p = 0.00025). However, as described in the methodology, the location of antennas is uncertain (it relies on assumptions made during the different cross-referencings). Those results need to be taken cautiously and the relation between the position of the antenna and the radial error of mobile phone geolocation needs to be further investigated. This does not negate the observation that angular orientation appears to be influenced by external factors.

6. Discussion

While this study illustrates the presence of large location errors (>2 km), not acknowledged in previous studies, it also demonstrates the variance of errors based on specific locations. While it is difficult to pinpoint the variables responsible for such variance in position error, the present study tested three hypotheses, which are discussed in section 6.1.

Although variance in geolocation error might have been documented in separate studies (for instance studies investigating urban canyons), the present results facilitate a discussion on forensic investigation protocols, as proposed in section 6.2.

Being able to communicate clearly the results of a forensic investigation needs communication tools. In the case of geolocation, statistics, scales and figures might be used to report the geolocation indicated by a mobile phone as well as to contextualise this information with the uncertainties that surround it. In section 6.3, we draw on our results to discuss the precautions that should be taken when using various previously illustrated tools (hypothesis tests, statistical metrics, graphs).

6.1. Impact of different variables: discussing the hypotheses

This study aims to assess three pairs of opposite hypotheses, with the first two of them focusing on the impact of network accessibility (GNSS, WiFi, Cellular Network) on geolocation, while the third one investigates the influence of urbanization on geolocation accuracy.

While the evaluation of these hypotheses through statistical tests was somewhat constrained by the presence of multiple duplicates in the data points, visual assessments tended to confirm the results of the statistical tests, as well did the comparisons between different metrics (minimum, maximum, median, mean, RMSE).

6.1.1. Impact of WiFi/GNSS accessibility (H1; H2)

The comparison of geolocation errors between measurements with WiFi/GNSS/4G access and those with 4G-only access reveals a significant difference, highlighting the influence of WiFi or satellite availability on geolocation accuracy. This significant difference is demonstrated using a Mann–Whitney test and a visual analysis of the data (scatter plots and bar charts). This finding tends to align with previous research by Merry and Bettinger (2019). However, differences in methodologies (such as device, location, and type of networks) make direct comparison difficult. Additionally, in the present study, WiFi access is limited in most locations, complicating the assessment of WiFi's impact on geolocation accuracy.

Nonetheless, there is a significant disparity between the range of radial errors reported in past studies and the values collected in the present experiment. Most previous studies report errors of less than ten meters, or at most, errors less than five hundred meters. In the present study, the maximum recorded radial error is 27 km (without satellite access) and 11 km (with satellite access), with median errors of 430 m and 42 m, respectively. This discrepancy with past literature underscores the need for precision and accuracy tests of devices involved in an investigation, without assuming a specific level of accuracy or precision.

The presence of WiFi/GNSS access does not seem to impact the angular orientation of the geolocation errors. However, the same mitigating factors that affect the discussion of radial errors need to be considered, preventing any conclusion to be reached.

6.1.2. Comparing urban and rural locations (H3; H4)

Using visual analysis (scatter plots and bar charts) as well as a Mann–Whitney test, the comparison between urban and rural locations within the 4G connection subset indicates a significant difference in error distributions. Results suggest that geolocation measurements in urban settings tend to be more accurate than those in rural environments, corroborating the findings of Yoo et al. (2020). Fig. 1 however illustrates that the distinction is not important. This could be due to (1) the definition used to differentiate urban areas from rural areas, or (2) other influences or variables that were not controlled in the experiment, such as antenna presence or the frequency of human activity in the location.

Errors observed in rural areas also exhibit greater dispersion than those in urban areas. However, the sample of locations in urban areas is smaller than that in rural areas. It is possible that the smaller number of input points impacts the statistical metrics. Additionally, the sampled urban locations may not represent areas that could significantly impact location precision. Indeed, not only accuracy, but also precision seems to be highly variable and depends on the location of the measurement. For the subset Standard Phone, 4G, not taking into account radial errors over 1847 m, RMSE (an accuracy measure) varies from 4 m (PORT) to 907 m (SGAF) and standard deviation varies from 11 m (ESCH) to 584 m (SGFA). However, since the number of measures at each location is limited, it is difficult to make any generalisation. Taking that into consideration, four categories can be roughly draft: places without any geolocation metadata, places with low variability (high precision, regardless of accuracy), places with high variability (low precision, regardless of accuracy) and places where geolocation metadata record extreme behaviors (high accuracy error).

In addition to the direct physical environment, the potential impact of antennas and geolocation databases (in this study, the Google database, measurements were made on an Android device) can vary from place to place. However, the number of data points per location is insufficient for any sort of correlation analysis. Further studies on the possible implications of neighboring antennas and the role of databases in cellphone geolocation would be valuable. Similar considerations apply when interpreting angular error.

6.1.3. Comparing network generation (H5; H6)

For this last set of hypotheses, the combined use of the Kruskal–Wallis and Conover–Iman tests highlights a difference between data collected while the phone was connected to a 2G cellular network and data collected with a 3G connection. However, while slight differences between the 3G subset and the 2G or the 4G subsets might be visible with the use of bar charts, the results of the statistical test are sensitive to slight changes in the outlier limit. Therefore, cautious interpretation of the variations in accuracy between measures taken on 2G, 3G, or 4G networks appears challenging, suggesting that the observed differences may be the result of random elements rather than an actual difference in accuracy.

Nonetheless, a difference emerges when comparing the high number of pictures without geolocation metadata taken while being connected to a 2G network and pictures taken while being connected to a 3G or a 4G network. It is difficult to pinpoint the reason behind this difference. The authors could not find previous studies testing the possible influence of cellular network generations on geolocation error.

6.2. Using geolocation in a case

The results obtained in this study show a larger degree of variance and higher errors than previous studies. Whilst the variance may be a result of a smaller sample size, the origin of the difference in observed error is still unknown. Potential reasons might be the mountainous terrain, specific device settings or the measurement protocol using photographs instead of an application running on the device. With the information currently available, there is no good way to discern between these potential explanations. However, these observations illustrate the risk associated with unconditionally trusting evidentiary localisations. No matter the reason of the higher observed error, this data shows that mobile devices can sometimes produce highly inaccurate geolocation traces.

Acknowledging this does not allow to conclude that geo-location traces cannot be used due to them being unreliable. Instead, it emphasises their use as one element of many. Indeed, it is rare to have only one single piece of geolocation information available in a case. Often, practitioners are presented with a series of localisations, generally with varying accuracy and precision. The ensemble of these data points generally makes it possible to identify locations with large error as evident outliers. These other available traces act as a safeguard. Typically, it is more reasonable to consider one outlier rather than assuming that all other traces are affected by large errors. Consequently, it is unlikely that errors as extreme as those observed in this study would lead to erroneous conclusions. However, caution must be exercised when an argument about location is based purely on a singular observation. In such cases, none of these safeguards are present, and the risk of an erroneous conclusion increases.

The observations presented here align with a series of previous studies that question the absolute reliability of mobile device locations (von Watzdorf and Michahelles, 2010; Rodriguez et al., 2018; Merry and Bettinger, 2019). The implication for localisations in criminal cases is clear: Categorical statements about the location of a device are not supported by existing research and should not be presented in court. Instead, more balanced approaches, integrating uncertainties, should be

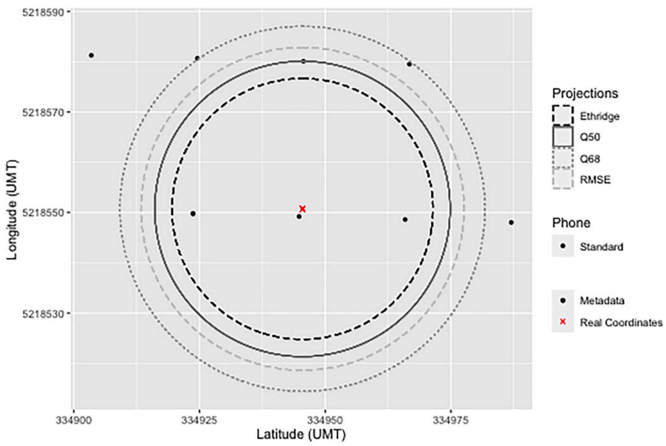

Fig. 3. Accuracy metrics projection for GACH. The grid-like positions illustrates the rounding of measurements to the 5th decimal.

chosen. So far, Bayesian evaluation is the only approach that has been proposed in this regard (Casey et al., 2020; Spichiger, 2022, 2023). This approach consists in evaluating the observed trace from the perspective of each proposed explanation of the trace. For each suggested location, the likelihood of the observations is explored. Ideally based on reference data, a probability is assigned to the statement *These location-traces were recovered from the device given its presence at location A*. This process is then repeated for each other suggested location, allowing for a differentiated conclusion to be presented. In Spichiger (2023), the feasibility of conducting such an evaluation for a single point location is shown. In this study the process of data generation is approached as a black box process. For each location of interest, an equivalent device is used to generate a series of reference data points. This approach allows to measure where a device is localised when it is at this given location (Spichiger, 2023).

This approach requires on-site measurements for each case anew, which can be resource-intensive. Developing a generalised model for localisation accuracy could lead to significant efficiency gains for practitioners. Currently, there are no suggestions on the potential structure or approach for such a model. However, the results presented here highlight potential challenges for developing such a model. As can be seen in Fig. 1, the distance between the actual location of the device and the coordinates recovered from the images vary quite heavily from one location to the other. The observed medians differ by two orders of magnitudes from the lowest to the highest. The spread of the quartiles vary heavily and the form of the distribution includes all from relative symmetry to being heavily skewed in either direction. This implies highly variable parameters that a model would need to account for and may suggest that different distribution models might best describe the observations at different locations. If other factors, such as different devices or environmental changes, are also considered, the complexity of a generalised model will increase even more. This complexity could reach a degree where generalised modeling is no longer feasible, leaving black box measurements as the only viable option.

6.3. Choice of results presentation

Three types of measurements are employed in this study to describe radial error: RMSE, CEP50, and the 50th and 68th percentiles. Comparing RMSE with the 68th percentile reveals differences in most locations, except for MOCH and JGAZ. However, in the case of MOCH, where only four data points were observed, all of the observations are identical, which distorts the calculations. Notably, in subsets like ARVR, PORT, and CRFO, the RMSE exceeds the Q68 considerably. This indicates that a substantial portion of data points in these subsets exhibit larger errors than suggested by the Q68. It implies the presence of outliers or instances of exceptionally large errors within these specific subsets, despite the defined outlier threshold of 1847 m.

Conversely, for other locations such as EGAD, CRET, and FFBO, the RMSE is significantly lower than the Q68. This suggests the presence of a subgroup displaying higher accuracy within these subsets. A closer scrutiny of the data reveals distinct clusters of data points that exhibit superior accuracy compared to the remaining subset, underlining the variability in measurements within these locations. Finally, in certain locations, there is a slight difference between the RMSE and the Q68, with the RMSE being marginally lower (indicating a relative error of approximately 10–15 %). In these cases, the RMSE appears to provide a representative indication of the radial error behavior. These results are coherent with those of Zandbergen (2008, 2009) acknowledging that in some cases the RMSE metric might not be representative of the error distribution.

Reporting the Circular Error Probable (CEP) values to the 50th percentile reveals that in most cases, the relative error is greater than 5 %, with the median absolute error at 38 m. Overall, the CEP appears to overestimate the accuracy of the measuring device (in this case, a mobile phone). However, this behavior varies from one location to another, and we are unable to pinpoint an explanation.

These results might be influenced by the number of measurement points. However, it seems reasonable to suggest that characterizing the geolocation error of a mobile device should not rely solely on a single statistical tool (such as RMSE or CEP). Including graphical representations allows for an illustration of both the accuracy and precision of a specific device in a specific location (see Fig. 3). A report that limits the evaluation of accuracy and precision to one metric or one type of communication media might not be the most effective way to convey that information to a third party.

7. Conclusion

The use of localisations recovered from mobile devices is nowadays central to many investigations and frequently used in court. The recent rise in awareness for errors and uncertainty linked to these traces underlined the need for a better understanding of the factors influencing the observed results and for means to quantify the observed uncertainties.

In this study, localisations are obtained at 29 different locations, over a time span of six months with the settings of the devices varying. The measurements reveal a large variety of precision in the data and suggest an impact of the urbanisation of the surroundings of the measurement location. Overall, distances between observed locations and the actual position of the device are found to be larger than observed in previous studies without it being clear where this difference comes from. The largest observed difference was over 11 km in standard conditions and more than 27 km when only based on visible cell towers (referred to as the *experimental* setting). The obtained results highlight the critical importance of considering the uncertainties associated with geolocation data obtained from mobile phones, as well as their high complexity. It is crucial to recognize that while mobile phones offer a valuable source of information, this data can be complex and prone to potential inaccuracies. The revealed significant variability suggests that finding a simple, universal model for evaluating location data is unlikely. A nuanced approach is needed for interpreting and utilizing this information in legal proceedings and for basing conclusions on field measurements.

It is worth noting that this research is not exhaustive, and there remain many unexplored factors that may impact geolocation data.

Acknowledgements

For the purposes of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Accepted Author Manuscript version arising from this submission.

A Appendix.

Table 4
Measurement points

Location	Latitude	Longitude	Environment
ARVR	47°02′14.53″	6°52′29.68″	Rural
BCEC	47°03′43.56″	6°44′42.82″	Rural
BELV	46°59′53.42″	6°55′36.90″	Urban
BOUD	47°01′40.02″	6°53′16.31″	Urban
CCHP	47°04′18.65″	6°44′59.19″	Rural
CHOU	47°05′41.92″	6°45′21.02″	Rural
CRCH	47°00′26.37″	6°54′13.62″	Rural
CRET	47°06′22.48″	6°48′45.42″	Rural
CRFO	47°00′41.12″	6°54′22.53″	Rural
EGAD	46°59′40.00″	6°56′07.36″	Urban
ESCA	47° 05′ 04″	6° 45′ 18″	Rural
ESCH	46° 59′ 31″	6° 55′ 38″	Urban
FAFO	47°03′22.50″	6°52′59.42″	Rural
FERM	47° 04′ 49″	6° 45′ 16″	Rural
FFBO	47°00′52.40″	6°54′30.31″	Rural
FRBD	47°01′10.60″	6°54′16.34″	Rural
GACH	47°05′56.05″	6°49′30.03″	Urban
GALO	47°03′27.31″	6°44′46.18″	Urban
GANE	46°59′46.15″	6°56′07.59″	Urban
GORG	46°59′22.81″	6°54′41.64″	Rural
HGGA	47°02′52.05″	6°52′29.25″	Urban
JGAZ	47°03′02.65″	6°53′12.61″	Rural
MAIL	47°06′11.13″	6°47′11.77″	Rural
MOCH	46°59′30.37″	6°55′42.93″	Urban
PORT	46°59′22.88″	6°56′09.91″	Urban
SAAB	47°04′12.28″	6°45′03.81″	Rural
SGAF	47°03′21.22″	6°52′46.54″	Rural
SPON	46°59′24.47″	6°54′32.02″	Rural
VREV	47°06′15.30″	6°48′14.21″	Rural

Table 5
Results of the Shapiro–Wilk Normality Test on radial error from *Experimental* (Exp.) and *Standard* (Std.) phones

Subsets	p-value
Exp. Phone 4G	1.55733e-06
Std. Phone 4G	3.38884e-17
Exp. Phone 2G	7.805305e-05
Exp. Phone 3G	6.107828e-08
Exp. Phone 4G	1.55733e-06
Std. Phone 4G Urban	2.170208e-08
Std. Phone 4G Rural	5.408977e-11
Limit value	1847m

Table 6
Results of hypotheses test on Radial error from *Experimental* (Exp.) and *Standard* (Std.) phones

Mann–Whitney	p-value
Standard 4G/Experimental 4G	5.547026e-13
Rural Std 4G/Urban Std 4G	0.002223755
Kruskal–Wallis	p-value
Experimental 2G/3G/4G	0.02960179
Conover–Inman	p-value
Experimental 2G/3G	0.004208075
Experimental 2G/4G	0.133197728
Experimental 3G/4G	0.049376780
Limit value	1847m

Table 7
Comparison of accuracy measures in meters (Per location, Standard, 4G). Locations with less then 4 datapoints do not appear.

Loc.	n	CEP 50 ± 0,2	Q50	RMSE	Q68
ARVR	9	81.13 ± 1.62	20.43	225.32	53.30
BCEC	5	262.27 ± 5.25	202.59	548.68	380.84
CRCH	4	313.48 ± 6.27	270.10	490.02	606.43
CRET	4	555.65 ± 11.11	592.91	744.46	1048.27
CRFO	4	88.09 ± 1.76	11.20	132.67	82.60
EGAD	5	19.33 ± 0.39	7.59	22.10	32.65
ESCH	6	23.30 ± 0.47	27.29	27.24	29.96
FAFO	6	153.98 ± 3.08	97.92	223.62	247.01
FFBO	8	173.23 ± 3.47	286.56	329.68	438.65
FRBD	5	321.64 ± 6.43	350.72	505.48	617.35
GACH	9	25.95 ± 0.52	29.38	32.12	36.31
GALO	7	53.55 ± 1.07	55.82	91.56	72.43
GANE	5	61.07 ± 1.22	66.31	82.68	76.14
GORG	6	117.60 ± 2.35	29.90	156.87	190.31
HGGA	7	64.55 ± 1.29	98.30	118.69	135.87
JGAZ	6	306.25 ± 6.13	248.28	505.20	506.90
MOCH	4	–	11.48	11.48	11.48
PORT	6	42.72 ± 0.85	4.14	79.89	18.96
SGAF	8	535.49 ± 10.71	1104.92	907.08	1119.75
SPON	5	31.71 ± 0.63	16.40	46.38	48.59

Errors >1847m removed; all CEP computed with Ethridge; MOCH is not computable.

B Equations

$$IQR = \widehat{X}_{0.75} - \widehat{X}_{0.25} \tag{1}$$

$$\text{Upper Limit} = \widehat{X}_{0.75} + 1.5 \times IQR \text{Lower Limit} = \widehat{X}_{0.25} \ - 1.5 \times IQR \text{With: } \widehat{X}_p\text{, estimation of } p\text{th percentile}$$

$$m = \frac{(p+1)}{3.p_k} = \ \frac{k - 1/3}{n + 1/3} p_k \approx median[F(x_k)] \tag{2}$$

References

Abdi, E., Sisakht, S.R., Goushbor, L., Soufi, H., 2012. Accuracy assessment of gps and surveying technique in forest road mapping. Ann. For. Res. 309–317.

Ahad, N.A., Yin, T.S., Othman, A.R., Yaacob, C.R., 2011. Sensitivity of normality tests to non-normal data. Sains Malays. 40, 637–641.

Bosma, W., Dalm, S., van Eijk, E., el Harchaoui, R., Rijgersberg, E., Tops, H.T., Veenstra, A., Ypma, R., 2020. Establishing phone-pair co-usage by comparing mobility patterns. Sci. Justice 60, 180–190. https://doi.org/10.1016/j.scijus.2019.10.005. URL: https://www.sciencedirect.com/science/article/pii/S1355030619300942.

Carlson, J.J., Beer, J.N., 2021. Performance Comparison of Circular and Spherical Error Probable Estimators. Technical Report. Sandia National Lab.(SNL-NM), Albuquerque, NM (United States).

Casey, E., 2019. The chequered past and risky future of digital forensics. Aust. J. Forensic Sci. 51, 649–664. https://doi.org/10.1080/00450618.2018.1554090 publisher: Taylor & Francis _eprint.

Casey, E., Jaquet-Chiffelle, D.O., Spichiger, H., Ryser, E., Souvignet, T., 2020. Structuring the evaluation of location-related mobile device evidence. Forensic Sci. Int.: Digit. Invest. 32, 300928 https://doi.org/10.1016/j.fsidi.2020.300928. URL: https://www.sciencedirect.com/science/article/pii/S2666281720300238.

Garnett, R., Stewart, R., 2015. Comparison of gps units and mobile apple gps capabilities in an urban landscape. Cartography Geogr. Inf. Sci. 42, 1–8.

Hyndman, R.J., Fan, Y., 1996. Sample quantiles in statistical packages. Am. Statistician 50, 361–365.

Jones, T.W., Marzen, L., Chappelka, A., 2015. Horizontal accuracy assessment of global positioning system data from common smartphones. Papers in Applied Geography 1, 59–64.

Karney, C.F., 2011. Transverse mercator with an accuracy of a few nanometers. J. Geodesy 85, 475–485.

Liu, J.C., Lin, C.H., Lin, C.Y., 2018. Accuracy evaluation of location service modes on mobile devices. Sensor. Mater. 30.

Merry, K., Bettinger, P., 2019. Smartphone GPS accuracy study in an urban environment. PLoS One 14, e0219890. https://doi.org/10.1371/journal.pone.0219890.

Miluzzo, E., Oakley, J.M., Lu, H., Lane, N.D., Peterson, R.A., Campbell, A.T., 2008. Evaluating the iphone as a mobile platform for people-centric sensing applications. In: Proceedings of UrbanSense, vol. 8.

Moriarty, K.M., Epps, C.W., 2015. Retained satellite information influences performance of gps devices in a forested ecosystem. Wildl. Soc. Bull. 39, 349–357.

Purfürst, T., 2022. Evaluation of static autonomous gnss positioning accuracy using single-, dual-, and tri-frequency smartphones in forest canopy environments. Sensors 22, 1289.

Razali, N.M., Wah, Y.B., et al., 2011. Power comparisons of shapiro-wilk, Kolmogorov-smirnov, lilliefors and anderson-darling tests. Journal of statistical modeling and analytics 2, 21–33.

Rodriguez, A.M., Tiberius, C., Bree, R.v., Geradts, Z., 2018. Google timeline accuracy assessment and error prediction. Forensic Sciences Research 3, 240–255. https://doi.org/10.1080/20961790.2018.1509187 publisher: Taylor & Francis _eprint.

Schaefer, M., Woodyer, T., 2015. Assessing absolute and relative accuracy of recreation-grade and mobile phone gnss devices: a method for informing device choice. Area 47, 185–196.

Specht, M., 2020a. Consistency of the empirical distributions of navigation positioning system errors with theoretical distributions—comparative analysis of the dgps and egnos systems in the years 2006 and 2014. Sensors 21, 31.

Specht, M., 2020b. Statistical distribution analysis of navigation positioning system errors—issue of the empirical sample size. Sensors 20, 7144.

Spichiger, H., 2022. The Evaluation of Mobile Device Evidence under Person-Level, Location-Focused Propositions. Doctorate Thesis. University of Lausanne, Lausanne.

Spichiger, H., 2023. A likelihood ratio approach for the evaluation of single point device locations. Forensic Sci. Int.: Digit. Invest. 44, 301512 https://doi.org/10.1016/j.fsidi.2023.301512. https://www.sciencedirect.com/science/article/pii/S2666281723000136.

Tart, M., Brodie, I., Patrick-Gleed, N., Edwards, B., Weeks, K., Moore, R., Haseler, R., 2021. Cell site analysis; use and reliability of survey methods. Forensic Sci. Int.: Digit. Invest. 38, 301222 https://doi.org/10.1016/j.fsidi.2021.301222. URL: https://www.sciencedirect.com/science/article/pii/S266628172100130X.

Tart, M., Pope, S., Baldwin, D., Bird, R., 2019. Cell site analysis: roles and interpretation. Sci. Justice 59, 558–564. https://doi.org/10.1016/j.scijus.2019.06.005. http://www.sciencedirect.com/science/article/pii/S1355030618303538.

Tomaštík, J., Chudá, J., Tunák, D., Chudý, F., Kardoš, M., 2021. Advances in smartphone positioning in forests: dual-frequency receivers and raw gnss data. Forestry: Int. J. Financ. Res. 94, 292–310.

Tomaštík Jr, J., Tomaštík Sr, J., Saloň, Š., Piroh, R., 2017. Horizontal accuracy and applicability of smartphone gnss positioning in forests. Forestry: Int. J. Financ. Res. 90, 187–198.

Wang, Y., Yang, G., Yan, D., Wang, Y., Song, X., 2014. Comprehensive assessment algorithm for calculating cep of positioning accuracy. Measurement 47, 255–263.
von Watzdorf, S., Michahelles, F., 2010. Accuracy of positioning data on smartphones. In: Proceedings of the 3rd International Workshop on Location and the Web. Association for Computing Machinery, New York, NY, USA, pp. 1–4. https://doi.org/10.1145/1899662.1899664. URL: https://dl.acm.org/doi/10.1145/1899662.1899664.
Weaver, S.A., Ucar, Z., Bettinger, P., Merry, K., 2015. How a gnss receiver is held may affect static horizontal position accuracy. PLoS One 10, e0124696.
Williams, C.E., 1997. A Comparison of Circular Error Probable Estimators for Small Samples. Air Force Institute of Technology. Master's thesis.
Yakimenko, O.A., 2013. Statistical analysis of touchdown error for self-guided aerial payload delivery systems. In: AIAA Aerodynamic Decelerator Systems (ADS) Conference, p. 1265.
Yoo, E.H., Roberts, J.E., Eum, Y., Shi, Y., 2020. Quality of hybrid location data drawn from GPS-enabled mobile phones: does it matter? Trans. GIS 24, 462–482. _eprint. https://onlinelibrary.wiley.com/doi/pdf/10.1111/tgis.12612.
Zandbergen, P.A., 2008. Positional accuracy of spatial data: non-normal distributions and a critique of the national standard for spatial data accuracy. Trans. GIS 12, 103–130.
Zandbergen, P.A., 2009. Accuracy of iphone locations: a comparison of assisted gps, wifi and cellular positioning. Trans. GIS 13, 5–25.
Zandbergen, P.A., Barbeau, S.J., 2011. Positional accuracy of assisted gps data from high-sensitivity gps-enabled mobile phones. J. Navig. 64, 381–399.